Conceptualising and Creating a Digital Edition

Jennifer Stertzer
Nick Wasmoen
Welcome to DHSI 2019!

Thanks for joining the DHSI community!

In this booklet, you will find essential course materials prefaced by some useful information about getting settled initially at UVic, finding your way around, getting logged in to our network (after you’ve registered the day before our courses begin), and so on.

Given our community’s focus on things computational, it will be a surprise to no one that we might expect additional information online for some of the classes - your instructors will let you know - or that the most current version of all DHSI-related information may be found on our website at dhsi.org.

Do check in there first if you need anything that’s not in this coursepak.

To access the DHSI wifi network, simply go into your wireless settings and connect to the “DHSI” network and enter the password “dhsi2019”.

And please don’t hesitate to be in touch with us at institut@uvic.ca or via Twitter at @AlyssaA_DHSI or @DHInstitute if we can be of any help....
DHSI Wi-Fi

Network name: DHSI
Passkey: dhsi2019
The 2019 schedule is just taking shape nicely! A very few things to confirm, add, etc, still but this is the place to be to find out what is happening when / where ...

Psst: Some Suggested Outings

If you’re here a day or two before we begin, or staying a day or two afterwards, here are a few ideas of things you might consider doing ....

Suggested Outing 1, Botanical Beach (self-organised; car needed)
A self-guided visit to the wet, wild west coast tidal shelf (and historically-significant former research site) at Botanical Beach; we recommend departing early (around 8.00 am) to catch low tide for a better view of the wonderful undersea life! Consider bringing a packed lunch to nibble-on while looking at the crashing waves when there, and then have an afternoon drink enjoying the view from the deck of the Port Renfrew Hotel.

Suggested Outing 2, Butchart Gardens (self-organised)
A shorter journey to the resplendently beautiful Butchart Gardens and, if you like, followed by (ahem) a few minutes at the nearby Church and State Winery, in the Saanich Penninsula. About an hour there by public bus from UVic, or 30 minutes by car.

Suggested Outing 3, Saltspring Island (self-organised; a full day, car/bus + ferry combo)
Why not take a day to explore and celebrate the funky, laid back, Canadian gulf island lifestyle on Saltspring Island. Ferry departs regularly from the Schwartz Bay ferry terminal, which is about one hour by bus / 30 minutes by car from UVic. You may decide to stay on forever ....

Suggested Outing 4, Paddling Victoria’s Inner Harbour (self-organised)
A shorter time, seeing Victoria’s beautiful city centre from the waterways that initially inspired its foundation. A great choice if the day is sunny and warm. Canoes, kayaks, and paddle boards are readily rented from Ocean River Adventures and conveniently launched from right behind the store. Very chill.

And more!
Self-organised High Tea at the Empress Hotel, scooter rentals, visit to the Royal BC Museum, darts at Christies Carriage House, a hangry breakfast at a local diner, whale watching, kayaking, brew pub sampling (at Spinnaker’s, Swans, Moon Under Water, and beyond!), paddle-boarding, a tour of used bookstores, and more have also been suggested!

9:00 to 4:00
Early Class Meeting: 4. [Foundations] DH For Department Chairs and Deans (David Strong Building C124, Classroom)
Further details are available from instructors in mid May to those registered in the class. Registration materials will be available in the classroom.

3:00 to 5:00
DHSI Registration (MacLaurin Building, Room A100)
After registration, many will wander to Cadboro Bay and the pub at Smuggler’s Cove OR the other direction to Shelbourne Plaza and Maude Hunter’s Pub OR even into the city for a nice meal.

Monday, 3 June 2019
Your hosts for the week are Alyssa Arbuckle, Ray Siemens, and Jannaya Friggstad Jensen.

7:45 to 8:15
Last-minute Registration (MacLaurin Building, Room A100)
Welcome, Orientation, and Instructor Overview (MacLaurin A144)
- Welcome to the Territory
- Welcome to DHSI: Ray Siemens, Alyssa Arbuckle
- Welcome from UVic: Jonathan Bengtson (University Librarian), Alexandra D'Arcy (Associate Dean Research, Humanities)

Classes in Session (click for details and locations)

1. [Foundations] Digitisation Fundamentals and their Application (Clearihue A103, Lab)
2. [Foundations] Introduction to Computation for Literary Criticism (Clearihue A102, Classroom)
3. [Foundations] Making Choices About Your Data (Digital Scholarship Commons, McPherson Library A308, Classroom)
4. [Foundations] DH For Department Chairs and Deans (David Strong Building C124, Classroom)
5. [Foundations] Developing a Digital Project (With Omeka) (Clearihue A031, Lab)
9. Out-of-the-Box Text Analysis for the Digital Humanities (Human and Social Development A160, Lab)
10. Sound and Digital Humanities (Cornett A128, Classroom)
11. Critical Pedagogy and Digital Praxis in the Humanities (Clearihue D132, Classroom)
12. Digital Humanities for Japanese Culture: Resources and Methods (McPherson Library A003, Classroom)
13. Conceptualising and Creating a Digital Edition (McPherson Library 210, Classroom)
14. Retro Machines & Media (McPherson Library 129, Classroom)
15. Geographical Information Systems in the Digital Humanities (Clearihue A105, Lab)
16. Introduction to IIIF: Sharing, Consuming, and Annotating the World’s Images (Cornett A121, Classroom)
17. Ethical Data Visualization: Taming Treacherous Data (Cornett A128, Classroom)
18. Linked Open Data and the Semantic Web (Cornett A132, Classroom)
20. Information Security for Digital Researchers (David Strong Building C114, Classroom)

12:15 to 1:15
Lunch break / Unconference Coordination Session (MacLaurin A144)
(Grab a sandwich and come on down!)
Discussion topics, scheduling, and room assignments from among all DHSI rooms will be handled at this meeting.

1:30 to 4:00
Classes in Session

Institute Lecture: Jacqueline Wernimont (Dartmouth C): “Sex and Numbers: Pleasure, Reproduction, and Digital Biopower”
Chair: Anne Cong-Huyen (U Michigan)
(MacLaurin A144)

4:10 to 5:00
Abstract: Drawing from Numbered Lives (MIT 2018), this talk will consider a long history of sex-number entanglement in Anglo-American Cultures. Drawing on historical and contemporary objects and practices, Wernimont will ask “in what ways do theories of biopower, critical gender and critical race studies, and media studies” suggest that we can understand this set of entanglements and their impacts. NB: While relevant, this talk will not include discussions of sexual trauma or violence. It will include frank discussion of sex acts and various ways of translating sexual behavior into numbers.

5:00 to 6:00
Opening Reception (University Club)

Tuesday, 4 June 2019
Wednesday, 5 June 2019

4:15 to 5:15
DHSI Conference and Colloquium Lightning Talk Session 2 (MacLaurin A144)
Chair: Kim O'Donnell (Simon Fraser U)
- Catherine Ryu (Michigan State U), “Tone Perfect: Developing a Multimodal Audio Database for Mandarin Chinese as an Open Source”
- Jessica Linzel (Brock U), “The Shopkeeper Aristocracy: Mapping Trade Networks in Colonial Niagara”
- Kirsten Painter (U Washington), “From Bogatyrs to Bread: Digitization & Online Exhibition of Rare Russian Children’s Books at the U Washington”

6:00 to 8:00
DHSI Newcomer's Gathering (Grad House Restaurant, Graduate Student Centre)
Come down, buy meal and a beverage, and make some new friends!

7:30 to 9:00
DHSI Conference and Colloquium Lightning Talk Session 3 (MacLaurin A144)
Chair: Kim O'Donnell (Simon Fraser U)
- Colleen Kolba (U South Florida), “What Comics can Teach our Students about Multimodal Literacy”
- Trish Baer (ETCL; U Victoria), “Preserving Digital Legacies: Archived Websites and Digital Discoverability”
- Suchismita Dutta (U Miami), “The Importance of Archival Transcription for Genre Building”
- Jeffrey Lawler (California State U, Long Beach), “Twining our way through the Past: Video Game Authoring as History Pedagogy”

Thursday, 6 June 2019

9:00 to Noon
Classes in Session

12:15 to 1:15
Bring your DHSI nametag and enjoy your first tipple on us! [A great opportunity for an interest group meet-up...]

1:30 to 4:00
Classes in Session

4:15 to 5:15
DHSI Conference and Colloquium Lightning Talk Session 3 (MacLaurin A144)
Chair: Kim O'Donnell (Simon Fraser U)
- Colleen Kolba (U South Florida), “What Comics can Teach our Students about Multimodal Literacy”
- Trish Baer (ETCL; U Victoria), “Preserving Digital Legacies: Archived Websites and Digital Discoverability”
- Suchismita Dutta (U Miami), “The Importance of Archival Transcription for Genre Building”
- Jeffrey Lawler (California State U, Long Beach), “Twining our way through the Past: Video Game Authoring as History Pedagogy”
Friday, 7 June 2019 [DHSI; ADHO Pedagogy SIG Conference Opening]

9:00 to Noon  Classes in Session

12:15 to 1:15  Lunch Reception / Course E-Exhibits (MacLaurin A100)

1:30 to 1:50  Remarks, A Week in Review (MacLaurin A144)

2:00 to 3:00  Joint Institute Lecture (DHSI and ADHO Pedagogy SIG Conference):
Matt Gold (CUNY Graduate Center and Association for Computers and the Humanities): “Thinking Through DH: Proposals for Digital Humanities Pedagogy”
Chair: Diane Jakacki (Bucknell U)
(MacLaurin A144)

Abstract: How do we teach digital humanities, and how should DH be taught? What, indeed, should we teach when we teach DH? This talk will present a proposal for grounding digital humanities pedagogical practice in the research interests of our students and the epistemological foundations of our methods rather than through an approach grounded more central in data and methods.

2:00 to 3:00  Joint Reception: DHSI and ADHO Pedagogy SIG Conference (University Club)
E-Poetry Event (Chris Tanasescu)
Watch this space for details, including how to participate!

DHSI Conference and Colloquium Poster/Demo Session
• Pia Russel (U Victoria); Emily Streml (U Victoria), “British Columbia’s Historical Textbooks Digital Library”
• Cody Hennesy (U Minnesota); Rachael Samberg (U California, Berkeley); Stacy Reardon (U California, Berkeley), “Finding the Haystack: Literacies for Accessing and Using Text as Data”
• Paula Johanson (ETCL; Independent Scholar), “Proving Seahorses and Juan de Fuca’s Travels in The Curve of Time”
• Tara Baillargeon (Marquette U); Elizabeth Wawrzyniak (Marquette U), “FellowsHub: J. R. R. Tolkien Fanzine Portal”
• Graham Jensen (U Victoria), “Canadian Modernist Magazines Project”
• Caterina Agostini (Rutgers U), “Art at the Time of Syphilis: A First-Person Medical Narrative in Benvenuto Cellini’s Vita”
• Lauren Elle DeGaine (ETCL; U Victoria), “Women at the Front: A Digital Exhibit of Victorian Frontpiece Illustrations”
• Adam Griggs (Mercer U); Kathryn Wright (Mercer U); Christian Pham (Mercer U); Gail Morton (Mercer U); Stephanie Miranda (Mercer U), “Digitizing Middle Georgia’s History of Slavery”

Saturday, 8 June 2019 [Conference, Colloquium, and Workshop Sessions]

8:00 to 9:00  Conference / Workshop Registration (MacLaurin A100)

The day's events are included with your DHSI registration. If you're not registered in DHSI, you're very welcome to join us by registering here as a Conference / Colloquium / Workshop participant. We'll have a nametag waiting for you!

Coffee, Tea, &c?  Looking for some morning coffee or tea, or a small nibble? Options and hours of operation for weekend campus catering are available here. Mystic Market usually opens around 10.00.

9:00 to 4:00  DHSI Conference and Colloquium Sessions
ADHO Pedagogy SIG Conference Sessions
Right2Left Workshop Sessions

9:00 to 4:00  All Day DHSI Workshop Session (click for workshop details and free registration for DHSI participants)
• 55. Introduction to Machine Learning in the Digital Humanities [8-9 June; All day, each day] (David Strong Building C124, Classroom)

9:00 to 9:10  Informal Greetings, Room Set-up (Lobby, outside Hickman 105)

Session 1
DHSI Colloquium and Conference (Hickman 105)
Digital Humanities & Literature, Chair: Kim O'Donnell (Simon Fraser U)
- Youngmin Kim (Dongguk U), “Transdiscursivity in the Convergence of Digital Humanities and World Literature”
- Caroline Winter (U Victoria), “Digitizing Adam Smith’s Literary Library”
- Kaitlyn Fralick (U Victoria); Kailey Fukushima (U Victoria); Sarah Karlson (U Victoria), “Victorian Poetry
9:10 to 10:30
ADHO Pedagogy SIG Conference (Hickman 110)
Chair: Katherine Faull (Bucknell U)
- Aaron Tucker and Nada Savicevic (Ryerson U), "Write Here, Right Now: An Open Source eTextbook for the Flipped Classroom"
- Heather McAlpine (U Fraser Valley), "Digital Meters: Using Text Encoding to Teach Literature in the Undergraduate Classroom"
- Tiina H. Airaksinen (U Helsinki), "Digital Humanities in Cultural Studies: Creating a MOOC course for University Students and A-Level Students"

Right2Left Workshop (Hickman 116)
Keynote - Nathan P. Gibson (Ludwig Maximilians U, München): "Thinking in -JTR: Reorienting the Directional Assumptions of Global Digital Scholarship"

10:30 to 10:40
Break

10:40 to Noon
Session 2
DHIS Colloquium and Conference (Hickman 105)
Digital Humanities & Society, Chair: Eleanor Reed (Hastings C)
- Joel Zapata (Southern Methodist U), "Uncovering the Southern Plains' Mexican American Civil Rights Movement"
- Ayo Oseisanwo (U Ibadan), "Online Newspaper Construction of Agitation for the Sovereign State of Biafra in Nigeria"
- Joseph Jones (U British Columbia), "Testbed for an Approach to Distant Reading: Fictions That Represent Vietnam War Resisters in Canada"
- Brendan Mackie (U California, Berkeley), "Visualizing Long-Term Cultural Change: An Example From The Birth of Civil Society"

ADHO Pedagogy SIG Conference (Hickman 110)
Chair: Laura Estill (St Francis Xavier U)
- Jane Jackson (Chinese U of Hong Kong), "Interrogating digital spaces for intercultural meaning-making"
- Ryan Ikeda (UC Berkeley), "Disrupting Digital Literacy: Situating Electronic Literature Among Public Education Initiatives"
- Christopher Church, Katherine Hepworth (U Nevada, Reno), "We’re STEAMed! A call for balancing technical instruction and disciplinary content in the digital humanities"
- Chelsea Milbourne (Cal Poly, San Luis Obispo), "Finding the Right Fit between Technology and Class Content: Reflections on Including Web Development in a Digital Storytelling Course"

Right2Left Workshop (Hickman 116)
- Edward “Eddie” Surman (Claremont Graduate U), "Qualitative Digital Text Analysis and #Right2Left Languages: A Demonstration of Atlas.ti using the Hebrew Bible"

Noon to 1:10
Lunch (We recommend Mystic Market on weekends!)

1:10 to 2:30
Session 3
DHIS Colloquium and Conference (Hickman 105)
Digital Humanities & Community, Chair: Claire Carlin (U Victoria)
- Pia Russel (U Victoria); Emily Stremel (U Victoria), "Mentorship and disability: Supporting disabled employees in digital humanities"
- Amy Lueck (Santa Clara U), "Virtually Emplacing Indigenous Memory"
- Md. Shehabul Alam (National U Bangladesh), "Integrating Library Service with Union Information and Service Center: A Joint Initiative towards Digital Bangladesh"
- Veronica Gomez (Instituto de Humanidades y Ciencias Sociales (HuCSo) - UNL-CONICET), "Latin American E-literature and Location: The Nation Revisited in Electronic Literature Organization (ELO)"

ADHO Pedagogy SIG Conference (Hickman 110)
Chair: Chris Tănăsescu (UC Louvain)
- Laura Estill (St Francis Xavier U), "One Assignment, Three Ways: Assessing DH Projects in a Literature Course"
- Felix Bayode Oke, Stella N. Kpolugbo (Anchor U Lagos), "The Multimodal Technique as a Pedagogical Tool in Pelu Awofeso’s White Lagos: A Definitive and Visual Guide to the Eyo Festival"
- Shu Wan (U Iowa), "A digital "historical gaze" of Chinese students in Iowa, 1911-1930"
- Francesca Giannetti (Rutgers U, New Brunswick), "So near while apart: Correspondence Editions as Critical Library Pedagogy and Digital Humanities Methodology"

Right2Left Workshop (Hickman 116)
- Najla Jarkas (American U Beirut) and David Joseph Wrisley (NYU Abu Dhabi), "RTL Software Localization and Digital Humanities: the Case Study of Translating Voyant Tools into Arabic"
2:30 to 2:40  
Break

<table>
<thead>
<tr>
<th>2:40 to 4:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session 4</td>
</tr>
</tbody>
</table>
| DHSI Colloquium and Conference ([Hickman 105](#))  
Digital Humanities & Media, Chair: Caroline Winter (U Victoria)  
- Olivia Wikle (U Idaho), "Listening with Our Eyes: Using Topic Modeling, Text Analysis, and Sound Studies Methodologies to Explore Literary Soundscapes"  
- Olin Bjork (U Houston-Downtown), "Dramatic Redundancy: Interactive Transcripts and Multimodal Performance Editions"  
- Ashleigh Cassermer-Stanfield (U Chicago), "Sonifying Hamlet and Reading the Room"  
ADHO Pedagogy SIG Conference ([Hickman 110](#))  
Chair: Aaron Tucker (Ryerson U)  
Youngmin Kim (Dongguk U), "Teaching Digital Humanities and World Literature in Class"  
Alice Fleerackers, Juan Pablo Alperin, Esteban Morales, Remi Kalir (Simon Fraser U, U Colorado Denver), "Online annotations in the classroom: How, why, and what do students learn from annotating course material?"  
Andie Silva (York C and Graduate Center, CUNY), "Keeping it Local: Undergraduate DH as Feminist Practice"  
Right2Left Workshop ([Hickman 116](#))  
- Joanna Byszuk (Institute of Polish Language, Polish Academy of Sciences, Warsaw/Computational Stylistics Group) and Alexey Khismatulin (Institute of Oriental Manuscripts, Russian Academy of Sciences, Saint Petersburg), "Attribution of Authorship for Medieval Persian Quasidas with Stylometry"  
- Ilan Benattar (New York U), "#Right2Left Biblical Translations in Jewish Textual History: Case Studies in Judeo-Arabic and Judeo-Spanish"

---

**Sunday, 9 June 2019 [Workshop Sessions]**

<table>
<thead>
<tr>
<th>8:00 to 5:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>The day’s events are included with your DHSI registration. If you’re not registered in DHSI, you’re very welcome to join us by registering here as a Conference / Colloquium / Workshop participant. We’ll have a nametag waiting for you!</td>
</tr>
<tr>
<td>Coffee, Tea, &amp;c?</td>
</tr>
<tr>
<td>Looking for some morning coffee or tea, or a small nibble? Options and hours of operation for weekend campus catering are available here. Mystic Market usually opens around 10.00.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9:00 to 4:00</th>
</tr>
</thead>
</table>
| All Day Workshop Sessions (click for workshop details and free registration for DHSI participants)  
- 55. Introduction to Machine Learning in the Digital Humanities [8-9 June; All day, each day] ([David Strong Building C124](#), Classroom)  
- 56. Pedagogy of the Digitally Oppressed: Anti-Colonial DH Methods and Praxis [9 June; All Day] ([Hickman 116](#), Classroom)  
- 57. Natural Language Processing and Network Coding Apps for Text & Textual Corpus Analysis in the Humanities [9 June; All Day] ([David Strong Building C114](#), Classroom) |

<table>
<thead>
<tr>
<th>9:00 to Noon</th>
</tr>
</thead>
</table>
| AM Workshop Sessions (click for workshop details and free registration for DHSI participants)  
- 59. 3D Visualization for the Humanities [9 June; AM] ([Cornett A229](#), Classroom)  
- 60. It’s All Relational: AbTeC’s Indigenous Video Game Workshops as Storytelling Praxis [9 June; AM] ([Cornett A121](#), Classroom)  
- 61. Spatial DH: De-Colonizing Cultural Territories Online [9 June; AM] ([Clearihue D130](#), Classroom)  
- 63. Creating a CV for Digital Humanities Makers [9 June; AM] ([David Strong Building C108](#), Classroom) |

<table>
<thead>
<tr>
<th>Noon to 1:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lunch (We recommend Mystic Market on weekends!)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1:00 to 4:00</th>
</tr>
</thead>
</table>
| PM Workshop Sessions (click for workshop details and free registration for DHSI participants)  
- 65. Indigenous Futurities in the Classroom and Beyond [9 June; PM] ([Cornett A124](#), Classroom)  
- 66. DHSI Knits: History of Textiles and Technology [9 June; PM] ([Fine Arts 109](#), Classroom)  
- 68. Linked Open Datafication for Humanities Scholars [9 June; PM] ([McPherson Library A003](#), Classroom)  
- 69. Stylo - WYSIWYM Text Editor for Humanities Scholars [9 June; PM] ([McPherson Library A025](#), Classroom) |

---

After the day, many will wander to Cadboro Bay and the pub at Smuggler's Cove OR the other direction to Shelbourne Plaza and Maude Hunter's Pub OR even into the city for a bite to eat.

---

**Monday, 10 June 2019**
Your hosts for the week are Ray Siemens and Jannaya Friggstad Jensen.

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:45 to 8:15</td>
<td>DHSI Last-minute Registration (MacLaurin A100)</td>
<td></td>
</tr>
<tr>
<td>8:30 to 10:00</td>
<td>Welcome, Orientation, and Instructor Overview (MacLaurin A144)</td>
<td></td>
</tr>
<tr>
<td>10:15 to Noon</td>
<td>Classes in Session (click for details and locations)</td>
<td></td>
</tr>
<tr>
<td>12:15 to 1:15</td>
<td>Lunch break / Unconference Coordination Session (MacLaurin A144)</td>
<td></td>
</tr>
<tr>
<td>12:15 to 1:15</td>
<td>&quot;Mystery&quot; Lunches</td>
<td></td>
</tr>
<tr>
<td>1:30 to 4:00</td>
<td>Classes in Session</td>
<td></td>
</tr>
<tr>
<td>4:10 to 5:00</td>
<td>Institute Lecture: Angel David Nieves (San Diego State U): &quot;3D Mapping and Forensic Traces of Testimony: Documenting Apartheid-Era Crimes Through the Digital Humanities&quot; Chair: Constante Crompton (U Ottawa) (MacLaurin A144)</td>
<td></td>
</tr>
<tr>
<td>4:10 to 5:00</td>
<td>Abstract: In 1989 the killing of a queer, 14-year-old youth in Winnie Mandela's house named Stompie Seipei (an event that few in South Africa are willing to recall, let alone discuss, in any detail) -- is perhaps one of the most glaring examples where the queer and activist community was suppressed or erased from anti-apartheid/liberation histories. Digital humanities may actually help both reconstruct and recover a history that is still very early in the telling, despite what is commonly believed about the liberation struggle and the contributions of queer activists in the dismantling of apartheid. Perhaps it could explain why a youth such as Seipei was killed -- or at the very least, provide a more complex and messy narrative that permits one to know more how the history of queer anti-apartheid activists was suppressed. This talk outlines a methodology for &quot;messy thinking and writing&quot; in the digital humanities that -- through a queer and feminist intersectional framework -- permits a more complex layering of oral histories and 3D historical reconstructions.</td>
<td></td>
</tr>
<tr>
<td>5:00 to 6:00</td>
<td>Reception (University Club)</td>
<td></td>
</tr>
<tr>
<td>9:00 to Noon</td>
<td>Classes in Session</td>
<td></td>
</tr>
<tr>
<td>12:15 to 1:15</td>
<td>Lunch break / Unconference</td>
<td></td>
</tr>
<tr>
<td>12:15 to 1:15</td>
<td>&quot;Mystery&quot; Lunches</td>
<td></td>
</tr>
</tbody>
</table>
### Wednesday, 12 June 2019

<table>
<thead>
<tr>
<th>Time</th>
<th>Event Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:30 to 4:00</td>
<td>Classes in Session</td>
</tr>
<tr>
<td>4:15 to 5:15</td>
<td>DHSI Conference and Colloquium Lightning Talk Session 4 (MacLaurin A144) Chair: Lindsey Seatter (U Victoria)</td>
</tr>
<tr>
<td></td>
<td>- Ashley Caranto Morford (U Toronto); Kush Patel (U Michigan); Arun Jacob (McMaster U), “#OurDHIs anti-colonial: Questions and challenges in dismantling colonial influences in digital humanities pedagogy”</td>
</tr>
<tr>
<td></td>
<td>- Luis Meneses (ETCL; U Victoria), “Identifying Changes in the Political Environment in Ecuador”</td>
</tr>
<tr>
<td></td>
<td>- Laura Horak (Carleton U), “Building the Transgender Media Portal”</td>
</tr>
<tr>
<td>6:00 to 8:00</td>
<td>DHSI Newcomer’s Gathering (Grad House Restaurant, Graduate Student Centre) Come down, buy meal and a beverage, and make some new friends!</td>
</tr>
</tbody>
</table>

### Thursday, 13 June 2019

<table>
<thead>
<tr>
<th>Time</th>
<th>Event Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:30 to 4:00</td>
<td>Classes in Session</td>
</tr>
<tr>
<td>4:10 to 5:00</td>
<td>Institute Lecture: Karina van Dalen-Oskam (Huygens Institute and U Amsterdam; Alliance of Digital Humanities Organizations): “The Riddle of Literary Quality: Some Answers” Chair: Aaron Mauro (Penn State, Behrend C) (MacLaurin A144)</td>
</tr>
<tr>
<td></td>
<td>Abstract: What is literature, and can you measure it? That is the key question of the project The Riddle of Literary Quality. “The Riddle” is a research project of the Huygens Institute for the History of the Netherlands (Amsterdam) in collaboration with the Fryeke Akademy (Leeuwarden) and the Institute for Logic, Language and Computation (University of Amsterdam). The Riddle combines computational analysis of writing style with the results of a large online survey of readers, completed by almost 14,000 participants. In my talk, I will go into...</td>
</tr>
</tbody>
</table>
some of the main results of the project.

**Friday, 14 June 2019**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 to Noon</td>
<td>Classes in Session</td>
</tr>
<tr>
<td>12:15 to 1:15</td>
<td>Lunch Reception / Course E-Exhibits (MacLaurin A100)</td>
</tr>
<tr>
<td>1:30 to 2:00</td>
<td>Closing, DHSI in Review (MacLaurin A144)</td>
</tr>
</tbody>
</table>

**Contact info:**
email: institut@uvic.ca  
P: 250-472-5401  
F: 250-472-5881
Conceptualising and Creating a Digital Edition - Coursepack

Instructors: Jennifer Stertzer and Nikolaus Wasmoen

This course will explore all aspects of conceptualizing, planning for, and creating a digital edition. It provides a basic introduction to the various types of digital editions, the practice of editing in the digital age, and a survey of the many digital tools available to serve project goals. Approaching a digital edition means taking time to think about how end-users will want to work with a particular edition. Beginning with the research and analytical needs of end-users in mind, editors are better able to develop effective editorial strategies that will result in a dynamic, useful, and usable, digital edition.

In this course, participants will engage in hands-on learning and group discussions related to project conceptualization, editorial policies and processes, and the selection and use of digital tools that can serve the needs of researchers and other end-users. Participants will bring a few sample materials they are working with. We will use these in a class project - creating a digital edition over the course of the week using skills learned in each session. Our goal is for participants to return to their home institutions ready and able to build upon, enhance, and transform these initial ideas into robust digital editions.

What to Bring

Bring digital copies of sample documents for the project(s) you are interested in working on. We recommend at least 10-20 pages, including:

- Select representatives of the more difficult and complex documents that you have
- Select a broad range of the kinds of text formats that you want to include
- If you have transcriptions of the texts or have already begun annotations, bring them.

A laptop computer

- Mac or PC is fine, but some Chromebooks may not be able to support some of the software we will be running.

Recommended Guides/Texts


This guide provides a deep understanding of the history and practice of scholarly editing, familiarity with the terminology editors employ, and advice on the major decisions to be made. It is chiefly addressed towards the creation of traditional print editions. Free online edition at [http://gde.upress.virginia.edu](http://gde.upress.virginia.edu)

*Editing Historical Documents: A Handbook of Practice*, Michael E. Stevens and Steven B. Berg (AltaMira Press, 1997)
This book provides examples of practice from a variety of different editing projects and covers all the major editing tasks--transcription, annotation, indexing, etc. Excerpts available on documentaryediting.org.

Schedule

*Please note, each session with begin with an overview of the topics. We will then, as a class, decide what topics to explore and discuss in-depth. We will also tailor the hands-on time to meet the needs of the participants.

Day 1


Late morning – Hands-on time – participants will share current state of project, their end-user goals, and working plans for their digital edition, as well as discuss the kinds of documents/items they are working with. The goal of this session is to help participants craft an effective mission statement as a basis for elaborating further editorial policies and objectives.

Afternoon – Overview of digital editions – what’s currently available and in-progress. What are the features, benefits, and drawbacks of each type? We’ll also examine the intersection of digital humanities and digital editing.

Late afternoon – Hands-on time – participants will compare examples from existing digital editions available on the web and be given access to some test environments in which they can experiment with various features/approaches.

Suggested Readings:


* Editing Historical Documents, Preface and Introduction


* Schulze, “Stories that Work: Our Philological Future” Presidential Address from Joint
Conference of the Society for Textual Scholarship and the Association for Documentary Editing, University of Nebraska-Lincoln, 2015.


Hughes, "Project Planning and Funding," in Digitizing Collections, 164-209.


Day 2

Morning – Editorial Processes and Policies: Transcription. We will discuss how editorial decisions about transcription policies and strategies can be made to respond to different types of end-user goals for different kinds of editorial projects. Discuss various tools and major publication platforms support for transcription, including connecting transcriptions to facsimiles or other images.

Late Morning – Hands-on transcription practice.

Afternoon – We will continue the morning’s conversation of editorial processes, policies, and practices based on participant's specific interests.

Late Afternoon – Hands-on transcription and verification practice.

Suggested Readings:


Editing Historical Documents, chapters. 3-5.


TEI, Gentle Intro to XML.
Day 3

*Morning* – Editorial Processes and Policies: Annotation. We will discuss different styles of annotation and how these are implemented within various digital editing toolsets and publication platforms. Attention will be paid to the variety of forms of annotation.

*Late Morning* – Hands-on practice annotating using TEI/XML markup.

*Afternoon* – Editorial Processes and Policies: Annotation and Databases. Discuss methods and tools for annotating large collections of documents and materials efficiently and accurately. Discuss differences between metadata formats, especially markup versus database structures.

*Late Afternoon* – Hands-on practice annotating using fields in Omeka test environment.

Suggested Readings:

*A Guide*, Chapter 7 (sections III–VII)

*Editing Historical Documents*, Chapters 6–7.

Day 4

*Morning* – Editorial Processes and Policies: Indexing and Prosopography. Discuss strategies for tracking and making accessible to users biographical, historical, geographical, social, financial, and other types of information across a large collection using techniques for indexing and prosopography. Explore how indexing and prosopography relate to information architecture including navigation and search.

*Late Morning* – Hands-on practice using a centralized prosopography within a TEI-XML publication stack and within an Omeka install or other database driven platform such as a Drupal-based digital edition.

*Afternoon and Late Afternoon* – Rapid prototyping of projects using test environments, integration of previous exercises into a package ready to be implemented in a given platform/technology stack.

Suggested Readings:

TEI P5: Guidelines for Electronic Text Encoding and Interchange, "Chapter 13: Names, Dates, People, and Places."

Day 5
Morning – How to evaluate the goals of the project, ideal workflow and editorial environment, and publication/user interface plan and make decisions about tools and platform. Discussion of how to plan, design, organize, layout, and manage your edition and site. Minimal approaches and rapid prototyping strategies discussed.

Course wrap-up as well as additional hands-on time. Discussion of what the future holds for digital documentary editions as well as the newest developments in visualizations, sound, images, and moving images.

Suggested Readings:

- Cameron Blevins, “Martha Ballard’s Diaries,” (4 blog posts), Cameron Blevins, 2009-2010.
- Gil, “The User, the Learner and the Machines We Make,” Minimal Computing (21 May 2015).

Links to Readings:

- Cameron Blevins, “Martha Ballard’s Diaries,” (4 blog posts), Cameron Blevins, 2009-2010.
- Quinn Dombrowski, “When Not to Use Drupal” [see coursepack] and “Choosing a platform for your project website” [http://digitalhumanities.berkeley.edu/blog/13/12/04/choosing-platform-your-project-website](http://digitalhumanities.berkeley.edu/blog/13/12/04/choosing-platform-your-project-website).

Alex Gil, “The User, the Learner and the Machines We Make,” *Minimal Computing* (21 May 2015). [http://go-dh.github.io/mincomp/thoughts/2015/05/21/user-vs-learner/]


Lorna Hughes, “Managing a Digitization Project,” in *Digitizing Collections: Strategic Issues for the Information Manager* (Facet Publishing, 2009), 164-209. [https://drive.google.com/file/d/1JmxILGeyJJKrH99eA5DXqc4KCZbrT1QS/view?usp=sharing]


MLA Committee on Scholarly Editing, “Considering the Scholarly Edition in the Digital Age: A White Papers of the Modern Language Association’s Committee on Scholarly Editions,” [https://scholarlyeditions.commons.mla.org/2015/09/02/cse-white-paper/]


Kenneth M. Price
University of Nebraska - Lincoln, kprice2@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/englishfacpubs
Part of the English Language and Literature Commons

http://digitalcommons.unl.edu/englishfacpubs/69

This Article is brought to you for free and open access by the English, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications -- Department of English by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Conceptualising and Creating a Digital Edition - Coursepack

Instructors: Jennifer Stertzer and Nikolaus Wasmoen

This course will explore all aspects of conceptualizing, planning for, and creating a digital edition. It provides a basic introduction to the various types of digital editions, the practice of editing in the digital age, and a survey of the many digital tools available to serve project goals. Approaching a digital edition means taking time to think about how end-users will want to work with a particular edition. Beginning with the research and analytical needs of end-users in mind, editors are better able to develop effective editorial strategies that will result in a dynamic, useful, and usable, digital edition.

In this course, participants will engage in hands-on learning and group discussions related to project conceptualization, editorial policies and processes, and the selection and use of digital tools that can serve the needs of researchers and other end-users. Participants will bring a few sample materials they are working with. We will use these in a class project - creating a digital edition over the course of the week using skills learned in each session. Our goal is for participants to return to their home institutions ready and able to build upon, enhance, and transform these initial ideas into robust digital editions.

What to Bring

Bring digital copies of sample documents for the project(s) you are interested in working on. We recommend at least 10-20 pages, including:

- Select representatives of the more difficult and complex documents that you have
- Select a broad range of the kinds of text formats that you want to include
- If you have transcriptions of the texts or have already begun annotations, bring them.

A laptop computer

- Mac or PC is fine, but some Chromebooks may not be able to support some of the software we will be running.

Recommended Guides/Texts


This guide provides a deep understanding of the history and practice of scholarly editing, familiarity with the terminology editors employ, and advice on the major decisions to be made. It is chiefly addressed towards the creation of traditional print editions. Free online edition at http://gde.upress.virginia.edu

Editing Historical Documents: A Handbook of Practice, Michael E. Stevens and Steven B. Berg (AltaMira Press, 1997)
This book provides examples of practice from a variety of different editing projects and covers all the major editing tasks—transcription, annotation, indexing, etc. Excerpts available on documentaryediting.org.

Schedule

*Please note, each session will begin with an overview of the topics. We will then, as a class, decide what topics to explore and discuss in-depth. We will also tailor the hands-on time to meet the needs of the participants.

Day 1


Late morning – Hands-on time – participants will share current state of project, their end-user goals, and working plans for their digital edition, as well as discuss the kinds of documents/items they are working with. The goal of this session is to help participants craft an effective mission statement as a basis for elaborating further editorial policies and objectives.

Afternoon – Overview of digital editions – what’s currently available and in-progress. What are the features, benefits, and drawbacks of each type? We’ll also examine the intersection of digital humanities and digital editing.

Late afternoon – Hands-on time – participants will compare examples from existing digital editions available on the web and be given access to some test environments in which they can experiment with various features/approaches.

Suggested Readings:


Editing Historical Documents, Preface and Introduction


Schulze, “Stories that Work: Our Philological Future” Presidential Address from Joint
Conference of the Society for Textual Scholarship and the Association for Documentary Editing, University of Nebraska-Lincoln, 2015.


Hughes, "Project Planning and Funding," in *Digitizing Collections*, 164-209.


**Day 2**

*Morning* – Editorial Processes and Policies: Transcription. We will discuss how editorial decisions about transcription policies and strategies can be made to respond to different types of end-user goals for different kinds of editorial projects. Discuss various tools and major publication platforms support for transcription, including connecting transcriptions to facsimiles or other images.

*Late Morning* – Hands-on transcription practice.

*Afternoon* – We will continue the morning’s conversation of editorial processes, policies, and practices based on participant's specific interests.

*Late Afternoon* – Hands-on transcription and verification practice.

Suggested Readings:


*Editing Historical Documents*, chapters. 3-5.

Rosenberg, “Documentary Editing,” in *Electronic Textual Editing*.

TEI, Gentle Intro to XML.
Day 3

Morning – Editorial Processes and Policies: Annotation. We will discuss different styles of annotation and how these are implemented within various digital editing toolsets and publication platforms. Attention will be paid to the variety of forms of annotation.

Late Morning – Hands-on practice annotating using TEI/XML markup.

Afternoon – Editorial Processes and Policies: Annotation and Databases. Discuss methods and tools for annotating large collections of documents and materials efficiently and accurately. Discuss differences between metadata formats, especially markup versus database structures.

Late Afternoon – Hands-on practice annotating using fields in Omeka test environment.

Suggested Readings:

A Guide, Chapter 7 (sections III–VII)

Editing Historical Documents, Chapters 6–7.

Day 4

Morning – Editorial Processes and Policies: Indexing and Prosopography. Discuss strategies for tracking and making accessible to users biographical, historical, geographical, social, financial, and other types of information across a large collection using techniques for indexing and prosopography. Explore how indexing and prosopography relate to information architecture including navigation and search.

Late Morning – Hands-on practice using a centralized prosopography within a TEI-XML publication stack and within an Omeka install or other database driven platform such as a Drupal-based digital edition.

Afternoon and Late Afternoon – Rapid prototyping of projects using test environments, integration of previous exercises into a package ready to be implemented in a given platform/technology stack.

Suggested Readings:

TEI P5: Guidelines for Electronic Text Encoding and Interchange, "Chapter 13: Names, Dates, People, and Places."

Day 5
Morning – How to evaluate the goals of the project, ideal workflow and editorial environment, and publication/user interface plan and make decisions about tools and platform. Discussion of how to plan, design, organize, layout, and manage your edition and site. Minimal approaches and rapid prototyping strategies discussed.

Course wrap-up as well as additional hands-on time. Discussion of what the future holds for digital documentary editions as well as the newest developments in visualizations, sound, images, and moving images.

Suggested Readings:


- Cameron Blevins, “Martha Ballard’s Diaries,” (4 blog posts), Cameron Blevins, 2009-2010.

- Gil, “The User, the Learner and the Machines We Make,” Minimal Computing (21 May 2015).


Links to Readings:


Cameron Blevins, “Martha Ballard’s Diaries,” (4 blog posts), Cameron Blevins, 2009-2010.


Quinn Dombrowski, “When Not to Use Drupal” [see coursepack] and “Choosing a platform for your project website” [http://digitalhumanities.berkeley.edu/blog/13/12/04/choosing-platform-your-project-website]
https://quod.lib.umich.edu/cgi/t/text/id/e/etlc/13607061.0001.001/1:8/--proofs-of-genius-collected-editions-from-the-american?g=dculture;rgn=div1;view=fulltext;xc=1

Alex Gil, “The User, the Learner and the Machines We Make,” Minimal Computing (21 May 2015). http://go-dh.github.io/mincomp/thoughts/2015/05/21/user-vs-learner/

[https://wp.nyu.edu/archivesandpublichistory/people-2/cathy-moran-hajo/scholarly-editing-in-a-web-2-0-world/]


Lorna Hughes, “Managing a Digitization Project,” in Digitizing Collections: Strategic Issues for the Information Manager (Facet Publishing, 2009), 164-209. 
[https://drive.google.com/file/d/1JmxILGeyJJKrH99eA5DXqc4KCZbrT1QS/view?usp=sharing]

[http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1193&context=libraryscience]


[http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1068&context=englishfacpubs]


Claire Rydell and Caroline Winterer, "Visualizing Ben Franklin’s Correspondence Network, 1757-1763," Mapping the Republic of Letters Project, Stanford University, October 2012.  
[http://republicofletters.stanford.edu/casestudies/franklin.html]

Kenneth M. Price <kprice_at_unlnotes_dot_unl_dot_edu>, University of Nebraska-Lincoln

Abstract

What are the implications of the terms we use to describe large-scale text-based electronic scholarship, especially undertakings that share some of the ambitions and methods of the traditional multi-volume scholarly edition? And how do the conceptions inherent in these choices of language frame and perhaps limit what we attempt? How do terms such as edition, project, database, archive, and thematic research collection relate to the past, present, and future of textual studies? Kenneth M. Price considers how current terms describing digital scholarship both clarify and obscure our collective enterprise. Price argues that the terms we use have more than expressive importance. The shorthand we invoke when explaining our work to others shapes how we conceive of and also how we position digital scholarship.

What are the implications of the terms we use to describe large-scale text-based electronic scholarship, especially undertakings that share some of the ambitions and methods of the traditional multi-volume scholarly edition? What genre or genres are we now working in? And how do the conceptions inherent in these choices of language frame and perhaps limit what we attempt? How do terms such as edition, project, database, archive, and thematic research collection relate to the past, present, and future of textual studies? Drawing on a range of resources, including the Walt Whitman Archive, I consider how current terms describing digital scholarship both clarify and obscure our collective enterprise. In addition, I will use the final term, thematic research collection, to discuss yet-to-be-developed parts of the Whitman Archive dealing with place-based cultural analysis and translation studies as a way to illustrate the expansive possibilities of this new model of scholarship.

Digital textual studies seem to me inadequately described by the terms now available. Project is amorphous; archive and edition are heavy with associations carried over from print culture; database is both too limiting and too misleading in its connotations; and digital thematic research collection lacks a memorable ring and pithiness. The terms we use have more than expressive importance. The shorthand we invoke when explaining our work to others shapes how we conceive of and also how we position digital scholarship. We need a new term that is vivid enough to be memorable, elastic enough to cover a class of like things, and yet restrictive enough to allow us to include some scholarly undertakings and not others. Ordinary readers and academics alike rely heavily on the work of editors, yet the standing of editors in the academy has for decades been shaky at best. For many people, electronic work is even more dubious: what relatively short history it has is marked by distrust, denigration, and dismissal. We all know the charges, however distorted they may be: digital work is ephemeral, unvetted, chaotic, and unreliable. When suspicion of the value of editing combines with suspicion of the new medium, we have a hazardous mix brewing. There is a danger that if humanities scholars do not undertake the key work of textual transmission, this work will be done by librarians and systems engineers — that is, it will be done by people with less specialized knowledge of the content. In the fraught circumstances of the academy, driven by a prestige economy, humanities scholars are well advised to be highly self-conscious about what we do and how we describe it.

Edition

What do we mean when we use the term edition? Even among print editions, there are a number of variations: selected editions, reader's editions, and some boldly claiming to be authoritative or definitive editions. The descriptive word "scholarly" has been applied to numerous approaches: authorial or social, critical or documentary, genetic, eclectic, or best text [Stauffer 2007]. Successful scholarly editions yield a text established on explicitly stated principles by a person or a group with specialized knowledge about textual scholarship and the writer or writers involved. What makes the edition scholarly, of course, is the rigor with which the text is reproduced or altered and the expertise deployed in the offering of suitable introductions, notes, and textual apparatus.
For those of us who work on prominent figures who have received previous treatment, our own textual work intervenes in an ongoing editorial tradition. A fundamental and often vexingly difficult question is, what should go in an edition? Like most digital editing endeavors, the Walt Whitman Archive must proceed with an awareness of the print past — in our case, especially of two significant attempts to present Whitman in scholarly editions: The Complete Writings of Walt Whitman (G. P. Putnam’s Sons, 1902) and The Collected Writings of Walt Whitman (New York University Press, Peter Lang, and the University of Iowa Press, 1961-2004). This awareness produces competing impulses: we want to benefit from and respond to past work, but we also want to avoid constraints on thought and action that were a result of print-based limitations. As editors, we acknowledge the ways of knowing that are enabled by our predecessors — they are the cultural history we inherit — but our job is also to extend their efforts and to produce new ways of knowing that are responsive to cultural, critical, and technological changes (as well as the discovery of documents and the development of new biographical insights) that have happened in the interim.

The language of the Walt Whitman Archive’s first grant application to the National Endowment for the Humanities (NEH), drafted in 1999, shows how we were thinking of our digital work as in dialogue with the print past. We wrote,

Our goal has been to build upon the strengths of the Collected Writings edition, most volumes of which were supported by grants from the National Endowment for the Humanities. The amount of Whitman’s work is so huge that no two scholars could hope to edit it effectively in a lifetime — fourteen scholars spent the better parts of their careers editing the materials that now make up the Collected Writings. But we do believe that developments in electronic scholarship have made it possible to enhance and supplement the Collected Writings by editing the materials that have not yet been included (and adding the materials that have come to light since the Collected Writings volumes were issued) and by digitizing and encoding the Collected Writings so that these disparate volumes — which often arrange material in confusing and contradictory ways — can function seamlessly and so that Whitman’s materials can be presented effectively in any number of new configurations: by genre, by date, by keyword, by subject. The electronic environment can also allow us to make available not just printed transcriptions of Whitman’s manuscripts, letters, and books, but to deliver actual facsimile images of the original documents. [Folsom and Price 1999]

It would be fair to acknowledge that Ed Folsom, co-director of the Whitman Archive, and I have had evolving views of the relationship between our undertaking and its most recent print predecessor, the Collected Writings of Walt Whitman. Our gradually shifting views have been shaped in part by discussions with publishers. At various times, we considered entering into agreements with two publishers — Primary Source Media and the University of Virginia Press — and in fact reached late stages of contract negotiations with each of them. Initially, we reasoned that if a publisher could secure the permissions for us to use the copyrighted material in the twenty-two volumes of the Collected Writings published by New York University Press, a significant amount of work, some of it meticulously done, could be preserved and extended.[1] Of course this line of thinking raised a key issue: if a new publisher had to pay for the permissions, the site, or some significant part of it, would need to be commercial in order to recover these and other costs, and perhaps make a profit as well. We were not absolute purists committed always to building a completely free site. In fact, there were extended periods when we were convinced that such an approach would not be possible for a poet like Whitman who left so much debris everywhere. We thought that editing such chaos would demand the combined resources and know-how of the scholarly, library and archival, and publishing communities. Gay Wilson Allen, a general editor of the Collected Writings, commented about editing Whitman, "Sometimes his exhausted editors almost wish that he had had two or three good house fires, and considering the houses he lived in, it is also astonishing that he did not" [Allen 1963, 8].

For us, then, a key question emerged: would we conceive of the Whitman Archive primarily as being the remediation of the Collected Writings? We recognized that our relationship to the Collected Writings is problematic: we have a half-century of valuable editorial work collected there, but the limits of a print format make this edition a trial to use. The Collected Writings has been the standard edition, the edition cited by American literary scholarship over the past few decades, but much of the work needs to be done again and the presentation re-conceptualized. We struggled to come to terms with a giant from the print past. And yet this monumental edition was both enormous and characterized by some inexplicable omissions, most notably Whitman's revelatory poetry manuscripts. As our initial grant application pointed out,

[We] have Whitman’s laundry lists in print; we have the business cards of his sidewalk repairman in print, but we don’t have the manuscripts of “Song of Myself” in print... His poetry manuscripts and periodical publications reveal, among other surprising things, a Whitman who devoted extraordinary time and care to the creation of a poetry that appeared to be quick and spontaneous; his manuscripts expose an artist whose casual, loafing persona was in fact the result of intensive and obsessive artistic labor. [Folsom and Price 1999]

In retrospect, it is clear that we have responded to the Collected Writings not by “digitizing and encoding” it but by prioritizing work on material not included there: photographs, bibliography, full texts of various editions of Leaves of Grass, archival guides to manuscripts, transcriptions of manuscripts, contemporary reviews of Whitman’s writings, and...
so on. If we were the first editors of Whitman, this order of development for an online resource would have been peculiar. Certainly some of Whitman's prose, *Democratic Vistas* or *Specimen Days*, for example, or his correspondence might rank ahead of some of these items in most people's sequencing list. But of course we do work within a historical context, and what has seemed most pressing (and perhaps most fundable) have been those things altogether neglected or poorly treated by the *Collected Writings*.

Sometimes we learn to be thankful for our failures, and I am certainly grateful now that our negotiations with publishers always went bust. I think — because of a recent NEH challenge grant to be discussed later — that the *Whitman Archive* is in an unusual position: we now have a team of people and the resources in place so that, with reasonable luck, we ought to be able to achieve a more expansive *Whitman Archive* than the already quite extensive site, and to keep it freely available. There are of course examples of other large, not to say gargantuan, free sites. But we should not underestimate the challenges attendant on making vast amounts of material freely available since "free" means no cost to the end user, not the creators.

It is reasonable to wonder why Whitman needs to be edited if there have been two previous scholarly editions. And it is reasonable to acknowledge, in response, motivations that have nothing to do with the electronic medium specifically. Editorial work is one way to engage in historical criticism and to help bring the past into the present so it may live in the future. Although the shelf life of a scholarly edition far exceeds that of a monograph, scholarly editions begun half a century ago for Whitman in one case, or a century ago in the other, now seem inadequate. Their approaches require rethinking, not to mention the need to add material and convey new discoveries. Editions of modern writers are almost always selective. Still, a selection ought to include the most important items. If asked to pick Whitman's most important single text, many would name the first publication of *Leaves of Grass* (1855). Here Whitman was at his boldest and most experimental, and the book has elicited some memorable reactions over the past 150-plus years: Ralph Waldo Emerson found it to be "the most extraordinary piece of wit and wisdom that America has yet contributed" [Emerson 1938-1994, 446]. William Carlos Williams called the first *Leaves* "a book as important as we are likely to see in the next thousand years" (Williams, quoted in Hindus 1955, 3). Clearly, this is a highly significant book. And we might expect the 1855 *Leaves* to be the highlight of an edition of Whitman's writings. Strangely enough, neither *The Collected Writings of Walt Whitman* nor the earlier *Complete Writings of Walt Whitman* bothered to include it.

How do we explain this omission? To a large extent, this odd result stems from twentieth-century editorial practices for establishing authoritative or definitive texts that encouraged the selection of a single text. The economics of print publishing — combined with the dominant editorial theories of the mid twentieth-century — made the so-called deathbed edition of *Leaves of Grass* the one most commonly featured in various commercial and scholarly editions. That final authorized printing of Whitman's book is in fact presented twice in the New York University Press edition: it serves as the basis of both the Comprehensive Reader's Edition and the *Leaves of Grass* Variorum. The deathbed edition is remarkable, but it could not be described as Whitman's most daring, most experimental, or even most coherent volume.

Print editions of Whitman tended to falter when dealing with multiplicity, whether of versions or of authorship. Whitman is well known as the writer who couldn't stop writing, revising, and reissuing *Leaves of Grass* (a book that appeared in six radically distinct American editions in his lifetime). Less well known is Whitman's involvement in collaborative enterprises. In fact, when we think of the great collaborators in literary history, Whitman hardly jumps to mind. Instead, we remember that Whitman was so self-reliant that for the first edition he more or less did everything: wrote the poetry, designed the book, set some of the type, distributed the book, and anonymously reviewed it. He appears to be dead set against even the largely invisible and ordinarily neglected forms of social authorship, a poet acting out the role of the solitary singer made famous in "Out of the Cradle Endlessly Rocking." Yet this poem also dramatizes collaboration, with one set of voices prompting another, bird song and human song, a single trill and the thousand responsive chords from a thousand different singers to follow.

Arguably the medium of print itself encouraged earlier editors to take a restricted view that often remained blind to the social aspects of textual production. It is easier, frankly, to exclude contributions made by book designers, copyeditors, typesetters, and others. Yet if we think longer and harder about Whitman's own career, the extent of his collaboration — almost entirely ignored by *The Collected Writings of Walt Whitman* — is striking. Whitman collaborated with typesetters, designers, and proofreaders, as he readily acknowledged, and also in his journalism, both as editor and writer; in his extensive though anonymous contributions to the early Whitman biographies by R. M. Bucke and John Burroughs; in heretofore uncataloged interviews (now being edited by Brett Barney); in his extensive conversations with Horace Traubel — a 5,000 page trove of information. In fact, his correspondence itself is fundamentally a collaborative undertaking involving (ordinarily) two-way engagements, though the strong authorial bias of the *Collected Writings* is clear in their featuring of just Whitman's outgoing correspondence.

**Project**

*Project* is a bigger, baggier term than edition and is far less specific in what it suggests about the type of work being undertaken. *Project* can describe everything from fixing a broken window on the back of a house to the Human Genome Project. In a literary context, *editions* and other results tend to emerge out of *projects*, but what constitutes the *project* is also the entirety of the undertaking: space, personnel, atmosphere, and the totality of all efforts. An *edition* might result...
from a project, without being the project, which includes all of the work conducted and records produced. The Whitman Archive, when regarded as a project, encompasses the compiled email discussion list that faithfully records the building of the Archive and the thinking that has gone into it. The documentation of a project, in our case, includes the behind-the-scenes Works-in-Progress page, with its assortment of information, including grant proposals, minutes from Whitman planning meetings over the years, a manuscript tracking database, an image warehouse, and project-related humor.

Project is not a favored word in every context. When I sent drafts of a "We the People" challenge grant application to NEH program officers, I was struck by how forcefully they discouraged me from using the word project, at least in the context of that competition. Their reasoning was that challenge grants were intended to fund permanent entities, unlike a project which they conceived of as having a finite temporal life. For me, "Whitman Project" and "Whitman Archive" were more or less interchangeable terms. I had to make a real effort to purge the document of all references to project. It was a neutral term to me: project was so natural as to be almost invisible in the drafts and certainly did not raise a red flag.

This story raises a larger issue: what happens when an undertaking becomes not just rhetorically but practically open-ended, when it has the good fortune or obligation to be an ongoing concern? We were successful with our challenge grant application, and we are now well along in building a $2 million permanent endowment for the Whitman Archive. Thanks to this remarkable turn of events, the Whitman Archive can now plan on an ongoing annual budget comparable to what one might expect annually from a major two- or three-year grant from a federal agency or foundation. And, remarkably, in this case, there is no end date to that support.

For the 2007 Digital Humanities conference at the University of Illinois, Urbana-Champaign, Matt Kirschenbaum coordinated a panel called "Done. Finished Projects in the Digital Humanities." He asked, "How do we decide when we're done? What does it mean to finish something? How does the 'open ended nature of the medium' (a phrase we all pay lip service to) jibe with the reality of funding, deadlines, and deliverables? What can we learn from finished projects, both successful and unsuccessful? For that matter, how do we define success and failure? Are 'we' the ones who ought to be defining it? If not, who?" These are good questions, and at the Whitman Archive we find ourselves concerned with them even as we face different considerations as well. What happens when work plans realistically could continue over generations? What is the best way to plan for that type of future? A theoretical possibility of digital scholarship — the indefinite expansibility — has become a lived reality in our case. We are only now absorbing the meaning of this grant, but one implication is that it provides us with the license, perhaps even the charge, to be as bold and ambitious as our talents and energies allow.

Database

How adequate is the term database for describing the type of large scale electronic projects we have been considering? Throughout this essay, I have used the Walt Whitman Archive as a testing point and illustrative example. To discuss the Whitman Archive in terms of database is especially timely now because PMLA recently featured an article about the Walt Whitman Archive by Ed Folsom, "Database as Genre: The Epic Transformation of Archives," and included a handful of responses (along with Folsom's response to the responses). The ensuing discussion made clear that people understand the term database in a variety of ways and attach different connotations to the word. These differences arise mainly from a distinction between 1) a strict definition of database — as a technical term in an electronic context database refers primarily to a collection of structured data that is managed by a database management system, most commonly based on a relational model; and 2) a looser use of database that employs the term on a more metaphorical level.

As the PMLA discussion of the Whitman Archive indicates, database can be a suggestive metaphor because it points to the re-configurable quality of our material (and that of similar sites). The term also conveys simultaneously "finished" and "unfinished" qualities; while a project can be logically thought of as "done" or "not yet done," we usually conceive of a database as usable as soon as it begins to exist, and we take as a given that the data will continue to proliferate, potentially indefinitely. The Whitman Archive resembles a database in that its content is discrete computer files that function atomistically: as functional units within a computing system each item is just as important as every other item.

If the Walt Whitman Archive resembles a database (without meeting the specifications of a technical or a literal definition), so, too, does Whitman's own process of composition. As Folsom notes, "Whitman formed entire lines as they would eventually appear in print, but then he treated each line like a separate data entry, a unit available to him for endless reordering, as if his lines of poetry were portable and interchangeable, could be shuffled and almost randomly scattered to create different but remarkably similar poems" [Folsom 2007, 1574-75]. At times, it almost seems as if Whitman were anticipating Raymond Queneau's Cent Milliard de Poèmes [One Hundred Thousand Billion Poems], a fascinating book in which the pages are cut horizontally so that each verse in each sonnet of the collection can be turned separately and all combinations of choices are poetically grammatical. (Queneau estimated that a reader would have to spend two hundred million years, working twenty-four hours a day, to read every combination.) Whitman's own cutting and pasting of lines, and his rearranging of poems to make other poems is not this extreme — nor is it as extreme as Samuel Beckett's experiments in Lessness — though there is some resemblance to both. Finally, though, what may appear random ordering in Whitman is best understood as restless experimentation, a
combinatory and recombinatory poetics, guided by Whitman’s recurrent drive to improve the effectiveness of his poems. Here, for those willing to use the term database metaphorically and to recognize non-electronic forms of databases, we can think of database as a key tool for Whitman himself: his storehouse of poetic lines, in both manuscript and print, was his working database for future compositions, one that he had always only partial access to because of the scattering of his documents but that nonetheless served as a means of composition.

If we turn to more literal uses of the word database and think about the *Whitman Archive*, we see that it is a complex composite structure that includes numerous databases and XML files. Folsom’s description of the *Whitman Archive* as “a huge database” is illuminating when taken metaphorically, though it is less helpful when taken literally, because the entirety of the *Whitman Archive* is not a single database any more than it is, as Jerome McGann asserts, merely XML files plus XSLT. In fact, the *Walt Whitman Archive* is comprised of numerous databases (some public and some not) along with many XML files including TEI, EAD, and XHTML files. McGann goes on to claim that the XML and XSLT work together to “allow users to access and—through an X-query-based search engine—manipulate The Walt Whitman Archive in the ways that Folsom rightly celebrates” [McGann 2007, 1588]. Ironically, though, in the course of denying the applicability of database as a term suitable to the *Whitman Archive*, McGann overlooks that our search engine is entirely dependent on translating the XML files into database form. At a more general level, McGann is perceptive in noting that any database represents an initial interpretation of the material. A database is not an undifferentiated sea of information out of which structure emerges. Argument is always there from the beginning in how those constructing a database choose to categorize information — the initial understanding of the materials governs how more fine-grained views will appear because of the way the objects of attention are shaped by divisions and subdivisions within the database. The process of database creation is not neutral, nor should it be.

**Archives and Digital Thematic Research Collections**

Having discussed *edition*, *project*, and *database* separately, I now turn to consider the final two terms together, *archive* and digital thematic research *collection*. In the past, an archive has referred to a collection of material objects rather than digital surrogates. This type of archive may be described in finding aids but its materials are rarely edited and annotated as a whole. In a digital environment, *archive* has gradually come to mean a purposeful collection of surrogates. As we know, meanings change over time, and *archive* in a digital context has come to suggest something that blends features of editing and archiving. To meld features of both — to have the care of treatment and annotation of an edition and the inclusiveness of an archive — is one of the tendencies of recent work in electronic editing. One such project, the *William Blake Archive*, was awarded a prize from the Modern Language Association recently as a distinguished scholarly edition.

Digital archives are often notable for their depth and breadth of coverage of whatever the stated thematic interest is. Such scope has not been common in editing. Indeed it is possible to see a tension in the very term *collected edition* because collecting and winnowing are two very different activities. Thomas Wentworth Higginson, in a review of the *Complete Writings of Walt Whitman*, might have been commenting on the *Whitman Archive* when he wrote, “[T]he present editors do not shrink from inserting not only the details of every change, but even the unprinted variations which have hitherto existed in manuscript only” [Higginson 1903, 400]. Of course, the more inclusive an edition becomes the more it may be dominated by the surviving “discarded” writings, especially for writers who kept many documents [Folsom 1982, 374]. Some feel that we do violence to the wishes of writers when we make their second-rate material available to the public, while others celebrate what they believe is made possible by inclusive editions: a new, deepened, and enriched sense of the artist’s process of composition, preoccupations, and achievements. Ultimately, the whole question of what is in keeping with the wishes of a writer is beside the point. We do not edit for writers themselves but for our own purposes as scholars and readers.

Peter Shillingsburg expresses skepticism about the advantage of the archival approach:

> The computer makes possible, we are told, the juxtaposition of all the relevant texts in their linguistic and bibliographic variant forms. Thus a library of electronic texts, linked to explanations and parallels and histories, becomes accessible to a richly endowed posterity. To the extent that such archives contain accurate transcriptions, high resolution reproductions, precise and reliable guides to the provenance and significance of their contents, and the extent to which they are comprehensive, to that extent they are “definitive” — until the next generation of critics and scholars with new interests notices some other aspect of texts that scholarly editors of the past (by then that will be us) took for granted and ignored. But already, information overload has set in. The comprehensiveness of the electronic archive threatens to create a salt, estranging sea of information, separating the archive user from insights into the critical significance of textual histories. [Shillingsburg 2006, 165]

Shillingsburg focuses on the limits of a form still being developed as opposed to the potential of that form. Nothing in the archive form intrinsically requires it to be “estranging” or alienating, of course. An electronic archive can be as welcoming as fresh water and as rewarding as the wit of its creators can make it. Having a lot of information is not
inherently more estranging than having less information. Nothing guarantees the effectiveness of selective treatment accompanied by "textual histories," and nothing guarantees effectiveness of more comprehensive treatment accompanied by textual histories. In each case, everything depends on the quality of the editorial work. Digital and print scholarship are equally embedded in history, and both share a vulnerability to aging.

Another term that is more or less synonymous with electronic archive is digital thematic research collection. Some prefer this term because it may avoid some of the misleading connotations of archive — ordinarily people assume that materials in a traditional print-based archive are unedited. Carole Palmer writes about thematic research collections, collections of all kinds can be open-ended, in that they have the potential to grow and change depending on commitment of resources from collectors. Most thematic collections are not static. Scholars add to and improve the content, and work on any given collection could continue over generations. Moreover, individual items in a collection can also evolve because of the inherent flexibility (and vulnerability) of "born digital" and transcribed documents. The dynamic nature of collections raises critical questions about how they will be maintained and preserved as they evolve over time. [Palmer 2004, 351]

Archive is a self-designated term, one adopted by the creators of resources. In contrast, digital thematic research collection is a term used by people describing the work created.

Thematic research collection may be the most accurate term for what many of us are attempting, but it has not gained currency because it is neither pithy nor memorable. Carole L. Palmer notes that a digital thematic research collection is the closest thing to the laboratory that we have in the humanities — the place where necessary research materials are amassed. I have argued elsewhere that in a "digital context, the 'edition' is only a piece of the 'archive', and, in contrast to print, 'editions', 'resources', and 'tools' can be interdependent rather than independent" [Price 2007, 435].

Does collecting — the emphasis in Palmer's description — qualify as research, as a scholarly genre? A digital thematic research collection possesses the virtues of a traditional scholarly edition while containing much more. We may nonetheless wonder about how helpful the term digital thematic research collection is to the uninitiated. Nothing in the term indicates editorial rigor and nothing points to the value added by scholarly introductions, annotations, and textual histories. The only thing that seems to separate it from a mass digitization project is the "thematic" element. However, one can imagine a mass digitization project that is thematic and that lacks editorial supervision and intervention in the reader's experience of the text. Can we find a better term that indicates this difference? Does digital thematic research collection communicate its meaning adequately?

If literary scholars who are assembling electronic texts are becoming fundamentally or solely "literary-encoders" and "literary-librarians," then, despite my own recognition of the inseparability of interpretation and encoding, I fear for the standing of their work when judged by faculty in humanities departments (Schreibman, as quoted in [Palmer 2004, 352]). Without care and forceful practical examples and theoretical essays, the same prejudices and misunderstanding that drove editing and bibliography from the center to the periphery of literary studies will continue to prevail. We also need descriptions of digital thematic research collections that highlight the editorial work and other types of scholarly value that are added to the raw materials populating the collection. In many circles, editing — whether it is print-based or electronic — is regarded as pre-critical work. Some editorially related tasks are fairly routine and do not require scholarly expertise (the same is true of critical work as well). And yet others clearly do, and we need to find ways to clarify how historical knowledge, theoretical sophistication, and analytical strengths are necessary to the creation of a sound text or texts and accompanying scholarly apparatus in a successful edition.

Some components of a digital thematic research collection or archive may stretch ordinary understandings of edition. Many thematic research collections or archives aim toward the ideal of being all-inclusive resources for the study of given topics. A good thematic research collection might begin with an edition conceived in inclusive terms. Digital thematic research collections go far beyond traditional editions in their presentation of many types of materials. They are often even more "organic" than print editions (despite their technological aspects) — that is, they grow, evolve over time, based very much on immediate circumstances. For the Walt Whitman Archive, new work on the Civil War is now underway because an expert on Abraham Lincoln at the University of Nebraska-Lincoln, Kenneth J. Winkle, and I perceived a scholarly need and are interested in collaborating on this undertaking. New work on translation — I will say more about both of these new endeavors later — developed because Matt Cohen, already associated with the Whitman Archive, was interested. Being published online but being simultaneously a work-in-progress allows for a flexibility in the Whitman Archive that print editions could never have. New scholars with new ideas may emerge at any time, creating new and unexpected additions to our work.

I mentioned earlier that the theoretical possibilities of digital scholarship might oblige us to boldness — the present moment, when electronic scholarship is still nascent and the boundaries are still capable of being moved, provides a mandate to innovate and expand possibilities. Ideally, a digital thematic research collection would also allow for the study of cultural contexts. In the case of Whitman, we might want to study him as a city poet. He once said that Leaves of Grass "arose out of my life in Brooklyn and New York from 1838 to 1853, absorbing a million people, for fifteen years
with an intimacy, an eagerness, an abandon, probably never equaled" (quoted in Reynolds 1995, 83). A life-long city-dweller, his work also emerged out of New Orleans, Washington DC, and Philadelphia/Camden, New Jersey. We would like for the site to enable and to promote interpretations of place-based writing that were not possible before. It would be useful to be able to study all of these areas with dynamic maps containing detail down to the block level. Period maps exist for Washington, DC, New York, Brooklyn, Philadelphia, and New Orleans. New discoveries will emerge once we can ask different questions because of having a great deal more information from census records, maps, health records, police reports, possibly even information on sexual subcultures, and so on.

I have recently begun work on a digital undertaking that may or may not become part of the *Whitman Archive*. Whether the project ultimately is folded into the *Archive* or remains a separate, stand-alone collection, it certainly grew out of my work on the *Archive*. We might think about it as budding off of an existing digital thematic research collection and taking on a life of its own. The project "Civil War Washington: Studies in Transformation" draws on the methods of many fields — literary studies, history, geography, computer-aided mapping — to create an experimental digital resource. The President and the poet both experienced the War from vantage points in the nation's capital, Lincoln striving to reunite the divided nation and Whitman caring for tens of thousands of wounded soldiers. Their activities and perspectives chronicle the War and provide insights into the large and complex forces that transformed Washington from a sleepy Southern town to the symbolic center of the Union and nation.

We are gathering uncollected factual data about an urban space that served as the center both of the Union's War effort and of a divided nation, where hospitals arose overnight, wounded men moved in and out, "contraband camps" of fugitive slaves developed, and temporary shelters were erected to house the city's swelling population, which tripled during the four years of the War. Washington was a noisy city during these years: the noise in the city was of construction as work on the Capitol continued; the noise just outside the city was of destruction as the Confederate army worked to tear it down. Even as bridges were defended and a ring of forts made this space the most heavily defended city on earth, Washington fostered vibrant life.

"Civil War Washington: Studies in Transformation" will situate Lincoln and Whitman in the midst of a rich field of geo-spatial and temporal data. At the heart of the project will be richly layered, interactive maps plotting both geographic and temporal data that clarify the transformation of Washington, DC. The maps and underlying databases will make it possible to analyze change over time as structures grew and the population swelled and developed a new ethnic and racial mix. We will make possible multifaceted and dynamic studies of Lincoln's and Whitman's activities during the War years, based on textual and statistical evidence and using the power of maps and graphs to illustrate historical change. Lincoln's and Whitman's routes can be plotted on a daily and sometimes hourly basis. We believe that by providing a rich backdrop of census, health, and hospital records; theater schedules; horsecar routes; and other factual data, we will make possible a better understanding of Lincoln's and Whitman's lives and their roles in the transformation of the nation and its capital.

Another extension of the *Whitman Archive* now being undertaken serves to expand trans-linguistic, cross-cultural understandings. Whitman scholarship offers rich opportunities because *Leaves of Grass* has been translated into every major language. One of the *Archive*’s objectives is to present editions of Whitman's work key to literary, cultural, and historical study of the poet and his work's effects. Thus Matt Cohen has taken the lead in tackling a digital edition of the first extensive translation of his work into Spanish. Álvaro Armando Vasseur's 1912 selection from *Leaves of Grass* is the work of a Uruguyan poet who translated Whitman not directly from English but via an earlier Italian translation. This fascinating text tells us a lot about the circulation of culture. Making a version of *Leaves* available to the Hispanophone world seems fitting given current trends in U.S. demographics and in light of the many calls to internationalize American studies.

We supplement the translation with a critical introduction and a sample back-translation into English in order to give those unable to read Spanish an opportunity to see how the text was altered in the process of translation. For example, consider the following lines as given by Whitman:

The disdain and calmness of martyrs,
The mother of old, condemn'd for a witch, burnt with dry wood, her children gazing on,
The hounded slave that flags in the race, leans by the fence, blowing, cover'd with sweat,

And here is how Vasseur rendered these lines as revealed in a literal back translation:

The mother of old condemned as a witch and burned over dry firewood, before her children’s eyes,
The slave, persecuted like an imprisoned woman, who falls mid-flight, all atremble and sweating blood.

Vasseur's direct comparison of the slave to a woman presumably is based on their common lack of power, but it also creates some cross-gendered possibilities that turn the passage in new ways. Whitman had distinct units — separate lines — for the witch and the hounded slave. An association could be made between them because of their juxtaposition, but that association is hardly insisted on in the English original. Vasseur turns the suggestion of a link into an unmistakable link. Now racial slavery has become associated with the irrationality of the inquisition and serves to remind the reader of the widespread support of slavery by the church in the U.S. (and in South America). While this
reading is only barely available in Whitman's original, in Vasseur's translation it appears on the surface. This passage clarifies that translating a text is interpreting it in another language. To ignore such interpretations is to ignore an enormous part of Whitman's reception in the world.

We have either in progress or the planning stages work on Whitman and other languages (German, Russian, Ukrainian, Portuguese, and Chinese). This will begin to better place him in a world context rather than situating him solely in Anglophone culture. The work will provide valuable texts to further Whitman studies and through associated commentary reflect the social, historical, and linguistic milieus of the nations in which the translations were done, thereby once again stretching the bounds of what a digital thematic research collection originally envisioned within much narrower parameters can do. These possibilities, the ever-emerging questions and new directions, go far beyond the ordinary edition in the pre-digital age.

As I have indicated, we do not have an adequate term to describe the digital scholarly work now underway in numerous projects. What is it that we want our descriptive word to capture: is it the physical thing? Digital sites, contrary to popular (and sometimes scholarly) opinion, are physical things after all — they take up space, can be created and destroyed, and so on. Is it the nature of the content? If so, we need a word that suggests what can be an infinitely extensible resource. Or should we emphasize, primarily, the way we make the thing, the collective that has come together in order to do work on a new scale in humanistic study?

Importantly, we should not strive to fit our work to one or another existing term but instead expect that, in time, terms will alter in meaning — or new ones will come into existence — so as to convey the characteristics of a new type of scholarship. I strongly agree with Peter Shillingsburg that a new term is needed, though I am not enthusiastic about his proposed term: knowledge site. (So many places and institutions could justifiably be called knowledge sites that the term seems unlikely to become identified with a particular genre of electronic scholarship.) I propose instead a not-immediately-intuitive but perhaps ultimately more promising alternative: arsenal.[10] The online etymological dictionary helps explain the appeal of the term:

arsenal

1506, "dockyard," from It. arzenale, from Ar. dar as-sina'ah "house of manufacture, workshop," from sina'ah "art, craft, skill," from sana'a "he made." Applied by the Venetians to a large wharf in their city, which was the earliest meaning in Eng. Sense of public place for making or storing weapons and ammunition is from 1579.

I like the emphasis on workshop since these projects are so often simultaneously products and in process. I also like the stress on craft and skill, a reminder that editing is not copyist work. The "public place for making" suits current aspects of the genre under discussion and will no doubt characterize it even more in this age of social networking. The dockyard connotations of arsenal are helpful in suggesting a kind of inclusiveness about all the vessels, sloops, ketches, and yawls that can hook up to it. (The wharf and dockyard are places of multilingual exchange.) The obvious objection to the term arsenal is that it seems militaristic in current usage. Yet we should recall that magazine once primarily meant a storehouse for weapons and ammunition. If the primary meaning of magazine can shift from being a storehouse of weapons to a storehouse of mixed content for periodical publication, who knows what could happen with arsenal[11]

We are, for better or worse, always entangled with force and power: the Internet itself has its origins in the military. Perhaps one step toward turning swords into plowshares is to seize a word like arsenal and make it our own. Can we imagine a world in which what is emphasized is not the created thing so much as the group of people who are now joined together for a common purpose?

Notes


[2] After the publication of the 1881-1882 Leaves of Grass, Whitman remarked, "All this is not only my obligation to Henry Clark, but in some sort to all proof-readers everywhere, as sort of a tribute to a class of men, seldom mentioned, but to whom all the hundreds of writers, and all the millions of readers, are unspeakably indebted. More than one literary reputation, if not made is certainly saved by no less a person than a good proof-reader. The public that sees these neat and consecutive, fair-printed books on the centre-tables, little knows the mass of chaos, bad spelling and grammar, frightful (corrected) excesses or balks, and frequent masses of illegibility and tautology of which they have been extricated" [Whitman 1978, 256]

[3] Issues involving long-term preservation come to mind, of course. A simple curation is not viable. That is, we cannot hand over to a library the present-day Whitman Archive and expect people fifty years from now to find its interface and technical underpinnings particularly easy to use. This is in sharp contrast to a book published fifty years ago and...
deposited in a library. For digital scholarship, we cannot foresee how maintenance, updates, and migration will work in the future.

[C'est somme toute une sorte de machine à fabriquer des poèmes, mais en nombre limité; il est vrai que ce nombre, quoique limité, fournit de la lecture pour près de deux cents millions d'années (en lisant vingt-quatre heures sure vingt-quatre)] [Queneau 1961, n. p.]

Beckett's short story Lessness was first published in French as Sans. In his enigmatic story Beckett experimented with random ordering of sentences in the making of fiction.

In a "Reply" to those who commented on his essay, Folsom observed that the Whitman Archive is in fact "several databases."

It should be noted that my view of archive differs here from that of some commentators. Peter Shillingsburg, for example, remarks that "the level of critical intervention is miniscule in the electronic archive" [Shillingsburg 2006, 156]

In a series of talks in the 1990s John Unsworth and Daniel Pitti began applying the term thematic research collection to the type of scholarship under discussion in this paper.

Recent work in archival theory by Heather MacNeill, Elizabeth Yakel, and Michelle Light and Tom Hyry, emphasizes the non-neutral nature of archives themselves and urges the adoption of language such as "archival representation" to highlight the mediating role archivists as they order, interpret, and develop information architectures within socially constructed practice.

Henry Wadsworth Longfellow famously explored the possibility transforming an arsenal into pipe organs of love. See his poem "The Arsenal at Springfield."

I am less concerned that arsenal catches on than I am that we recognize the fresh features of new work underway and that we are self-conscious about what we want any new term to convey.

Works Cited


Editing Historical Documents

A Handbook of Practice

Michael E. Stevens
Steven B. Burg

Published in cooperation with the American Association for State and Local History, the Association for Documentary Editing, and the State Historical Society of Wisconsin.
Chapter One
What, Why, How, and for Whom?

Why Edit Historical Documents?
1.1 People edit and publish historical documents because they believe these materials have enough significance to merit the time, energy, and money needed to make them accessible to a wider audience. Whether intended as research tools, study aids, or simply pleasurable reading, the basic mission of historical editions is to provide easy access to the unique information contained in original documents. Editors reproduce documents through the creation of facsimiles or transcriptions, make them understandable by providing annotation, facilitate access to them through devices such as indexes, and publish these documents via microforms, computer networks, CD-ROMs, or the printed page. Historical editors do not simply reproduce texts; they also provide readers with the information needed to understand the content of historical documents.

Scope
1.2 However small or grand a project, from a single document in a historical journal to a multivolume book edition or a multireel microform edition, documentary editors begin by defining the scope of the project. Whatever the topic, the editor should articulate a clear and well-defined statement of scope that includes a definition of who or what is being documented, the dates under consideration, and the reasons for undertaking the project. By clearly specifying a project’s scope, editors use resources more efficiently and bring coherence to the work of collecting, editing, and publishing documents.
1.3 A project may use an individual as its subject. For example, the editors of the Papers of Woodrow Wilson document the thought, character, and career of Wilson in their edition:

The editors hope to publish a comprehensive edition that will include all important letters, articles, speeches, interviews, and public papers by Woodrow Wilson. These volumes, when complete, should make available to readers all the materials essential to understanding Wilson’s personality, his intellectual, religious, and political development, and his careers as educator, writer, orator, and statesman. The editors also hope that these volumes will be useful to scholars and others in various fields of history between the 1870’s and the 1920’s.

Many editors define the project’s scope more narrowly. For example, the editors of the Papers of George Washington: Revolutionary War Series set chronological and topical guidelines for winnowing the huge number of documents associated with the Continental Army under Washington’s command:

Fortunately the task immediately at hand is not to collect and print every surviving document relating to the Continental army. Rather, it is to identify and print those surviving documents written by and to George Washington between 1775 and 1783 that deal directly with Washington as commander in chief of the Continental army or simply as a man.

An editor may use aspects of individuals’ lives, such as their professional careers, to illustrate a broader historical topic. The editors of the Papers of Daniel Webster: Legal Papers used Webster’s legal career to explore the nineteenth-century legal profession:

It is the purpose of these two volumes to provide a sense of the texture of the practice of law in early nineteenth-century America. Webster presents a rare occasion for shaping that portrait because his practice evolved from a struggling rural environment to an elite urban setting. As Webster moved, he touched various strands and aspects of the profession. In order to provide a view of the actual practice, in part to fill what we have perceived as a gap in the historical literature for this period, we have included several variables that constitute practice, such as clients, legal doctrine, statutes, social environment, and economic conditions. In the history and the legal profession, both are a product of every individual and every era.

The editors of the Papers of Frederick Douglass and the Papers of the Reconstruction Era have used the opportunity to examine subjects of more contemporary interest. The Douglass papers, neither a found in American life in the early twentieth century, deal with African Americans as a subject for African American studies. The reconstructed history reveals the role of black Americans as a leading black American political leader.

The Washington papers, on the other hand, deal with this controversy with a broad interest in American culture in all its manifestations. Even in the early twentieth century, not adopted by Washington, but because of social and political needs. Because he was a black American, the Washington papers present an opportunity to explore the “American Dream” that there is a black American hero. Because he was a black American, the Washington papers present an opportunity to explore the “American Dream”

Fig. 1.3a Link, Wilson, I:xiv

Fig. 1.3b Chase, Washington: Revolutionary, 1:xxx
ments. For example, the editors sought, character, and career

The editors of the Booker T. Washington Papers sought to document both the man and an era of United States race relations:

Unlike the subjects of most earlier enterprises in historical editing, Washington was neither a founding father nor a national political leader. Yet his importance in American history is firmly established, and his significance as a subject for an editing project stands securely on the wealth of social history revealed in his papers as well as the fact that Washington was a leading black American.

The Washington Papers project does not seek to build a monument to this controversial figure or his social philosophy but to reveal black culture in all its complexity and rich human interest and the story of American race relations, black and white, in the late nineteenth and early twentieth centuries. This critical approach to the central figure is not adopted because he was black or because he was Booker T. Washington, but because of the editors’ belief that history should serve the cause of social criticism instead of exalting a people or an individual hero. Because Washington was a middleman between the white and black America of his day, his papers also shed light on interracial relations in the “age of accommodation.” But the reader should be warned that there is an imbalance in the Washington Papers, for with the exception of occasional mention of the crop-lien system or lynching or discrimination, Washington rarely admitted that there were deep racial problems in America. He generally took the hopeful view on the ground that it was more constructive, an approach that soothed white southerners and loosened the purse strings of philanthropists. By including incoming correspondence the editors hope to broaden the focus of the papers beyond Washington’s restricted perspective. This wealth of sources transcends Washington’s private and public life, and the editors seek in the selection of documents to portray not only Washington’s life but as far as possible his age as well.

Fig. 1.3d Harlan, Washington, 2000
1.4 The papers of two or more individuals or correspondents may be edited together if their personal interaction or the nature of existing documents makes such a compilation logical. The editors of the microfilm edition of the Papers of Elizabeth Cady Stanton and Susan B. Anthony chose to publish the two women’s papers together because their work was so interconnected:

The edition presents the papers of both women as a single, interfiled collection. Their good friend Theodore Tilton once described how difficult it could be to separate the individuals from the remarkable partnership between Stanton and Anthony.

Elizabeth Cady Stanton, or, as she is sometimes called, Susan B. Anthony, is a celebrated lady with snow-white and auburn hair, plump and slender figure, Grecian and Roman nose, and lives simultaneously in two houses—one at Tenafly, N.J., and the other at Rochester, N.Y. Indeed, Mrs. Stanton, or, to call her by her maiden name, Miss Anthony, is a ubiquitous personage, and not only attends all the women’s meetings wherever they are held, but also has been known on certain occasions to be writing the resolutions, and at the self-same moment to be delivering the speech, in their support. It has been sometimes suspected that Mrs. Stanton and Miss Anthony are two distinct persons, united by a cartilage like the Siamese twins, but in the absence of any medical or other scientific proof of this hypothesis, I remain of the opinion that, like Liberty and Union, they are “one and inseparable.”

Two distinct personalities emerge from this edition, but Tilton put his finger on a problem for the researcher or editor who tries to treat one woman independently of the other. Through half a century apiece of activism, they worked within the same realms of public life toward closely interconnected goals. To separate their stories is, in a sense, to distort how they themselves regarded the relationships between different reforms.

By combining their papers in a single edition, the papers of each woman are enriched by association, by being available within a broader context of related reform. In terms of editorial work, the search for documents for each woman independently would lead to endless redundancy. The richest lodes of documentation produced the papers of both women. When they worked in tandem, the problem is more practical. Whose papers are those that were written together over the dining-room table at the Stanton household? When Stanton accepted invitations to speak at events she had no intention of attending and Anthony faithfully delivered Stanton’s address in her stead, whose papers should document the meeting?

Had either woman received a large corpus of papers about her private life, the editorial decisions might have differed. That which distinguished the two women would have competed with that which they shared, rendering use of the edition somewhat awkward. Because most of the papers in this edition treat political activity and issues arising in public life, that outer world of agitation dominates the narrative. Two characters and all their co-workers move across a common stage, growing older, adapting their schedules and their domestic habits, but fighting to the last for acceptance of woman’s self-sovereignty.

Fig. 1.4a Holland, Stanton-Anthony, 2

The editors of the Adams Family Correspondence combined all personal correspondence among family members, an arrangement that highlights their interaction but also separates an individual’s private and public papers.

1.5 Since the significance of a manuscript, and so detailed in the case of a correspondence, is lost if it is flattened out as an early stage, the microfilm editors compiled it as a unit by a characteristic account of its advantages and limitations, and replies to questions of a rather than a formal character. The woman’s general education, as it enables the writer to express the interests of being objective, while still being significantly influenced by correspondence.

The editors of the League and Pan-African Garvey organization's manuscripts decided to focus on Garvey's actions and statements.

The editors of the Legal League and Pan-African Garvey organization's manuscripts decided to focus on Garvey's actions and statements.


or correspondents may be
the nature of existing docu-
tors of the microfilm edition
B. Anthony chose to publish
rk was so interconnected:
ingle, interfiled collection. Their
could be to separate the individ-
Anthony.

n B. Anthony, is a celebrated lady
Sorea and Roman nose, and lives
other at Rochester, N.Y. Indeed,
ny, is a ubiquitous personage, and
held, but also has been known on
same moment to be delivering the
Stanton and Miss Anthony are
twins, but in the absence of any
f the opinion that, like Liberty and

Hilton put his finger on a problem
pendently of the other. Through
reals of public life toward
sense, to distort how they them-
of each woman are enriched by
ated reform. In terms of edito-
lessly would lead to endless re-
papers of both women. When
se papers are those that were
in household? When Stanton
attending and Anthony faith-
ould document the meeting?
her private life, the editorial
two women would have com-
much awkward. Because
ues arising in public life, that
cers and all their co-workers
chedules and their domestic
overignty.

in Holland, Stanton–Anthony, 2

bined all personal corre-
that highlights their inter-
public papers:

Since the record of these constants and these changes is so extended
and so detailed, and since the letters that the Adamses wrote each
other have been so carefully preserved, the editor in chief decided at
an early stage in his planning to present the family correspondence
as a unit by itself. A disadvantage of this arrangement is that it adds
another chronology to a pattern already complicated enough. But its
advantages seem much greater. For one thing, it will present the letters
and replies of the Adams statesmen to each other in the same sequence,
rather than widely separated in the volumes devoted to each states-
man’s general correspondence and other writings. For another thing,
it enables the women of the family, both those born into it and those
who married into it, to take their places beside the Adams men instead
of being obscured by them. Finally, and, as the editors believe, most
significantly, certain kinds of evidence furnished by the family cor-
respondence are not too abundant in print elsewhere and are better
recognized today than formerly by historians as peculiarly precious.

Fig. 1.4b Butterfield, Adams Family, I.xxv

1.5 The significance of some individuals is so closely tied to a particular
institution or organization that it may make sense to edit the documents of the
individual and the organization together. The editors of the Marcus Garvey
Universal Negro Improvement Association Papers created their project with the
intent of studying both the man and the organization:

The Marcus Garvey and Universal Negro Improvement Association Papers
are based on a comprehensive survey of all the presently available historical
manuscripts and records pertaining to the life and work of Marcus Mosiah
Garvey as well as to the popular worldwide organization that he founded and
led from its inception in 1914 until his death in 1940. As the record of the only
organized international mass movement of persons of African descent and as
the history of a mass social phenomenon, these Papers go beyond a preoccu-
pation with the fortunes of a single, even if major, historical figure. Their
focus is as much on the participation of members and supporters as on
Garvey’s activities as the political leader of the movement.

Fig. 1.5a Hill, Garvey, I.xxi

The editors of the microfilm edition of the Papers of the Women’s Trade
Union League and Principal Leaders found it necessary to collect the papers of the
organization’s leaders to fully document the organization’s activities:

cooperation of seven different repositories, it has assembled in one edition the principal
manuscript collections pertaining to the League.
The collections vary in size and in scope. Two represent the surviving files of the National WTUL. When the League disbanded in 1950, the main portion of its files went to the Library of Congress and a smaller segment to the Schlesinger Library. The edition also includes the only other substantial body of official records still extant, those of the New York WTUL.

Fortunately for the history of the WTUL, these institutional records could be supplemented and fleshed out by substantial collections of personal papers. Foremost among them is the large and comprehensive collection of the papers of Margaret Dreier Robins, the guiding force of the League during its formative and peak years. Of considerable depth and richness also are the papers of Leonora O’Reilly, which illuminate both the League’s New York activities and O’Reilly’s own special intensity and inspiration. The Rose Schneiderman Papers, though relatively sparse, further document the New York scene; the Agnes Nestor Papers do the same for League activities in Chicago. Mary Anderson’s papers, although confined to her years as head of the federal Women’s Bureau, thus omitting her earlier experience as an organizer for the Chicago and National WTUL, record her continuing participation in the League and some of its inner concerns. Smaller collections give glimpses into the history of the Boston and Chicago Leagues and into the life of Mary Kenney O’Sullivan, co-founder of the National WTUL.

1.6 An organization or institution may be the subject of an editing project. The editors of the Documentary History of the Supreme Court of the United States deal with the court’s first ten years:

This multi-volume documentary history will provide scholars and lawyers with the first accurate record of all cases heard by the Supreme Court between 1790 and 1800. It will also present an ample selection of contemporary comment about the justices and their duties, the business of the Court, and the function of the Court in the constitutional framework. A compilation of official records, private papers, and other primary sources, this series will bring together and make readily available hitherto unpublished materials. They will document the Court’s important role in creating administrative procedures for the American judicial system and establishing the legal precedents that enabled the new government to prosper.

1.7 Editors may collect documents illustrating a particular topic, theme, or event. A topic may be concrete or abstract as long as it is defined well enough to guide the collecting and editing of documents. For example, the editors of Freedom: A Documentary History of Emancipation delineated specific topics that they would and would not include in their work on the transition from slavery to freedom in the southern United States:

1.8 Each edition has focused on a particular group of slaves and free blacks and on the economic, social, and political concerns of the period. The editors and their makers inform us that the volumes published by the Bureau of Census were used as a base for the narrative, but not for the sources. The documents, nonetheless, document the reign of slavery and the transition to freedom. The editors have reconstructed the records of the Bureau of Census and the correspondence that surrounded it.

Above all, the editors have sought to show the process of the transition from slavery to freedom.
The editors found it imperative from the outset to be selective. They have focused their attention upon the wartime and postwar experiences of slaves and ex-slaves, but have also sought to illuminate the social, economic, and political setting of the emancipation process. The formation of federal policy, for example, is not central to the project’s concerns, except insofar as the preconceptions and actions of policy makers influenced the shape that freedom assumed. Therefore, the volumes published by the Freedmen and Southern Society Project will not undertake a history of the Freedmen’s Bureau, the U.S. Army, the Bureau of Colored Troops, or any other governmental agency; nonetheless, documents about the operations of these agencies will be prominent when they describe activities of freedpeople and shed light upon the context in which former slaves struggled to construct their own lives. Throughout the selection process, the editors have labored to reconstruct the history of the freedpeople rather than the institutions that surrounded them.

Above all, the editors have sought to delineate the central elements of the process by which men and women moved from the utter dependence of slaveholders to freedom but never fully received, to the independence freedpeople desired but seldom attained.

Form of Publication

Each editor decides the form of publication most appropriate for the selected materials, including both the medium in which the documents will be published and whether the documents will be presented as facsimiles, as transcriptions, or in both forms.

Edited documents can be distributed through several different media. Printing documents in a historical magazine or another type of periodical makes them accessible to a wide audience, although the number of publications willing to print historical documents is limited, and they must of necessity limit the quantity of documentary materials. Self-publishing one or more documents in a pamphlet is a fairly easy and affordable way to reach an audience, especially for an institution like a library or a historical society that has a gift shop.

Editors of larger projects may want to publish the results of their work in a book edition, be it a single volume or multivolume set. Book editions are easy to read, compact, and portable and may be read without expensive viewers or computer hardware. When produced on acid-free paper they can have a shelf life of hundreds of years. Yet books are expensive to produce and require a publishing house willing to take on the costs of printing and distributing the work. The potential market for the edition or the availability of a subscription or endowment will affect both the number of volumes and the size of the press run.
Microform editions provide a compact and less costly means of reproducing and preserving large numbers of documents, but viewing microforms can be hard on the eyes, can produce images that are difficult to read and copy, and furthermore, requires the use of machines. Editors may also consider using computer and scanning technology to put their documents onto computer disks, CD-ROMs, or the World Wide Web (the Internet). The immediate appeal of electronic media is their ability to store vast amounts of information that can be disseminated at a low cost. However, electronic media have an uncertain shelf life, with carefully stored disks and electronic tapes holding data for ten or fewer years and CD-ROMs predicted to last from fifty to 100 years. Furthermore, the fast rate of change in computer technology threatens all electronic media with rapid obsolescence. Purchasers of electronic editions must adapt them to new systems, preserve the machines and software needed to use the existing format, or lose access to the information contained in the editions.

1.10 Whether publishing documents on the printed page, on microform, or via electronic media, editors must decide whether documents should be presented as facsimiles, as transcribed texts, or in both forms. While many microform and electronic editions present only facsimiles—thus saving the time, labor, and expense of transcribing and annotating documents—these editions can be hard to read, difficult to understand, and impossible to search unless the editors mark the texts or provide thorough indexes. Each user, in essence, approaches the documents afresh, without any of the expertise of the editor. In editing documents, the greatest service editors can provide readers is a clean, readable, transcribed text, followed closely by a convenient method for gaining access to the information in the documents and explanatory annotation.

Audience

1.11 Editors should identify their intended readers and consider the knowledge and expectations they are likely to bring to the work. Accounting for the specific needs of different kinds of readers affects decisions editors make about the selection, annotation, and presentation of documents.

1.12 Editions aimed at scholars and professional communities will often be tailored toward a knowledgeable audience well versed in the language and background of the subject. Scholarly and professional audiences use documentary editions as research tools and therefore require full and authenticated texts, notes that provide information beyond common historical knowledge, and exhaustive indexes. For scholars, the narrative flow and readability of an edition are secondary to the information contained in the documents.
pact and less costly means of reproducing documents, but viewing microforms negates that are difficult to read and of machines. Editors may also conclude to put their documents onto Wide Web (the Internet). The ability to store vast amounts of low cost. However, electronic media stored disks and electronic tapes ROMs predicted to last from fifty of change in computer technology sciences. Purchasers of electronic reserve the machines and software access to the information contained the printed page, on microform, or whether documents should be pre-n both forms. While many micro-facsimiles—thus saving the time, tating documents—these editions and impossible to search unless gh indexes. Each user, in essence, y of the expertise of the editor. In rs can provide readers is a clean, a convenient method for gaining id explanatory annotation.

Readers and consider the knowl- to the work. Accounting for the cts decisions editors make about documents. nal communities will often be ll versed in the language and sional audiences use document require full and authenticated common historical knowledge, ative flow and readability of an ed in the documents.

1.13 Editions aimed at audiences of students and general readers are characterized by brevity, readability, and a preference for using narrative rather than academic forms of annotation. Editors of documents aimed at students and general readers strive to make their works readable without sacrificing the accuracy of their content. In place of footnotes, an editor may, for example, use introductions and headnotes to provide a basic historical context and to define unfamiliar words or phrases. Editors also frequently abstract or excerpt documents to maintain the flow of the narrative and to keep readers interested.

1.14 Editions intended for different audiences present documents on the page in different ways. In Benjamin Franklin: Writings, an edition intended for general readers, Franklin's autobiography is presented without editorial markings (top), while the Autobiography of Benjamin Franklin: A Genetic Text created a scholarly edition that used editorial symbols extensively to mark textual details (bottom):

My Time for these Exercises & for Reading, was at Night after Work, or before Work began in the Morning; or on Sundays, when I contrived to be in the Printing House alone, evading as much as I could the common Attendance on publick Worship, which my Father used to exact of me when I was under his Care:—And which indeed I still thought a Duty; tho' I could not, as it seemed to me, afford the Time to practise it.

Fig. 1.14a(1) Franklin, Writings, 1320

My Time for these Exercises was <after Work at Night> at Night after Work, or before <it> Work began in the Morning; or on Sundays, when I contrived to be in the Printing House alone, evading as much as I could the <usual> Attendance on publick Worship, which my Father used to exact of me when I was under his Care:—And which indeed I still thought a Duty; tho' I could not, as it seemed to me, afford the Time to practise it.

Fig. 1.14a(2) Lemay, Franklin: Genetic Text, 14

The editors of Freedom: A Documentary History of Emancipation (top) republished many of the same documents in a volume they created for students and general readers entitled Free at Last: A Documentary History of Slavery, Freedom, and the Civil War (bottom). The editors adapted their documents for general
73: Officers of a Black Union Association to the Commander of the Department of the Gulf

New Orleans Dec 22th 1862

In obedience To Th High chift an Command of Th Head quarters Department of Th Gulf Maggor Gen N P Bank We Th members of Th union association Desir Th & Respectfully ask of you Th priviliges of Salabrating Th first Day of January th 1863 by a Large procension on that Day & We Wish to pass th Head quarters of th union officers High in a authority that is if it Suit your approbation & We also Wish to Give a Grand union Dinner on th Second Day of January that is if it so pleas you & th profit of th Dinner Will Go To th poor people in th Camp th Colour Woman & childrenen Your Most Hombre obedient servant

J M Marshall th president of th union association
Henry clay th Superintender of th Dinner


Fig. 1.14b(1) Berlin, Freedom, ser.1,1:235

While planters trembled at the prospect of the Emancipation Proclamation, slaves and free people of color prepared to celebrate the Day of Jubilee. In New Orleans, the officers of a "union association" asked General Banks to allow its members to observe the day in a manner befitting an event their people had awaited more than two hundred years.

New Orleans [La.] Dec 22th 1862

In obedience To Th High chift an Command of Th Head quarters Department of Th Gulf Maggor Gen N P Bank We Th members of Th union association Desir Th & Respectfully ask of you Th priviliges of Salabrating Th first Day of January th 1863 by a Large procension on that Day & We Wish to pass th Head quarters of th union officers High
presenting the documents in larger clearly annotation at the end of the
association to the Commander of the Gulf

New orleans Dec 22th 1862 ommand of Th Head quarters P Bank We Th members of carefully ask of you Th f January th 1863 by a Large ass th Head quarters of th is it Suit your approbation Dinner on th Second Day profit of th Dinner Will Go Woman & servant lent of th union association under of th Dinner

P. Bank, 22 Dec. 1862, M-62 yards - Banks' Expedition, Dept. and both signatures are in the letters sent in the records of the he Union Association.

1.14b(1) Berlin, Freedom, ser.1:1:235

Emancipation Procla-
pared to celebrate the of a "union association" to observe the day in a awaited more than two

1.15 Editors decide which genres of documents (e.g., letters, diaries, etc.) will be included in their editions as well as the sources from which those documents will be obtained. Establishing clear guidelines regarding the types of documents that will be used and identifying the locations of relevant groups of documents early in the planning process allow an editor to search for documents in a systematic, thorough fashion.

1.16 When deciding which types of documents to include in an edition, an editor takes into account both the scope of the project and the nature of available sources. Editors may choose to edit only a single genre of documents (such as an individual's letters or diary) or to compile a much larger selection of genres (such as speeches, memoirs, professional papers, telegrams, or published works). In addition, editors may want to specify the authorial origin of the documents that will be collected, deciding whether to limit collection to documents produced by the subject, received by the subject, written about the subject, or some combination thereof. A precisely defined scope and a familiarity with the content of major collections of documents will help editors make these decisions.

1.17 Editors should consider how many documents a project will collect, how thoroughly the editor or project staff will be able to search appropriate manuscript repositories and archives for relevant documents, and how many documents will be published. In all cases, the editors' aspirations should be matched by financial resources sufficient to see the project through to completion.

1.18 One document can stand alone as the focus of an editing project. A particularly interesting letter, a vivid diary, or an important published work may warrant editing and publication on its own. For example, see the short diary edited by John Hammond Moore and published as A Plantation Mistress on the Eve of the Civil War: The Diary of Keziah Goodwyn Hopkins Brevard (Columbia: University of South Carolina Press, 1993). A single
1.19 Documents drawn from a single manuscript collection or a solitary repository can provide sufficient sources for an editing project. Whether editors want to publish only a few documents or a great many, using the holdings of a single repository eases the collecting process, reduces the work of acquiring permission to publish documents, and may bring a sense of unity to the project, especially if the documents share a common history. Because many institutions have created or actively collected documents on particular subjects, a single library, historical society, or archive often may have a rich collection of documents on a given topic. For example, the editorial staff of *Freedom: A Documentary History of Emancipation* conducted all its research in the holdings of the National Archives.

In the fall of 1976, with a grant from the National Historical Publications and Records Commission, and under the sponsorship of the University of Maryland, the Freedmen and Southern Society Project launched a systematic search of those records at the National Archives that promised to yield material for a documentary history of emancipation. Over the course of the next three years, the editors selected more than 49,000 items, which represented perhaps 2 percent of the documents they examined. Indexed and cross-referenced topically, chronologically, and geographically, this preliminary selection constitutes the universe from which the documents published as *Freedom: A Documentary History of Emancipation* are selected and annotated, and from which the editors' introductory essays are written.
1.20 A documentary editing project may collect documents from multiple institutions. Many of the largest projects have completed worldwide searches spanning several decades that took the editors from historical archives to the holdings of collectors and manuscript dealers to the recesses of family barns and attics. Editors of smaller projects may also find themselves culling multiple collections at different institutions to acquire a larger or more diverse selection of documents. Rarely are both sides of a mutual correspondence located at a single institution. Likewise, some of the extant correspondence produced by a subject probably resides in the hands of recipients, their heirs, or various repositories. A thorough search in multiple archives may provide the only means of assembling a comprehensive collection of documents. Even though electronic cataloging, photocopying, and microfilming have made access easier than in the past, the cost of travel and lodging or of hiring researchers makes multiple-institution searches time-consuming and costly. Consider, for example, the difficulties incurred by the editors of the *Papers of Benjamin Franklin* in conducting their extraordinarily thorough worldwide search to gather the documents for their edition:

The editors’ first task was to locate and photocopy Franklin manuscripts and printed works in the scores of institutions and private collections where they are preserved. The major bodies of Franklin materials are well known to historians: the American Philosophical Society, the Library of Congress, the National Archives, the Historical Society of Pennsylvania, the University of Pennsylvania, Yale University, the Massachusetts Historical Society, and the French Foreign Office each owns more than 300 manuscripts. Scholars have long been acquainted with the Franklin papers in Harvard College Library, the William L. Clements Library, the New York Public Library, the New-York Historical Society, the Pierpont Morgan Library, the Henry E. Huntington Library, the Library Company of Philadelphia, the Pennsylvania State Records Office, and, abroad, in the British Museum, Public Record Office, Royal Society, and Bibliothèque Nationale. These twenty libraries and about a score more composed the first list of probable owners which the editors prepared when they began their work in 1954.

Obtaining photostats or microfilms from the larger libraries was, for the editors, relatively easy. (In a very few cases, for some special reason, manuscript transcripts had to be made.) By correspondence and personal visits single letters were located in smaller libraries and collections. University libraries, state and municipal libraries, and state and local historical societies were routinely
canvassed. The indexes prepared by the Federal Historical Records Survey of the Works Progress Administration provided helpful information, as did, in Great Britain, the Reports of the Historical Manuscripts Commission. The Catalogue Général des Manuscrits des Bibliothèques Publiques de France led to a dozen or more provincial libraries, and the admirable centralization of library facilities in France enabled us to examine and copy these manuscripts in the Bibliothèque Nationale in Paris. The director of the National Historical Publications Commission had a search made of the Continental Congress Papers and other promising collections in the National Archives at Washington and provided microfilms. Institutions which Franklin served, like the Pennsylvania Hospital and the Associates of the late Reverend Dr. Bray, almost always had a letter or two. Colleagues and strangers reported the existence of Franklin letters in their friends' libraries or in places to which we would probably never have addressed an inquiry, like the Riverdale Country Day School. Sometimes, as in the Salem County, N.J., Historical Society and in Christ Church in Philadelphia, while searching for a manuscript we believed to be there, we found others we had not known about. No place, it turned out, was too improbable to hold a Franklin letter: there are eight at Windsor Castle (seven of them presented to the Prince of Wales, later Edward VII, during his American tour in 1860), and two (each to a president of Yale College) in the Karl Marx University at Leipzig. We began as scholars, but have become sleuths and venturesome serendipitists as well.

Descendants of Franklin and his principal correspondents, especially in Philadelphia, without exception took a lively interest in the work, as their parents and grandparents did in the work of Smyth and Sparks, and allowed copies of their manuscripts to be made, or, if they had no manuscripts themselves, sent us to aunts and cousins who did. Autograph collectors, many of them members of the Manuscript Society, also responded to our requests and cordially and promptly allowed copies to be made of their treasures. As the first volume goes to press, about 220 institutions and some 110 private owners have given permission to print their manuscripts in this edition. All seem to regard The Papers of Benjamin Franklin, in words J. Francis Fisher of Philadelphia used to Jared Sparks, as "a national work," which they wish to promote as they can. Nor is this cooperation limited to owners in the United States. Photocopies have reached us not only from thirty-one states from Maine to Hawaii, and the District of Columbia, but also from twelve foreign countries extending from Canada to the Soviet Union. This response suggests that an edition of Franklin's writings may now be regarded as an international work.

Carrying on the publisher's policy of providing archival manuscripts, the first two volumes of the Pennsylvania Edition of the Papers of Benjamin Franklin end the last complete year, 1782. The editor was unable to write the memoirs of Franklin that are found in the first volume of the edition of his papers published in the 1850's by S. and Smyth. While the editor acknowledges that the truncated memoirs of his predecessor, which had been read to the Franklin Institute and the Philadelphia Society for the Promotion of History and Belles Lettres, are useful in the search connecting paper and page, he considers them anachronistic.

1.22 A comprehensive description of the project's scope and purpose is found in the introduction to the first volume. The editor has sought an edition large enough to embrace the entire Franklin papers; an edition large enough to contain as many original manuscripts, letters, drafts, or fragments as the editor believes necessary to facilitate the study of Franklin's papers.
Carrying the search to yet another level, the editors examined the published correspondence of Franklin’s contemporaries, printed archival collections, the periodicals of historical societies and associations; and found other letters, from now lost manuscripts, printed in whole or in part in the book and autograph sales catalogues of Goodspeed, Henkels, Maggs, Parke-Bernet, and others.

Even this exhaustive search is, of course, not ended. It will continue throughout the preparation of this edition, and even then manuscripts will continue to turn up. Just as Sparks, for some reason, did not or could not see the Franklin-Mecom correspondence that was preserved in Boston at the time he was working; and Smyth could find no trace of Franklin’s letters to Ingenhousz, which had been sold only a few years before; so we have been unable to verify in any way a report that Franklin letters are in possession of descendants of Joseph Galloway in Eire. So the search continues, and with it an appeal to those who read these pages to inform the editors (in care of the publishers) of any Franklin manuscripts of whose existence they know.

Fig. 1.20 Labaee, Franklin, 1 xxxvii-xxx

1.21 Because many interesting and important documents reside outside the collections of repositories, editors should consider other locations where documents might be found. Newspapers and magazines can provide texts of speeches, news accounts, public announcements, advertisements, and letters. Collectors, local historical societies, or individuals may own important documents or family papers that have escaped public attention. The files of businesses, newspapers, and organizations may also hold rare clippings or documents. While it may be beyond the means of documentary editors to locate all documents within a project’s scope, editors should think creatively when searching for documents and should seek advice from other editors and scholars who have already completed similar searches.

Comprehensiveness

1.22 A comprehensive edition includes all known documents within a project’s scope. A project with a narrowly defined scope and few pertinent documents may produce a short but exhaustive article or edition, while a project with a broad scope and many relevant documents may produce a very large although not necessarily comprehensive edition. While some comprehensive editions whose scopes encompass thousands of documents have been published as book editions spanning dozens of volumes, few topics are important enough to warrant the expense of reproducing routine papers, variant drafts, or form letters in such a format. More common are book editions that print the most informative letters in full but provide a comprehensive
overview of other documents (including information about where a reader could obtain those documents in the original) through devices such as extracts, abstracts, and calendars. (For examples of extracts, abstracts, and calendars, see Chapter 2.) When the reproduction of a large number of routine documents is essential to the scope of a project, it is common to produce comprehensive microform editions in which all documents are reproduced in facsimile form but not transcribed. Such an edition can stand alone or serve as a companion to a selective book edition.

1.23 A selective edition does not publish all in-scope documents discovered in the course of the document search. Most documentary editions are selective because editors will deem some documents unworthy of publication. Editors may omit documents to increase readability, to save time, to reduce printing costs, to meet space limitations, or because certain documents are inaccessible because of privacy or copyright considerations. Editors of selective editions should develop explicit criteria for selecting documents and then describe them clearly for their readers. (For a discussion of selection criteria, see Chapter 2.)
Managing a digitization project

Introduction

Understanding the digitization process is the best way to develop a sound basis for digital project management. This requires an understanding of all aspects of the digitization process, both technical (what equipment to buy, what standards to use) and administrative (managing people and workflows, and assessing quality). In particular, it will necessitate a precise understanding of all the issues involved at each stage of the imaging process. This knowledge will enable the project manager to make informed decisions about selecting appropriate digitization strategies for their own collections.

Project planning requires the development of comprehensive workflows appropriate to the original source materials and the desired objectives and project goals. Project management will necessitate the implementation of these plans within the constraints of the time and budget that are to be invested in the project. In addition to the expected workflow and components of running a digitization project, selection decisions and related issues (such as cataloguing, assessing the condition of originals, etc.) will add to the costs of such projects and the time they will take to complete.

The first part of this chapter focuses on identifying the core goals of the project and how they will affect the decisions that are made at each point of the digitization workflow. The second part outlines the technical workflow, technical choices, and management of the digiti-
zation life cycle. It is important to recognize that the digitization chain is not a straightforward linear series of actions – many activities will happen in parallel and at different stages of the project – and no single set of guidelines can be appropriate for all circumstances. There will be many components of the digitization chain, and these will change over time. However, the core elements of a technology project that have to be managed are:

- the handling, preparation and intellectual content of the original source materials
- the technology and processes that enable the source materials to be digitized and delivered to end-users
- the long-term maintenance and distribution of the electronic resources to ensure that they are useful over the long term.

Other sources of advice

Much excellent work has been produced on the detailed technical aspects of digitizing images and it is not the intent of this book to duplicate this advice. Instead, the reader is encouraged to consult some of these existing resources. These include *Moving Theory into Practice: a digital imaging tutorial* (Cornell University Library, Research Department, 2002). This tutorial offers base-level information on the use of digital imaging to convert and make accessible cultural heritage materials. It also introduces some concepts advocated by Cornell University Library, in particular the value of benchmarking requirements before undertaking a digital initiative. The reader will find technical information, formulas and reality checks on all aspects of digital imaging. A book accompanies the tutorial and expands on many of the key themes, including an integrated approach to digital imaging, and an emphasis on managing projects for preservation and access (Kenney and Rieger, 2000).

Other essential reading includes the recent *Digital Imaging: a practical handbook* by S. Lee (2001a) and *The NINCH Guide to Good Practice* (NINCH, 2002). Other online resources include *An Introduction to Making Digital Image Archives*, by the Technical Advisory Service for
Project planning and technology

The best-managed projects have clear goals. Brainstorming, the first phase of project management, is the time to talk about outcomes. 'Starting at the end' is an effective way to ensure smooth beginnings. Too often there is a tendency to dive right into the questions of technology - e.g., which scanner should I buy? - before articulating the purposes that digital reformatting must serve.

(S. Chapman, 2000)

Articulating the purposes of the project, and the way that the digital imaging processes chosen will create resources that will fulfill these goals, is the best way to plan a digitization project and will make it easier to narrow the choices of scanning technologies and techniques from a constantly expanding selection of equipment and software. Technology should always be assessed with an eye to the objectives of the project, rather than framing the goals of the project in terms of what is technologically available or possible. The factors that will influence these decisions are:

- the reasons for digitization
- the materials to be digitized, especially their condition and informational content
- the level of fidelity to the original that is required
- the technical and financial resources that are available to the project, and the scale of the project
- the potential uses and users of the digital objects
- any specific desired outcomes for the physical objects that are to be digitized.

These considerations will influence the decisions taken throughout the digital life cycle, especially questions of usage and the quality that is required of the digital objects. A frequent concern is how to achieve
the proper balance between quality and rate of production. Having a clear vision of the use of the digital materials and the quality required will help to focus such decisions. It is always preferable to reach a well defined and objective measure of affordable quality for the whole digitization chain, and to determine how many objects can be digitized at an acceptable rate of quality in the time available, allocating a longer time span to digitization if necessary. For all projects, ‘it is better to set modest goals than to create unrealistic expectations’ (Serenson Colet, 2000).

There are a number of approaches to representing original content that correlate the outcomes of the project to the requirements of the end-users and the level of fidelity to the original.

The first is often referred to as benchmarking and is a stringent approach that supposes that end-users of digital resources will have the same requirements and expectations as the users and creators of the original source materials. As such, this requires a careful assessment of the originals to ensure that digital reproductions will replicate the unique attributes of the originals as carefully as possible, by faithfully conveying such qualities and attributes as organization and presentation, size and dimension, and detail, tone and colour (S. Chapman, 2000).

Another approach is to anticipate the needs of the end-users, and to base digitization decisions on these criteria. This will necessitate predicting the hardware and software systems and network access available to the end-users, and implementing digitization decisions based on these factors. For example, are the users researchers who need to magnify or zoom in on materials for detailed image or palaeographical analysis? Or will they need to print out copies, or compare several versions of a document or image on screen at the same time? How legible do the materials need to be? It is important to remember, however, that if projects are conceptualized from the perspective of the users of today, it may be difficult to expand the scope of the project for other uses or purposes that may arise.

The best approach is one that will enable the project manager to ‘scan once and for all purposes’ by creating a fully documented high-quality ‘digital master’ from which all other versions (e.g. compressed versions for accessing via the web) can be derived. This digital master file should be created at the highest resolution and bit depth that is
tion. Having a quality required to reach a well the whole digitized, allocating a be digitized jects, ‘It is bet- be digitized jects, ‘It is bet-
ing content ements of the

t is a stringent will have the reators of the assessment of ate the unique tully conveying tion, size and
}), nd-users, and ill necessitate etwork access ion decisions rackers who nage or pale-
, or compare he same time? nt to remem- e perspective scope of the

t manager to mented high-


both affordable and practical, preserving to the greatest extent possible the authenticity and integrity of the original information. The digital master file will become an archive version of the data – it remains as pure a representation of the original as possible. Ideally, multiple copies should be stored on different media types and in more than one geographical location, thus providing a degree of protection against data corruption, media failure and physical damage to equipment. Creating a digital master minimizes the need to digitize again in the future and thus saves the original from excessive handling. It also prolongs the usefulness of the digital archive, which should serve the widest possible range of users and uses, including some that may be unforeseen or unknowable (Frey and Süsstrunk, 1999). This master file should be the source for every other digital copy required, such as low-resolution surrogate files for web delivery or high-resolution copies distributed for archival quality printing. All images used for purposes like viewing and printing should be produced as surrogates from the original archived images. Enhancements such as colour corrections and resizing should be performed on these surrogate images to ensure that they are appropriate for the uses intended. While surrogate images can be produced as part of the capture process, this can require large amounts of storage and processing time. Alternatively, a separate surrogate image production environment can be set up as an extension of the image capture process. From the user’s perspective, working with lower resolution web files allows browsing and fast access. The lower resolution and image quality also means that the images cannot be distributed illegally for commercial publication.

As such, this comprehensive approach is consistent with the ‘digital ecology’ movement, which recognizes that decisions taken at the very outset of a project will inform the long-term use and preservation of the digital materials. This is based on the understanding that digital projects have a ‘life cycle’, and that there will be a need to plan for not just the creation of the digital files, but their long-term management and possible conversion to another format for re-use (Lawrence et al., 2000).

Digital preservation issues must also be observed when producing digital content, and declaring an intended life-span for the objects will also help the decision-making process.
The project manager

The project manager will be responsible for balancing all these requirements and overseeing all aspects of a project: the ‘product’ or outcome; the schedule, which will have start dates, end dates and a series of milestone deadlines in between; and its resources: people, budgets, equipment, and source materials. These elements will have to be co-ordinated from the planning stages to implementation of the finished product. Activities will include developing and organizing the project team, supervising all elements of the project workflow, managing the project’s resources (including staff and budget) and co-ordinating outreach activities with all stakeholders, including the steering committee or advisory board and funders. Project managers will have to set targets for all aspects of the project and ensure that these are met. They will often be a conduit for the technical and academic aspects of the project, responsible for interpreting academic or project goals and ensuring that the technical specifications meet these goals, and that the technology chosen will be suitable for cultural heritage collections. Consequently, they will be responsible for most of the decision points throughout the project workflow and for allocating all activities to other members of the project team, such as technical staff, evaluation teams and content experts.

The project will be more successful if the project manager is able to build, support and maintain a team that can work well together, which can be difficult when project staff are located in disparate venues, or when participants from different backgrounds with no experience of working together have to collaborate (such as computer scientists and humanists, web programmers and conservation staff, and librarians and IT staff). Good project management is essentially about communication, and sometimes translation – if all participants in the project understand the intended outcomes of the project and the importance of their role, and have a respect for all other activities required to accomplish the initiative, there will be a greater impetus to work together to achieve these goals (S. Chapman, 2000). Regular meetings and discussions, open communications and working to agreed targets, deadlines and expenditures will be crucial in accomplishing such collaboration, as is maintaining staff interest in the project.

Managers should monitor every step of the process, and document
all aspects in order that the invaluable experience gained by doing digital projects becomes a resource for the whole institution. Quality documentation outlining the rationale, methodologies, systems, staffing models, costs and lessons learned from a project should be an outcome of every digital project, enabling future managers of digital collections to interpret why things were created in a particular way and what needs to be done to maintain, or even to improve, the digital objects (S. Chapman, 2000). Having documentation will also address the issue of ‘legacy projects’ – projects created by staff who have since left the institution, or which have languished for the want of attention since they were launched. Developing good documentation will mean that each digital project can be a ‘feasibility study’ (or a risk assessment exercise!) for the next.

Consistent staffing and management is the best way to ensure the longevity of digital projects. A project manager may manage many smaller projects simultaneously and therefore develop a broad understanding of the institutional factors affecting the implementation and operation of digitization initiatives, and the factors that will facilitate or hinder efforts to mainstream such activities. Consequently, individuals who fill such positions are an extremely valuable resource for the whole organization. Nonetheless, many project managers are often hired on short-term, externally funded contracts, and leave at the end of a project. Emphasis should always be placed on finding, hiring and, most importantly, retaining a good project manager. Their skills will be crucial to the successful completion of digital projects and the successful implementation of digital programmes.

The steering committee or advisory boards

Projects should consider assembling a steering committee or advisory group consisting of stakeholders and experts from within and outside the host institution. The group can consist of representatives of all aspects of the project: subject or discipline specialists, senior administrators, IT staff, conservation and preservation staff, evaluation experts, and experts who have participated in or managed similar projects at other institutions. It is important to think strategically about who should be on the steering committee – it is an excellent
opportunity to include senior members of staff at the host institution in the ongoing progress of the project. Including specialists in areas such as evaluation, dissemination and conservation is a good way to obtain ongoing expert advice throughout the course of the project. It is advisable to identify some key members of this group at the very early planning stages of the project and there should be an indication of who will be on this committee in any applications for funding. Project budgets should include the full costs of committee meetings, including travel costs and stipends where appropriate.

The functions of this group should be set out at the start of the project. These may include overseeing the project on behalf of funders, general project advice and oversight, reviewing progress reports, acting as advocates for the project and promoting it, and providing expert assistance where required. The group should meet on a regular if infrequent basis, and these meetings will be a useful opportunity for project staff and managers to report on progress and expenditure and to receive guidance and advice on various aspects of the project. A designated chair of the group should convene the meetings and ensure that the group is fulfilling the functions that are required of it.

Many funding agencies will mandate that such a group is convened, and some will require that a representative of the granting organization or agency is present at their meetings. Other funding agencies will ask that the project manager and other representatives meet with them on a regular basis to provide updates and progress reports.

Meetings with a project advisory board or funding agency should not be seen as an onerous requirement, but as a valuable opportunity to air ideas and gain support for particular approaches, and to obtain input and guidance from experts who will have a broad perspective on project management and implementation. The group will often be able to identify any concerns before they become problems, troubleshoot any issues, and strategize about political or institutional concerns such as dissemination and sustainability. Regular reports written for the group can provide a valuable record of the project.
Why digitization projects fail

As we have seen in the discussion on risk management (see Chapter 4), one of the most important aspects of planning a project is anticipating areas that may cause problems throughout the course of the project.

In a survey of risk factors for technology projects, the following reasons were identified as the cause of project failure (defined as going over budget, failing to complete the project on time, or otherwise failing to meet stated expectations):

- 32% - inadequate project management and control
- 20% - lack of communication
- 17% - failure to define objectives
- 17% - lack of familiarity with project scope and complexity
- 14% - incorrect technology, project size & other.

(Tanner, 2001b).

Digitization projects present many classic management issues: they are complex and costly, and often have an alarming tendency to manifest 'project drift' away from the original goals and objectives. This is a problem inherent in the very nature of technology-based projects – it is extremely likely that during the anticipated time to be spent on a project (be it six months or several years), some exciting new technology will come out that will benefit the project enormously. However, even minor changes in equipment, software, personnel and procedures between pilot projects and production projects can cause unexpected results. The project manager will have to steer this path carefully. Because of the complexity and expense of digitization projects, and because they take up so much management time, they can easily skew other priorities. A common perception is that other programmes and initiatives may suffer if the digitization programme or project starts to take up more resources than originally intended.

Effective project management is the key to keeping things on track and to ensuring that the project does not deviate from its stated goals or schedule. Some of the skills and activities that will help this process include:
• communication
• close management of staff resources
• implementing a good steering or advisory committee
• developing specifications based on the best advice obtainable, and fully tested with the people who will do the work
• regular and efficient quality control
• using cross-discipline teams to manage projects, to draw on a range of views and experience (Webb, 2000).

Another potential problem is loss of momentum from project conceptualization to funding to completion – there may be long time lags, and the project manager will have to keep enthusiasm for the project going as people become distracted by other priorities. There are a number of strategies to combat this by making the work of the project staff more interesting. The University of Michigan reports that staff engaged in the relatively tedious and repetitive tasks involved in the scanning of journals for the Making of America project were encouraged to work on different parts of the project’s workflow, in order to break some of the monotony of only working on, say, scanning or metadata creation (Bonn, 2001). Other approaches are to encourage and support faculty use of the emerging digital materials in teaching, and by emphasizing as many technological ‘advanced research’ aspects of the project as possible in order to keep participants interested and engaged.

Managing the digital workflow/digitization process

The digital life cycle

The digital life cycle has many components. It is necessary to understand all these elements, and how they are connected, in order to see how a digital project can be managed. This approach is known as the ‘holistic’ approach to digitization – decisions taken about one aspect of a project will affect the entire digital life cycle. The components of running a digitization project cannot be presented as a straightforward ‘recipe’ to be followed, via a linear series of actions – many activities will happen in parallel and at different stages of the project, and many activities can overlap. The project manager will have to co-ordi-
nate these activities and the people responsible for them, and maintain the project's momentum if there are long gaps between activities. It will also be important to understand how these activities impact on one another - for example, if copyright clearance will take a long time to complete, it will be unwise to purchase equipment for scanning until permissions have been cleared. There will be local collections and institutional considerations that affect the sequence of many of these activities, especially policies and procedures related to conservation and physical control of the original materials.

The workflow for a digitization project will consist of all or some of the following activities:

- Articulation of the project goals and intentions.
- Selection of materials for digitization.
- Copyright clearance, rights management research, plans and strategies for managing rights and permissions, development of rights and permissions metadata (administrative metadata).
- Preparation of source materials. This may include moving them to the digitization centre, conservation assessment, or treatment, such as re-housing or remounting.
- Creating catalogue records, finding aids or other pointers to a digital object or collection (descriptive metadata).
- Digital image production. This will include choosing digital capture devices, setting up imaging equipment, scanning source materials to create digital masters and associated technical metadata, and processing master images to create derivatives for screen or print (using image-processing tools, or optical character recognition tools where appropriate).
- Quality control for source materials and digital images, transfer of sources to original or new location, re-housing materials, updating catalogue records as necessary.
- Creation of structural metadata.
- Creation of full text and mark-up where required.
- File management; loading and delivering content to a database or repository.
- Integration of digital images and metadata into an image database.
- Delivery to users, using techniques ranging from developing web
pages to using an automated delivery system such as a database. 
- Advertising, promotion, user evaluation.
- Long-term preservation, including migration and conversion (S. Chapman, 2000).

The project manager will have to allocate a cost to each of these activities. Costs can often be projected using the results of prototyping and feasibility tests, as outlined in Chapter 6. Selection decisions and related issues – cataloguing, assessing the condition of originals, etc. – will add to both the costs and the schedule. What will be harder to anticipate than monetary costs will be the amount of time that is to be spent on each activity.

Possible risks associated with each of these activities will also have to be identified and managed. These include managing intellectual property issues, the abilities and experience of staff, the available technologies and their capacity and limitations (including tools for delivery of the digital content), the amount of work to be done versus the available time or money, and the possibilities of delays and how they affect the overall life cycle of the project if activities are interdependent.

Delivery of materials

Before starting a digitization project, it is essential to consider access and delivery options. Technical decisions taken early on in the planning process will affect the quality of information that can be delivered to remote users, and this is an area where testing and end-user evaluation is crucial. Many projects make the mistake of not assessing anticipated users until very late in project – often shortly before the project is launched – rather than evaluating users at the outset. It will also be necessary to consider questions of security, authenticity and authentication. Not all content will be for public access, and copyright and licensing restrictions may mean that restrictions have to be placed on content delivery. Tools such as those emerging from Internet 2’s Shibboleth project may be useful in this regard. Shibboleth is developing architectures, policy structures, practical technologies and an open-source implementation to support inter-institutional sharing of online resources subject to a series of access controls (http://shibboleth.internet2.edu/).
ch as a database.

nd conversion (S.

ach of these activi
ties of prototyping
tions and
1 of originals, etc.
: will be harder to
of time that is to
es will also have to
; intellectual prop-
available technolo-
gs for delivery of
ersus the available
ow they affect the
pendent.

 onsider access and
1 in the planning
un be delivered to
nd-user evaluation
essing anticipated
ere the project is
It will also be nec-
nd authentication.
ight and licensing
aced on content
's Shibboleth pro-
ing architectures,
source implement-
ourses subject to
.edu/).

Key issues related to the delivery of materials include:

- **Graphic design.** Is a project team member able to design a user interface that is appealing and easy to use, and enables all the 'scholarly primitives' end-users may need: discovering (browsing, searching), annotating, comparing, referring, sampling, illustrating and representing (Unsworth 2000). For example, the maps at www.davidrumsey.com fulfill many of these criteria.

- **Amount of information.** How much information is to be delivered? Are there large images, audio or moving image files? In which case, how will the users download these files? Are thumbnails to be presented for ease of browsing – which will necessitate creating an additional surrogate image as a JPEG, GIF or PNG file.

- **Organization of information.** How is access to the information presented? Is it a logical approach to the content? How many 'layers' of information do end-users have to go through in order to reach the core content (see Goldman and Wadman, 2003)?

- **Types of metadata.** What metadata is used? Is it sufficient for end-users to be able to find the digital resources and navigate their way throughout the information?

- **Types of media.** Are plug-ins or additional tools required? Will users find it easy to access these? Is content to be streamed or downloaded? If streaming is to be used (for copyright reasons, for example), can access to the site be guaranteed, especially if the content is to be used for teaching or other activities that require certain access? Are standard formats used for downloads (e.g. MPEG2 or MPEG4, QuickTime)? Do search tools meet the needs of end-users?

- **Accessibility.** Are there multilingual issues? – Is a mark-up system like Unicode required to support multilingual searches? Is accessibility for the disabled or partially sighted anticipated through the use of tools like the Bobby standard?

A useful overview of the attributes of delivery systems for digital materials that should be considered is the International Council of Museums (ICOM) guide to *Multimedia Evaluative Criteria* (ICOM, 1997). Although developed with a museums audience in mind, it is an
extremely comprehensive overview of all the factors that will affect the delivery of digital content.

Evaluation

It will be important to ensure that strategies are in place to evaluate end-users of the digital materials. This should take place:

- before the project has begun, to assess possible audiences for such materials and their expectations and abilities (front-end evaluation)
- throughout the development of a project, to assess the content, design and usability of the resources (formative evaluation)
- at the completion of a project, to assess the effectiveness of the delivery system in conveying the desired message of the content to the desired audience (summative evaluation).

Any evaluation should involve groups of users who approximate the intended audience for the finished project, whether they be students, schoolchildren, members of the public, or academics.

The attributes of delivery systems that should be assessed include:

- ease of use and accessibility
- functionality: how well does the system work? Are the interfaces intuitive?
- consistency of content and finding aids
- navigation and control for the end-user
- help and documentation
- the time taken to access, download and search the material.

Quantitative targets and models can help in setting parameters for these objectives, and monitoring their development throughout the course of the project (Keene, 1998).

Assessment and preparation of materials for scanning

The shape, size and condition of the original source materials will
rs that will affect place to evaluate place:

udiences for such front-end evalu-

asses the content, evaluation)

tfectiveness of the ge of the content

approximate the they be students,

assessed include:

re the interfaces

he material.

g parameters for throughout the

ing

nce materials will influence the decisions about which capture method to choose. Primary source materials in museums, archives and libraries consist of a variety of data formats (texts, images and sound or moving image), and the digital capture of that data needs to be carefully considered in relation both to the content and to the substrates on which the material is carried. Sound, moving image and fragile or rare material will be addressed in more detail in Chapters 8 and 9.

Anything that can be photographed can be digitized, and some materials can be photographed or digitized with more accuracy to the original than others (Deegan and Tanner, 2002). However, the attributes of the original materials will influence the decisions made about the type of digital imaging equipment and methods to be used:

Format. Materials can be bound or loose leaf, printed or handwritten, and colour or black and white, or a combination of these, such as pages with both print and images. A collection may include materials in a variety of formats, including manuscripts, three-dimensional objects and images. Can materials be laid flat for scanning, or is some sort of cradle required? Can the original materials be scanned, or does a surrogate image (e.g. a photograph or microfilm) have to be created?

Size. Materials will come in a variety of shapes and sizes, from small manuscript fragments to oversized maps or posters, banners or flags.

Detail. There will be questions concerning the level of detail in the originals, such as tonal range, that must be captured, and the level of detail that is to be delivered to the end-users.

Preparation of materials. When is the best time to repair and replace materials? Should this be factored into the workflow? How can a balance be struck between usability, cost and the best representation of the artifact? (For example, should blank pages at the beginning and end of each volume be included?) Post digitization, should the original volumes be kept? For how long? And what is the best way to collaborate with others on cost-effective physical and virtual preservation and giving access to digitally reformatted volumes?

If the collection to be digitized consists of a large variety of original materials, it may be appropriate to outsource the digitization of some materials, and do the rest in house. For example, if only a small number of materials need special attention (e.g., they are oversized, or
three dimensional), and the rest can be scanned on a flatbed scanner, the unique materials may be outsourced and the rest done in house. Exceptions, such as taking the time to digitize the occasional image, or to handle special preservation concerns, always increase costs.

Handling and preparation of original materials

Preparation and handling costs add a significant cost to the project budget – in some cases, as much as 30-50% of the project total (Tanner, 2001). If these costs are not anticipated, they will jeopardize a project’s budget.

The movement of materials, especially if they are to be digitized off site, can take a great deal of preparation. They will have to be inventoried and packed, and insured against damage if they are especially rare or valuable. The materials may also have to be disbound, unstapled, cleaned or remounted prior to digitization (for example, the Afghan Papers project at NYU is carrying out preservation treatment on every item in the collection prior to digitization).

Each item to be digitized should be catalogued and a unique identifier assigned, which may correspond to the file name of the digital object. It may also be necessary to develop indexing, cataloguing or metadata schema. If applicable, the cost of clearing copyright and obtaining permission to digitize the collections can be the largest preparation cost. For an example of a quantification of how much time and money can be spent on copyright clearance, see Cave, Deegan and Heinink’s (2000) report on copyright clearance workflows. They estimate that the cost of simply doing the administrative work of clearing permissions was £5-6 per document during the first year, dropping to £2-3 per document in subsequent years. The time spent on the project was that of a staff member two days a week. Recent assessments by staff on this project suggest that these original cost estimates are probably too low. Similarly, New York Public Library is still developing rights management methodologies for its African Diaspora digitization project, and a half-time staff position is dedicated to clearing and managing copyright for these materials.
The role of surrogate images in the digitization workflow

At the project planning stage, the project manager will have to address two key questions regarding the role of surrogate image formats in the digitization workflow:

- In cases where the original material exists, should digitization be from this original (e.g., black and white or color photographs, negatives, glass plate negatives, manuscripts, drawings, prints or maps) or from a surrogate image (slides, transparencies, copy prints, microfilm)?
- What, if any, is the role of microfilming in the project workflow?

It may not be possible to digitize directly from the original – its condition or accessibility may preclude the material being handled or exposed directly to scanning or digital photography for conservation reasons. Similarly, digital imaging technologies may not be adequate to cope with certain types of originals, such as oversized maps, banners or posters or three-dimensional objects.

Sometimes, only a secondary image will be available. This is often the case in projects to reconstruct or reunify ‘lost’ collections, and is frequently the case in photography or slide libraries, where slides are often retained over original photographic prints for quality reasons (the art historical community prefers slides to prints). The quality of negatives can vary greatly, although they are the first generation of an image, there may be substantial differences between the negative and the print, especially in fine arts photography, where artists will spend a great deal of time in the darkroom creating prints (the work of Ansel Adams is one of the best examples of this). The results of this work will be lost if the negative, not the print, is scanned (Serensen Colet, 2000).

However, working with an intermediary may result in a loss of quality. Any copying process may result in the loss of quality between the original and the surrogate, as an operator will have to make decisions about focus and colour based on the individual’s ability to perceive the full light spectrum. A scan from a surrogate will only be as good as the surrogate, not the original, and this proposition is demonstrated by some useful comparisons and research carried out by the Digital Image Archive of Medieval Music, a collaborative project
between the University of Oxford and Royal Holloway, University of London (www.diamm.ac.uk/).

In cases where it is nonetheless necessary to create an intermediary for scanning, the advances in digital capture (especially digital photography) mean that a digital record can be created that is at least equal to, and occasionally better than, a good photographic copy (and eliminates the expense of this additional step). If direct digital capture is selected as an option, it will be important to create an imaging environment that replicates the care and attention of preservation photography – for example, camera angles, lighting and exposure rates should be given as much attention as with traditional photography, and a professional photographer should be employed to operate the digital capture device.

What, if any, is the role of microfilming in the project workflow?

There will often be a role for microfilming in the project workflow. Stuart Lee (2001a) concludes that there are four situations in which it is necessary.

- if the collection already consists of a good stock of film surrogates, and if the quality of these is deemed to be acceptable
- to match preservation needs, if a conservation decision has been made to make a film copy of an item
- if traditional photography is deemed to pose less of a threat to the original
- if the original cannot be digitized using the equipment available, for example because of constraints of size in the case of maps, etc., or if there are time efficiencies to be gained by filming the materials rather than scanning, as may be the case if working with bound volumes.

Microfilm provides a stable format, with a very long life (estimates range from 200–500 years if the materials are stored and handled properly); it is platform independent and accessible via lens-based hardware. It can be used as a source for re-mastering or digitizing throughout the course of its life cycle, and will not require migration or refreshing.
Filming is a well established technique, with existing workflows and reliable vendors to choose from, and is considered to be the production process that is most appropriate with regard to the integrity of the originals. This may therefore be a better solution for some projects rather than having to set up a scanning workstation that takes into account preservation conditions and special handling requirements. The use of preservation microfilm standards (notably *RLG Guidelines for Microfilming to Support Digitization* (2003b)) will produce microfilm that can be scanned by most microfilm digitization service bureaus with less handling, and less cost, because it is uniform and therefore easier to scan. The *RLG Guidelines* cover areas such as preparation, targeting and indexing of the film. They suggest a technically rigorous approach to preservation microfilming to make it more amenable to digitization, which will help libraries and archives ‘achieve the viable option of using microfilm for preservation and digitization of the film for enhanced access’ (Robin Dale, writing in the preface to the 2003 *Guidelines*).

What is emerging as a suggested strategy is hybrid approach, in which a combination of both microfilming and digitization of materials is carried out. This approach recognizes that microfilm has the edge as a preservation format but that digital technologies offer advantages for accessing the content, notably via indexing, searching and browsing.

In practical terms, many projects are finding that both scanning and microfilming often go hand in hand, which can significantly add to the project’s workflow and costs. The University of Iowa Libraries reports that a project to digitize part of their special collections had four distinct workflows: preservation photocopying, keying and encoding textual information for full-text searching, cataloguing the individual materials, and digital imaging of the same materials. The pieces resulting from each of these separate workflows had to be re-assembled in physical and digital format in the final phase of the project (Hughes, 2000).

**Scan first or film first?**

If the hybrid approach of scanning and microfilming is deemed to be necessary, there are two possible approaches, both developed in the context of research carried out for the Brittle Books project, funded by the NEH in the USA:
the model developed at Cornell University Library, which is to create digital images directly from the originals, and then produce film from the digital images using Computer Output Microfilm or COM (i.e. to scan first)

- the model developed at Yale University Library, which is to first microfilm the original source materials, then scan from the film, i.e. to film first. This model would suit an institution with a great deal of experience in preservation microfilming. One project that is following this model is the Australian Cooperative Digitisation Project 1840–45, also known as Ferguson 1840–45, a collaborative project between the University of Sydney Library, the State Library of New South Wales, the National Library of Australia and Monash University Library to digitize Australian serials and fiction.

For a discussion of both approaches, and the economics of scanning first versus filming first, see Chapman, Conway and Kenney (1999a) and the CLIR Working paper by the same authors (Chapman, Conway and Kenney, 1999b), which provides a decision tree to assist institutions in determining the circumstances under which one would scan first or film first.

**Equipment**

The types of digitization equipment available include:

- flatbed scanners, which may have sheet-feed attachments for batch scanning of disbound documents
- book, or overhead scanners
- microfilm or transparency scanners
- digital cameras.

These can be categorized as using 'contact' and 'no-contact' methods of image capture, and the choice of equipment will depend on the type of material to be digitized. Contact equipment, such as flatbed scanners, transparency scanners and drum scanners, will necessitate placing an original flat against the scanbed in order to scan the image. This approach will only work if the original is flat (for example, photographic materials, sheets of paper) or can be pressed flat
rary, which is to
s, and then pro-
per Output Micro-

which is to first un from the film,
ition with a great One project that ative Digitisation 5, a collaborative library, the State of Australia and erials and fiction.
s of scanning first y (1999a) and the Conway and Ken-titutions in deter-first or film first.

---

Scanners

It is necessary to ensure that the capability of the selected scanner is suitable to the materials to be scanned and to the project’s objectives. Both scanners and digital cameras use a CCD (Charge Coupled Device) array as their core component. CCD is a a collection of tiny light-sensitive diodes that sweep across an image during capture and, when exposed to light, generate a series of digital signals that are converted into pixel values. The most important factor is the resolution that the scanner is capable of. The scanner will often be listed with a maximum optical resolution and an interpolated or software resolution. The optical resolution is the significant figure: interpolated resol-ution uses software to ‘guess’ the values of pixels that are between those that the scanner can optically register. Interpolation should be avoided in an archive-quality scanning exercise. Where resolution is listed as, for example, 600x1200 dpi the maximum optical resolution will be 600 dpi. For more information on these requirements, and many other questions about scanning, see www.scantips.com.

The dynamic range of the scanner describes the tonal density of the information that can be captured. The higher the dynamic range
the better, particularly for originals with a large amount of tonal detail (e.g. shadow, or other subtle detail), such as photographic prints and transparencies. Production-level flatbed scanners usually have either an A4- or an A3-sized scanning area; scanners are available for larger materials, but they are more expensive. Transparency scanners are available which can scan strips or mounted 35 mm negative or positive transparencies to high resolutions. Adaptors are available to batch scan up to 50 slides at a time. Scanners range in price from tens to thousands of pounds, and the investment made should be appropriate to the requirements of the project.

Kenney and Rieger, in their digital imaging tutorial (2000), discuss these issues in greater detail, and suggest that the following questions should influence the selection of a scanner:

- Is the scanner compatible with the source material and can it handle the potential range of sizes, document types (single leaf, bound volume) and media (reflective, transparent) and the condition of the originals?
- Can the scanner produce the requisite quality to meet the needs of the project? It is always possible to derive a lower-quality image from a higher-quality one, but no amount of digital magic can accurately restore detail that was never captured to begin with. Factors affecting quality include optical resolution, bit depth, dynamic range, and signal-to-noise ratio (see ‘Images: technical details’, page 187).
- Will the scanner support the production schedule and conversion budget? Are the manufacturer’s throughput claims valid? What are the document-handling capabilities, duty cycle, MTBF (Mean Time Between Failure) and lifetime capacity? What kind of maintenance contracts and service agreements are available (does the manufacturer or distributor offer on-site repair and/or 24-hour replacement to eliminate potential downtime in the schedule if there is an equipment failure (Kenney and Rieger, 2000))?  

Scanner specifications can be difficult to interpret and often lack standardization, making meaningful comparisons impossible. For detailed, specific information about selecting digital imaging equip-
ment, see the feasibility study for the JISC Image Digitisation Initiative (Tanner and Robinson, 1998). See also the RLG, CLIR and DLF Guides to ‘Quality in Visual Resource Imaging’, particularly Guide number 2, Selecting a Scanner’ (Williams, 1998).

The software that runs the scanner is also important. It should be straightforward to use and incorporate an ability to run batch scans to save capture time, where applicable (NOF-Digitise, 2002a).

Digital cameras

Digital cameras are developing for both the home and professional market and are priced from several hundred to thousands of pounds. Home-use cameras are aimed at non-professional users for general photography. There are two kinds of professional digital camera. The first has developed from medical and industrial uses and has all the parts (lens and CCD array) integrated into one single unit, which is connected to, and operated directly from, a computer. The second has the appearance of a traditional camera, but replaces film with a digital CCD array with internal processing and memory, which stores the images until the memory is full, at which point the images can be transferred to a computer. This is known as a digital scanning back. Digital scanning backs are becoming an acceptable replacement for traditional film cameras for professional use. One of their advantages is that they use the lenses and camera body of a traditional professional camera. Both types of digital camera are used in digital imaging projects, although the first type of camera will cost considerably more than the second.

Professional digital camera set-ups will require the operator to at the very least understand the basics of photography and for projects where high-quality images are crucial it is recommended that they have a professional photography background. For information on selecting a digital camera, see the TASI section on digital cameras at www.ukoln.ac.uk/nof/support/help/papers/digitisation_process/.

Network assessment

It will be necessary to establish if the institution’s network capacity is
adequate to support a large-scale digital imaging project. Networks are probably the least visible portions of the technical infrastructure, managed by unseen systems departments. But nothing can bring a digital imaging initiative to a halt faster than a network that is undersized, too slow, or unreliable, especially if large image files have to be moved around the network for storage and access. A heavily used digital image collection will place even greater demands on your network. Institutions may find that a digital imaging initiative makes demands on an existing network that have implications for the entire organization. A discussion with network administrators about the anticipated network demands (including long-term storage on institutional servers) should take place early in the project-planning stage (Kenney and Rieger, 2000).

Equipment procurement

It will also be necessary to address any weaknesses in your institution’s procurement process that may slow down or complicate the acquisition of this equipment. It may be necessary to spend a lot of time researching available scanning equipment, requesting samples, and ensuring that expert advice is taken about what equipment to choose. This process can take a lot of time, and the market for this technology changes very quickly, meaning that purchasing decisions will have to be made quickly. Also factor in the time to set up and test an imaging workstation, and to set targets for image capture.

Other equipment

Other equipment that will need to be considered will include computer hardware (CPUs and monitors) and printers. It is always advisable to buy the highest-quality equipment that is affordable, and to maximize available processor speed, RAM, and disk space. Requirements will vary, depending upon the needs of a particular project, and what other purposes (if any) the equipment is to be used for. (While dedicated workstations should be allocated to digital imaging projects this isn’t always possible.)
Images: technical details

Image quality

*Resolution* refers to the number of pixels used to display an image. Resolution affects the level of detail that can be shown in the image file. Lower resolution means fewer pixels will be used to describe the image, leading to blurring of edges and loss of overall detail in the image (Beamsley, 1999). Resolution is usually expressed in dots per inch (dpi) or pixels per inch (ppi), and relates to the density of information that is captured by the scanning equipment. Broadly speaking, the higher the dpi, the more detail is being captured. The amount of resolution required to obtain a useful image of an item is determined by the size of the original, the amount of detail in the original and the eventual use for the data. For example, a 35 mm transparency will require a higher dpi than a 5x4 print because it is smaller and more detailed. An A4-sized modern printed document that is intended to be processed into a searchable text will need less resolution than a similar-sized colour photographic original. There are also upward limits on resolution - file size is one (increasing resolution will increase the file size) and another is preventing the capture of extraneous information. For example, postcards are often printed on poor-quality paper and if they are scanned at too high a resolution the texture of the paper will be captured and can obscure the content. There is also a point where putting more resolution into the capture process will no longer add value to the information content of the digital output.

*Bit-depth*, or *dynamic range*, relates to the level of colour that will be captured. A bit is the binary digit that represents the tonal value of the pixel. As an overview, a 1-bit image is black and white (the pixel has 1 bit and is therefore black or white with no shades in between), an 8-bit image has 256 shades of either grey or colour ($2^8 = 256$ shades), and a 24-bit image has millions of shades of colour ($2^{24} = 16,777,216$ shades).

Essentially, the digitized images should be of high enough quality to meet most of the institution’s foreseeable needs and to satisfy the demands of researchers seeking access to the collections using high-resolution monitors or advanced printing techniques (NOF-Digitise, 2002a).

Comparisons of suitable resolutions for digital master files for various media types are discussed in the JIDI feasibility study (Tanner and Robinson, 1998), which contains a useful table of baseline standards.
of minimum values of resolutions according to original material type. A detailed discussion of resolution and of binary and bit depth can be found on TASI's web pages (www.tasi.ac.uk). A comparative analysis of the different image qualities used in a variety of cultural heritage projects can also be found in the Cornell digital imaging tutorial (Cornell University Library, Research Department, 2002).

**File size and compression**

Production of high-quality images is much more expensive and takes up much more storage space. File size will affect the amount of work that can be done in a specific time and will affect workflow, especially files that are over 100 Mb in size. However, if the intention is to create an archival master file, compromises cannot be made on file size. Reducing image sizes by a small amount, e.g. from 100 Mb to 80 Mb will not significantly change the number of images one can capture in a given time, as most of the time spent is in other aspects of the workflow, such as setting up the source materials on the scanner or setting up a digital camera, editing the work, processing images and backing up the files. Efficiencies and savings should be gained by assessing where the workflow can be made more efficient, rather than by reducing the image size. However, it will be necessary to research the storage required for back-up and archiving of files, and to assess the cost of storing and backing up the digital images over the long term.

Digital image projects will create expensive requirements in terms of hard disk space and processing power, in order simply to open and edit a file and to view multiple images simultaneously. This will create cost considerations that will have an impact on the overall project budget (Sersenzen Colet, 2000).

The digital file can be compressed to reduce its size. Compressed files are classified as using either lossy or lossless compression techniques. Lossy compression will reduce the size of the file by a ratio from 10:1 to 50:1, but this reduction is at the expense of some of the file's original data, which will affect the integrity of the image. Lossless compression guarantees that the original data can be restored exactly, and so this is the format that should always be used for compressed master images. Typically, lossless compression results in a compression ratio of about 3:1 or 2:1.
Formats and technology standards

The end products of digitization will be files in a variety of formats, depending on their medium. These will include images in many file formats, electronic text files, digital video or audio, virtual reality or 3-D materials. Managing these files will be easier over the long term if conformity to standard file formats is observed from the outset of the project.

A wide range of technology standards exists, reflecting the long history of the computer industry, which develops and adopts new standards on an ongoing basis. It will be necessary for project managers to acquaint themselves with the community and industry standards for digital media. These standards will change over time, and it is important to maintain an understanding of the field in order to keep up to date with changes in technology. The use of relevant standards will be critical if a project is to share and publish the digital files that it has created. Formats used will also be dependent on what sort of compression is required.

Image standards and formats

TIFF (Tagged Image File Format)

This format is used when creating digital images for long-term use and for maximum information capture. It does not use compression, or, if a compressed form of TIFF is used, it will be lossless, so files saved in the TIFF format will be large, high-resolution and high-quality files. Any scanner or digital camera, and the software used in association with them, will provide the option of saving to the TIFF format. This format is recommended as the acceptable format for saving archival master copies of digital images (Kenney and Rieger, 2000). Because TIFF files are so large, it is impractical to transfer them over networks unless over a specially allocated high-speed network such as Internet 2, or to use them for day-to-day project work.

JPEG (Joint Photographic Experts Group)

This standard is used to deliver images across networks, such as the inter-
net or local networks. JPEG files use lossy compression to reduce the file size, while retraining a reasonable-looking image. Consequently, this file format is not appropriate for master archival files. The use and display of JPEG files is supported by all web browsers and by a large number of desktop applications. JPEG files can be created by image-processing software, which will import a TIFF image and export a JPEG one. The JPEG format is suitable for scanned photographs and complex images.

GIF (Graphics Interchange Format)
This format is used to deliver images over networks, usually the internet. Although the display of GIF images is supported by all web browsers and many applications, it is a proprietary file format, covered by a patent. GIF files utilize lossless file compression to reduce the size of the file. This is the file format recommended for cartoons, icons and similar graphic images. GIF files are created by image-processing software, which imports a TIFF file and exports a GIF image. Because they are compressed, both JPEG and GIF files may be considerably smaller than TIFF files (they can be measured in kilobytes).

PNG (Portable Network Graphics)
PNG images are supported by recent versions of web browsers. PNG files have more efficient compression than GIF or JPEG, and images may look better in this format, but they are a larger file size. Most image-processing applications will offer PNG as a format in which TIFF images can be saved.

Audiovisual standard formats
The formats listed here are the most commonly used for web delivery of audiovisual materials, but this is an area where there is a good deal of commercial development affecting standards. The MATRIX project and the AHDS Performing Arts Data Service provide regularly updated information on developments in audio moving image technologies. There is also an excellent comparative chart of file formats for time-based media in the NINCH Guide to Good Practice (NINCH, 2002).
Audio formats

Audio can be un compressed, compressed without loss, or compressed with loss. Lossless compression will only reduce the file size by a factor of about 2:1, so if compression is used, it is usually lossy. For uncompressed audio it is necessary to know the number of channels, usually stereo, i.e. two, or mono, i.e. one; the number of bits per sample; and the sampling rate. The CD standard is stereo, 16 bit/sample, 44100 samples per second. In terms of storage, that requires about 10 Mb per minute. (For archival purposes 'CD format' stereo at 16 bits and 44100 samples per second is the minimum accepted, because it corresponds almost exactly to our measured ability to hear.) A .wav file at the same sample depth and rate will take about the same amount of space, as will any uncompressed, or loss-free compressed, audio format.

For 'lossy' compression of both audio and visual recordings, it is difficult to draw useful comparisons about file sizes. While these algorithms are based on theory, there is no a priori way to compare the alternatives. Companies that are in this business use extensive listening tests to verify what they think should work. While it is possible to compare the alternatives in terms of file sizes, there is no way to know the cost of lossy compression. A file twice as large may not be twice as better sounding.

For some guidelines of comparisons, see the National Institute of Standards and Technology's publication, Digital Media File Types: survey of common formats (http://www.itl.nist.gov/div895/isis/filetypes.html).

WAV

WAV is the standard Windows audio file format, and is supported by recent versions of Windows via the inbuilt Windows Media Player (this comes pre-installed on most Windows computer systems). All audio CDs are in WAV format.

AC3

AC3 comes from Dolby and it is the multi-channel surround sound format commonly used in DVDs. It is sometimes called 5.1 audio.
PART 2: DIGITIZING COLLECTIONS

MP3

Part of the MPEG family of multimedia standards, MP3 files are small in size and produce a quality of sound that will be dependent on the compression selected. Although MP3 does not reach true uncompressed CD quality, the quality will be adequate for many users and this is the preferred format for the delivery of popular music over the internet (such as by Apple’s iTunes Music Store, which sells music for downloading). MP3 files can be used by the Windows Media Player. Note that MP3 is a lossy compression technique. AAC, a new audio format that is part of the MPEG4 standard, may well replace MP3 by offering better audio quality and smaller files, thanks to improved compression technology (for more information, see www.apple.com/mpeg4/aac).

Real Audio

This is a commonly used file format, owing to the free availability of Real Audio player software, which can be downloaded from the Progressive Networks website (www.real.com).

Digital video standards

The entertainment and technology industries have a vested interest in developing efficient and interoperable formats for the customer. Consequently, digital video formats are rapidly changing. In addition to the existing standards listed below, the reader is advised that there are emerging standards on the horizon, including DivX (see www.divx.com/), which is an up-and-coming video format, and CODEC (COmpression and DE-Compression tool).

MPEG (Motion Pictures Expert Group)

This format is popular for the publication of video on the internet, as it uses a short download time. A large range of player software, including the Windows Media Player, supports MPEG files. Sound and video may be combined in one file (such as in a motion picture or narrated documentary). MPEGs use psycho-acoustic algorithms to encode the most 'prominent' information only and so MPEG files can be highly
MP3 files are small dependent on the each true uncompa-nany users and this usic over the inter-ls music for down-media Player. Note new audio format x MP3 by offering roved compression om/mpeg4/aac).

free availability of dded from the Pro-

vested interest in he customer. Con-ag. In addition to advised that there ding DivX (see deo format, and on the internet, as r software, includ-

Sound and video picture or narrated ms to encode the files can be highly

compressed, while still maintaining a reasonable quality. MPEG is becoming a dominant standard in the field and MPEG formats are used for DVD video and digital television broadcasts. Like MP3, the quality is selectable depending on the acceptable size of the final file. High-quality DVD can use data rates of up to approximately 7 Mb per second (Mbps). There are different types of MPEG files available: MPEG-1, an early (and now fading) standard; MPEG-2, which is used with DVDs and uses data rates of typically about 8 Mbps, but as low as 2 Mbps for VHS quality; and the emerging MPEG-4 format, used for low bandwidth yet high-quality applications such as cell phone video or streaming broadcasts, anywhere from 64 Kbps to 4 Mbps.

Real Video
This is a proprietary format, developed and supported by progressive networks. It produces a good-quality picture, and files in this format can be played on free RealPlayer software. The quality of the image can be adjusted to take into account the desired file size, which means that the material may be compressed and made available to users with a slower connection.

AVI (Audio Video Interleave)
This is currently the most common format for audio/video data on the PC, and is also very common for audio video delivered over the internet. AVI files are supported by the Windows Media Player and RealPlayer.

QuickTime
QuickTime, originally a dominant video format for the Macintosh platform, is now a multi-format standard that supports and includes many software CODECs, including many of those mentioned in this section.

Quality assurance and quality control
The project manager must develop a rigorous quality control process
for three stages of the project: first, during the preparation of specifications for the digital-imaging workflow, defining all imaging objectives (regardless of whether the work is done in house or outsourced); second, during the key stages in the project, such as selection, handling and digital capture; and finally during the image delivery project, to evaluate download times, on-screen resolution, etc. Such quality assurance procedures will guarantee the integrity and consistency of the image acquisition process, and minimize the variation between different operators and different scanning devices. The person responsible for the scanning should not carry out the quality assurance. It will be equally important to verify any recorded information and metadata created against the information accompanying the original.

The quality of digital files is affected by many factors, each of which can be controlled independently: original material, capture device (scanner or technique used), operator, scanning resolution, dynamic range of scanner, post-processing and image manipulation, formats, compression algorithms and final display device. Ultimately, digital image quality will be a tradeoff between final image quality and the capture time required, as well as storage available and file size – that is between final image quality and cost. Higher image quality means accompanying increases in resolution and colour depth, storage and file size, all of which place a larger demand on the whole digitization chain.

Quality control procedures should address the following issues:

- Is the digital file named for the correct original object?
- Does the image include all the information in the original image?
- Does the image accurately represent the qualities of the original?
- Does the image conform to the agreed-upon file standards in the specification?

Other technical considerations to review will include: orientation; cropped and border areas, missing text, page numbers; alignment of image; size, resolution and bit depth of image; file format; details in highlights and shadows; tonal values, including brightness, contrast and sharpness.

An overall evaluation should address the quality of the digital
objects as a whole, and assess this quality against the original source materials. Is the digital copy unacceptable, adequate but of diminished quality, comparable in quality to the original or of improved quality?

It will also be necessary to assess the quality of the digital files when they are delivered to the end-user by assessing the images under different display or access conditions (e.g. on a lower resolution monitor, or files delivered via a modem).

The importance of consistency

Consistency is the key to ensuring the quality of digital files. It is necessary to develop a consistent series of processes to ensure that there are no variations in quality, regardless of different devices used for different stages of the digitization workflow, and to achieve consistent results. This can be achieved by using standard tools, such as the image quality charts produced by Kodak, and developing routines for testing the image quality of monitors or display devices, scanners, digital cameras and printers, and for calibrating the imaging equipment. For more detail on image quality, see the work of Frey (1997), which is seminal in this area. In particular Frey suggests that four targets be used for evaluating the results of digitization: tone reproduction, colour reproduction, detail and edge reproduction, and noise. Satisfactory performance in output tests for these four targets will ensure that an acceptable digital surrogate of image information has been created. This is especially important if several parties are responsible for the digitization or image capture. If a vendor is to be used, these results should be agreed in advance (CHIN, 2002).

Management of digital assets

Digital files, including images, metadata and associated materials, will often be stored in a database, image management system, or digital asset management system. There are many commercially available tools that can be used, and selection should be based on the needs of the project. For a discussion of these tools, and suggestions for some tools that can be used, see the NINCH Guide to Good Practice section on digital asset management tools (NINCH, 2002). See also the TASI
website (www.tasi.ac.uk) for a detailed discussion of image management systems and how they can be used to manage the workflow of digitization projects by providing versioning controls when several staff members are working on the same part of a project.

**Metadata**

Metadata is, simply, information about information. In the digital context, it is information that describes digital objects and enables users, both present and future, to find, manage, and use digital objects. Metadata is more than just cataloguing information; it should represent the total historic record of the digital object, and should represent the totality of information about the object, including its creators, structure, format, etc. Good metadata creation will be a key component of developing digital materials that are usable and useful for the long term, and therefore a significant investment should be made in the design of metadata schema at the project-planning stages and to ensure that accurate metadata is captured throughout the project. There is an excellent discussion on metadata for resource discovery, description and use in Deegan and Tanner’s *Digital Futures* (2002), and the reader is encouraged to consult this for a comprehensive overview of the topic.

Typically, metadata will be stored in some type of database, using a controlled vocabulary or series of keywords developed following conventional guidelines for the content, such as thesauri, classification schemes and authority term lists. Examples include the Library of Congress subject classification schema (http://lcweb.loc.gov/catdir/cps0/lcco/lcco.html), terminology from the Art and Architecture Thesaurus (AAT) and Thesaurus of Geographic Names developed by the Getty Information Institute (www.getty.edu/research/conducting_research/vocabularies/aat/index.html), the International Council of Museums International Committee for Documentation (CIDOC) standards for museums documentation (www.cidoc.icom.org). Using a controlled vocabulary will ensure that users of the digital resources are able to search and browse the resources in a consistent fashion.

A schema that has been used widely as a basic framework for many different types of metadata is the Dublin Core Metadata Element Set
MANAGING A DIGITIZATION PROJECT

... (DCMES; see http://dublincore.org/). This is a generic set of 15 elements applicable to a variety of digital object types. Dublin Core has been adapted by a number of communities, including museums, scientific ("Darwin core") and cultural heritage organizations and moving image archive repositories. Dublin Core was originally intended to facilitate the retrieval of web-based resources, but has since grown in scope to become able to describe almost any object.

The three main types of metadata are:

- descriptive metadata
- administrative/technical metadata
- structural metadata.

METS

The METS (Metadata and Encoding Transmission Standard) schema is an improved new standard for encoding the descriptive, administrative and structural metadata of digital objects, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation. For more information see www.loc.gov/standards/mets/.

For an example of a technical metadata schema, see the NISO Metadata for Images in XML (NISO MIX) project, at www.loc.gov/standards/mix).

Descriptive metadata

This is used for resource discovery and providing intellectual and physical access to the collections. Typical descriptive metadata formats are the MARC (MAchine Readable Cataloguing) format (for more information, see http://lcweb.loc.gov/marc/) and the EAD Document Type Definition (DTD), a standard for encoding archival finding aids using Standard Generalized Markup Language (SGML) and eXtensible Markup Language (XML). For more information see http://lcweb.loc.gov/ead/.
PART 2: DIGITIZING COLLECTIONS

Other descriptive metadata schema are the TEI Header, part of the TEI (Text Encoding Initiative; www.tei-c.org/Guidelines2/index.html) which is primarily concerned with describing electronic texts, and the VRA Core Categories, Version 3.0, which is a project of the Visual Resources Association Data Standards Committee. The VRA Core Categories, Version 3.0, consist of a single element set that can be applied as many times as necessary to create records to describe visual materials. See www.vraweb.org/vracore3.htm.

Administrative/technical metadata

This is information that allows a repository to manage its digital collection, e.g. date of scan, resolution, rights information. The appropriate technical and administrative metadata (such as catalogue details, copyright status, capture information, file format, resolution, etc.) will assist the project manager in maintaining and preserving the digital collection. This is particularly important with collections that contain large numbers of surrogate images created from a digital master. Each and every digital object will need its own individual metadata; the use of these tools will help manage the workflow of a digitization project.

Structural metadata

This is metadata relevant to the presentation of a digital object to users, and to the relationships of digital objects to one another and to external resources. The METAe project (the Metadata Engine project) is one initiative that is undertaking research into structural metadata (http://meta-e.uibk.ac.at).

Standards and best practices for digitization

The best thing about standards is that there are so many to choose from. (Gill and Miller, 2002)

The idea of following 'best practices' - accepted, documented methodologies for doing things efficiently and well - is long established in the library, archival and museum world, as well as other sectors, including industry. In relation to the digitization of cultural heritage materials, best practices, like their analogue antecedents, are simply working practices and procedures that can provide a safety net for practitioners as they navigate all aspects of the digitization workflow. Following these markers will give a reasonable expectation of success and add value to projects in terms of their long-term consistency, reliability and interoperability.

Developers of digital collections will have to face many difficult decisions in the course of steering a project to completion. Choices will have to be made about every aspect of the project, including the selection of materials; cataloguing, displaying, archiving and preserving digital materials; technical specifications about hardware, software and file formats; and metadata. Any decisions made will be dependent on many factors - the size of the project's budget, the number of available staff, the scope of the project, the time available and the nature of the institution. Publishers, universities, libraries, archives and government organizations will all have different perspectives and priorities for achievable practice within realistic considerations. Having a single set of canonical standards to guide the decisions that have to be made would be helpful to the project manager, and there is an increasing understanding across the cultural community of the need to apply recognized standards and guidelines to ensure the consistency, reliability and longevity of digital resources. However, technology and institutional requirements change too rapidly for any single set of standards to be applicable. The best solution, as outlined in a series of papers developed by the African Online Digital Library (www.aodl.org/bestpractices.html), is not to work towards rigid standards, but to develop sets of best practices, or current practices, around key topics. These can be then
applied or modified according to the needs of specific institutions, collections and projects.

The user will find that there are many standards and guidelines for best practice aimed at different sectors of the cultural heritage community, for example:

- The UK's New Opportunities Fund has established a set of guidelines and working practices for projects funded under its digitization initiative (www.ukoln.ac.uk/nof/support/intro.htm).
- The Arts and Humanities Data Service, another centrally funded initiative in the UK, has developed a series of Guides to Good Practice for humanities digitization projects (www.ahds.ac.uk/creating/guides/index.htm).
- The Digital Library Federation (DLF) maintains a comprehensive list of standards which it has endorsed (www.diglib.org/standards.htm).
- The Canadian Heritage Information network also maintains a list of standards (www.chin.gc.ca/English/Standards/introduction.html).
- One recent initiative to evaluate current practices and use them to develop a set of guidelines for practitioners was The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials (NINCH, 2002). NINCH, the National Initiative for a Networked Cultural Heritage, brought together an interdisciplinary working group and a research team from the University of Glasgow's Humanities Advanced Technology and Information Institute (HATII). Their research underscored that much of what we recognize as good practice has been developed at a grassroots level within the community - often through trial and error.

Examining existing guidelines will allow the extrapolation of the best and most recent standards currently available that can be modified to fit the intended purpose, institution and budget. Using community-accepted standards also allows the possibility of collaboration with other institutions in the future. Many consortia-based projects, such as the Making of America or NOF-Digitise projects, have followed shared sets
of technical standards and guidelines, and this can be yet another advantage to working within a collaborative project. The approach of funders to the use of prescribed guidelines varies greatly, although some grant-awarding bodies will mandate the use of a particular set of guidelines or practices. For example, the New Opportunities Fund requires mandatory conformance to a set of specified standards and guidelines as a condition of funding (Miller, Dawson and Perkins, 2001).

Ideally, using best practice should ensure that any investment in digitization will create the best results for the least investment by avoiding redundancy, adding value to the resources created and ensuring that investment in digitization creates content of the broadest use and appeal to multiple and diverse audiences. One way to conceptualize this is to focus on interoperability between and among source materials from different repositories or digitization projects by the use of community-appropriate and widely deployed standards. These standards include means of describing data such as MARC, Dublin Core or the TEI header, means of representing information digitally such as Unicode, JPEG and MPEG, and means of controlling data values, such as AAT or LC subject headings. Long-term access to the materials should ensure the re-use of primary materials beyond the context within which they were first digitized and encourage interdisciplinary re-use and interchange of information across disciplines, methodological boundaries and institutional types. Needs of users with disabilities should be considered and addressed (by using the W3C’s Guidelines for Web Site Accessibility or similar national guidelines – see www.w3.org/WAI/Policy/).

The implication will therefore be that users can focus on research and enquiry, rather than on the technology, that research results are replicable, and that future enquiry can build on past discoveries To achieve this, reliable, permanent resources must be available (NINCH, 2002).

Identifying good processes

Best practice may well be impossible or impractical for a particular project or institution. There is a broad spectrum of 'acceptable' practice for digitization, encompassing everything from no-frills, low-cost (or no-cost) projects aimed purely at making a collection accessible
through to nationally funded, high-profile projects with generous provisions for staff and budget. Doing things the ‘best’ way can add unacceptable costs or time constraints to the project. This diversity was recognized in a recent IFLA/UNESCO Survey on Digitization and Preservation, which noted an ‘almost complete lack of consistency in the handling of digitisation projects, from the type of material selected for digitisation, through the technical processes used, to the methods of consultation and the handling of the digitised collections’ (IFLA/UNESCO Survey, 1999).

Although the report urges practitioners to look to existing standards and guidelines, it emphasizes that such standards as exist will vary according to the type of material to be digitized and that this is still a relatively new area in which best practice continues to evolve and to be defined. Decisions will be influenced by many local conditions, priorities and considerations, as acknowledged in some pragmatic digitization principles developed at the National Library of Australia, which state that ‘Decisions about quality and quantity will be bounded by goals, priorities and available resources . . . Digitisation of collection materials can be a complex and expensive process’ (National Library of Australia, 2000).

In a recent essay discussing the Palestine Post digitization project, Ron Zweig, the project director, invoked Voltaire’s observation that ‘the best is the enemy of the good’ in describing the dilemma faced by many project developers (Zweig, 1998). Trying to achieve ‘best’ practice may mean delaying a project until technical standards are agreed upon, until large sums of money are available for digitization, or even until new technologies are invented. It might mean never digitizing, or delaying a project to the frustration of all concerned. Utilizing ‘best’ practice, or even good practice, may be counterproductive, and it may even constrain the creativity or innovation sometimes required to make things work on a budget. For example, Carnegie Mellon University’s Million Book Project was established with the goal of digitizing over 250 million books at a very low cost (see www.library.cmu.edu/Libraries/MBP_FAQ.html) and the project is actively researching cost-effective and efficient OCR methodologies in order to achieve this goal.

What can be more significant is working with the idea of ‘good
ith generous pro-
way can add unac-
This diversity was

Digitization and
of consistency in
type of material
sses used, to the
juised collections

to existing stan-
ards as exist will
d and that this is
ntinues to evolve
any local condi-
ed in some prag-
dional Library of
and quantity will
rces . . . . . Digitisa-
xpensive process

itization project, ob-
servation that
le dilemma faced
to achieve 'best'
al standards are
es for digitization,
mean never digi-
l concerned. Uti-
y be counter-
ty or innovation
et. For example,
was established
a very low cost
l) and the proj-
t OCR method-
ne idea of 'good

process' – developing a comprehensive workflow for project manage-
ment that ensures that guidelines are followed and that agreed prac-
tices are adhered to. Such guidelines will not just affect digitization –
y they will concern work methods, staffing, resource management and
all aspects of institutional infrastructure. Management will require a
focused perspective in order to implement processes to check that
work is being done to agreed standards, by regularly examining work
that is undertaken, overcoming potential barriers and evaluating out-
comes. Managers should plan to use quality assurance processes to
identify further refinements to programmes, or possible adaptations
of methodologies to suit changing needs, expectations or technolo-
gies. The implementation of documented and transparent processes
for negotiating a clear path through all aspects of the digitization life-
cycle will also be a valuable service to the diverse groups of stake-
holders and project participants who will have a vested interest, and
ongoing role, in the creation and preservation of digital materials.

The key to doing this type of work well and integrating it into the
institutional infrastructure is not by merely learning the technology,
but rather by applying the technology in specific settings, knowing
how to make good decisions when there are alternative paths to fol-
low (which is almost always the case), and working in a more con-
certed way with colleagues. Again we see that, above all, good
management practices, including developing good documentation,
are key. For example, the Forced Migration Online Project is a multi-
site, international project. To ensure consistency in all aspects of digi-
tization and delivery of electronic materials, regardless of the
location of the activity, the project has produced carefully docu-
mented guidelines for partner institutions to follow (www.forcedmi-
gration.org).

A simple, local solution can be to establish an internal standards
committee to review digitization projects and initiatives. Such a group
can keep up to date with existing standards and guidelines, ensure
that projects are not using proprietary systems and applications, and
check to see if proposed projects have already been done elsewhere.
A committee of this nature could have representatives from many
parts of the organization and therefore provide another venue for
interdisciplinary collaboration and communication.
There will ultimately be key elements of good practice that should not be compromised, and the first of these are preservation and access. Every digitization project should ensure widest possible access to the digital materials, including using metadata solutions to ensure that resources can be found by the broadest audience and search tools. They should also incorporate solutions for accessibility by the disabled and test that projects conform to W3C standards. Users can employ tools like Bobby, developed by the Center for Applied Special Technology (CAST; www.cast.org/) in the USA, to test web pages for accessibility. There should also be a clearly stated policy on long-term preservation of, and access to, digital resources – even if resources are not being created for the long term, this should be made obvious to users of the material. In many cases (such as when obtaining European Union funding), these conditions have become mandatory. A long-term commitment by the institution to preserve and maintain the materials that are created is the best way to guarantee the longevity of the project.

In this, as in many other aspects, best practices will be dictated by what an institution is capable of, and will be accomplished by the implementation of key processes by management. Choices will need to be made at each stage in the life cycle of the project, and managers must be able to make these informed choices based on an understanding of the technology involved, an understanding of prioritization of key decisions and planning tasks, and resource and skills management (National Library of Australia, 2002).

**Preservation of digital assets**

Preservation is a process to be managed, not a problem to be solved.

(Peterson, 1997)

One crucial aspect of planning is ensuring the long-term preservation and continual access to the digital materials. Hardware, software and the underlying network infrastructures change so rapidly that it is difficult to anticipate how the data of today will be viewed and accessed in the future. Anyone who has ever tried to access word-processed documents created in an early generation word-processing software such as WordStar, or to read databases created on early generations of
good practice that should not be preserved and access. %

videst possible access to the tica solutions to ensure that licensing and search tools. They

sibility by the disabled and

ists. Users can employ tools Applied Special Technology

es web pages for accessibility.

on long-term preservation resources are not being cre-

ide obvious to users of the

ng European Union fund-

uty. A long-term commit-

r materials that are

ey of the project.

actices will be dictated by

be accomplished by the

ertainment. Choices will need the project, and managers

aces based on an under-

nderstanding of prioritiza-

, and resource and skills

, 2002).

A problem to be solved.

(Peterson, 1997)

the long-term preservation

Hardware, software and

e so rapidly that it is dif-

ll be viewed and accessed

cess word-processed doc-

processing software such

on early generations of

Dbase software, will recognize that future users of digital data will have frustrations in reading data created on today’s systems. All too often, a lack of preservation strategies means that future users of digital data have to resort to extremes of ‘Digital archaeology’ in order to read files created on older operating systems or software applications (Ross and Gow, 1999). The best way to ensure long-term access to digital data is to use standard formats and open systems (such as OAKS) wherever possible, and to have a preservation and sustainability strategy for the project. See Deegan and Tanner, 2002, for a detailed discussion of the preservation question; for more information on managing preservation of digital materials, see the Preserving Access to Digital Information (PADI) project pages (www.nla.gov.au/padi/).

There are presently several technical approaches to managing long-term access to digital data. These include:

- **Refreshing.** This involves periodically moving a file from one physical storage medium to another in order to avoid the obsolescence or degradation of the storage medium – for example, moving files from 8-inch disks to 5.25-inch disks, and then from 3.25-inch disks to a network storage device. Refreshing is an ongoing process that will be necessary as long as storage media change.

- **Migration.** This approach requires periodically moving files from one encoded format to another that is more consistent with a more recent computer environment; for example, migrating datasets from Dbase to Mysql, or word-processed files from Word 5 to Word 97, then to Word 2000. Migration will gradually bring files into a narrower variety of standard, contemporary file formats.

- **Emulation.** This is similar to migration, but focuses on the applications software, rather than the files containing the data. Emulation will seek to develop new tools that will re-create the conditions under which the original data were created – by mimicking early operating systems and software applications. The objective is that under emulation, older datasets (say, databases created in Dbase 5) will run on a contemporary computer (Lawrence et al., 2000).

The long-term costs of digitization should be maintained and supported by the host institution, which should make a commitment to
the long-term maintenance of digital data at the outset of any digitization initiative. Institutions which commit to preserving information in digital format must also commit to a substantial investment in keeping up with technology. While international standards address physical formats for digital information, this is not the case for software components, including operating systems, databases for search capability, viewing software, or for logical formats where a certain level of risk management still remains (Canadian Council of Archives, 2002).

In addition to the costs of software and hardware to support preservation strategies, it is necessary to develop an economic framework for preservation activities, by developing costing strategies for long-term access and maintenance, as well as the physical costs of storing data in a long-term repository (such as the OCLC digital archive). This will involve factoring in all related costs – upfront, hidden and ongoing – but will enable institutions to develop a better understanding of long-term preservation needs that fit their institutional goals (Sanett, 2003).

The role of good metadata is crucial in helping users of digital collections to find and understand their contents. Besser (2000) outlines three ways in which metadata, if properly used, can assist in the long-term preservation of digital data:

- Metadata can help identify the work, who created it, migrated or reformatted it, and other descriptive information.
- It can provide unique identifying information and links to organizations, files and databases that may have detailed information about the digital content.
- It can describe the technical environment in which the digital files were created, including the equipment used, software, operating systems, compression schema, etc.

Metadata should be comprehensive and detail every aspect of the project, as it will be the key to the long-term discovery of resources. Unfortunately, it is often the case that staff responsible for developing digitization projects are on short-term contracts and are let go at the end of a project’s funding cycle, thus ‘orphaning’ the digital materials. Projects are normally funded only for the creation of digital content, rather than for the long-term life of the data. This makes it extremely
utset of any digitization activities. Investment in keeping up with advances in technology is for software vendors for search capabilities as for software vendors for search capabilities. A certain level of software functionality is required to support preservation frameworks for geographic database and data storage solutions. This will require ongoing understanding of long-term needs (Sanetti, 2003).

ers of digital color (2000) outlines assist in the long-term storage of digital data (in archives). The migration of assets is not straightforward. The links to organized information are complex. The digital files, software, operating systems, and the landscape of the project resources are often at odds with each other. Developing an inventory of digital materials, with digital content, makes it extremely difficult to implement any kind of long-term preservation plan, or to be aggressive about digital asset management. This is another argument for situating digitization activities as closely as possible to the core institutional activities, involving permanent members of staff:

Preservation can only be successfully managed if it is perceived as a core task throughout the institution and if preservation experts are consulted in all activities, including digitization initiatives undertaken in the name of access. It is of paramount importance that the preservation field keeps up the dialogue about the preservation of every initiative. They can bring a perspective of continuity to the discussion and make it clear that there is more to access than turning documents and images into tiffs and terabytes. (Lusenet, 1999).

Thinking strategically about digital assets from the outset is the best way to manage the preservation process.

Conclusion

An understanding of the essential attributes of original source materials, both intellectual and artefactual, and an ability to evaluate which of these attributes can be delivered to the end-user, must underpin all aspects of the digital life-cycle. Considering the uses and users of digital content will make it possible to make decisions about all aspects of the project, including quality, presentation, and long-term maintenance and access. These considerations will have to be balanced against the resources that are available for digitization (staff and money, as well the institutional framework that will support the project) in order to make decisions that are appropriate for a particular collection, institution, or situation. Judgements will also be influenced by the nature of the materials to be digitized, the required quality of the digital reproductions, and specific institutional policies and priorities. Setting goals based on these considerations will make it easier to select from a wide range of technologies and methodologies. Subsequent chapters explore these issues in relation to texts, images, audiovisual materials and rare and fragile materials, by describing practices which have been implemented and tested by project managers in a number of prominent initiatives.
Digital Editions: Scholarly Tradition in an Avant-Garde Medium

Andrew Jewell
University of Nebraska at Lincoln, ajewell2@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/libraryscience
Part of the Library and Information Science Commons, and the Other Arts and Humanities Commons

http://digitalcommons.unl.edu/libraryscience/183

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Digital Editions: Scholarly Tradition in an Avant-Garde Medium

Andrew Jewell

I come to you from the digital world. I am president of the Digital Americanists; I recently received a Digital Humanities Start-Up grant; I edit a digital archive on the life and work of Willa Cather; I am a faculty fellow at the Center for Digital Research in the Humanities; and I even have the word "digital" in my official job title: Assistant Professor of Digital Projects. I don't begin this way in order to impress you with my credentials, but as a confession: my current professional identity is absolutely entangled with the digital medium. That said, I want to confess something further: I am neck deep in the digital not because I have any particular interest in computers, but because our present—and future—academic environment is intertwined with this medium. The computer is a tool and a way to seize an opportunity to be what I really desire to be: an editor, a scholar, and teacher of literature.

In this way, I come from the digital world not really for idealistic reasons, but for circumstantial and pragmatic ones. In our current professional environment, there is a lot of energy and attention paid to the digital humanities and the dreamy new world it is ushering in, and it is becoming increasingly difficult to finance and publish large, sophisticated scholarly editions in print. Funding agencies are now demanding that editions be published in digital format, and the success of certain editorial projects in drawing in funds and attention—Ken Price’s Walt Whitman Archive, for example—suggests that future developments in the field will likely require sophisticated engagement with computers.

Much of the rhetoric surrounding the new medium, however, is misleading, as it suggests the world as we know it is being fundamentally transformed. For example, a talk given by Brett Bobley, director of the NEH’s Office of Digital Humanities, calls the presence of technology in the humanities “game-changing.” The transformational rhetoric is important to the agendas of funding agencies and university administrators who need to convince constituents of their bold visions. And, to some degree, it is true: the digital medium does
indeed transform important elements of our scholarly work. However, it is also possible to see the trend toward digital humanities as a reclamation of scholarly traditions. G. Thomas Tanselle, in his insightful foreword to *Electronic Textual Editing*, writes:

Even those engaged in textual criticism and scholarly editing have sometimes been swept along by the general euphoria and lost their sense of perspective. Their concerns, after all, are at the heart of the new developments, for what the computer offers . . . is a new way of producing and displaying visible texts. It can be of such great assistance to editors and other readers that they would be foolish not to make use of it and be excited about it. But when the excitement leads to the idea that the computer alters the ontology of texts and makes possible new kinds of reading and analysis, it has gone too far. The computer is a tool, and tools are facilitators; they may create strong breaks with the past in the methods for doing things, but they are at the service of an overriding continuity, for they do not change the issues that we have to cope with.

Tanselle’s point has been borne out in my own educational and professional experiences: my work with digital editions has simultaneously forced me to learn new technologies and established traditions. Is has been an act of learning how to put a contemporary tool to the service of an established scholarly need. In fact, it was the digital humanities that introduced me to scholarly traditions that had no visibility in my undergraduate or graduate work in literary study. Until I worked applying XML markup to Walt Whitman’s poetry manuscripts as a Graduate Research Assistant and engaged in debates about proper editorial policies, I had not been asked to confront elements of textual criticism: What is the role of authorial intention? What textual features are worthy of special editorial apparatus? What is the most effective form of annotation? How does one determine document order when leaves have become physically separated? Or, even more fundamentally: what is the most accurate transcription of this messy, handwritten document? The dominance of cultural studies and other theoretical models in the literary studies curriculum I encountered meant that work with texts and textual history was largely invisible. In fact, I’m embarrassed to say, I did not even know what a scholarly edition was until my graduate work was well under way.

Though my evidence is anecdotal, I believe that the excitement surrounding digital humanities has enabled a small surge in textual scholarship. At the University of Nebraska-Lincoln, where I work, one of the best-funded and most often-celebrated humanities initiatives is the Center for Digital Research in the Humanities. The institution expends significant resources to produce scholarly
works in a digital medium. Though these works of digital scholarship are widely varied, in most cases they involve some degree of documentary editing: transcription, markup, page scanning, proofreading, and more. The growth at UNL isn’t unique, of course: digital humanities centers are popping up around the world in different forms, funding agencies are prioritizing digital work, and University presses are looking (sometimes boldly, sometimes not) to reclaim their sagging bottom lines and sense of purpose using digital technology. In that sense, the digital medium is creating an atmosphere in which more people are engaging with textual and documentary editing; or, to put it crassly, digital technology has helped people rediscover that textual work is really cool.

All of the labor required for digitizing has meant that significant numbers of undergraduates, graduate students, library staff members, and faculty members in a variety of departments are engaging in some aspect of documentary editing. Though it would go too far to claim that each person who encounters one of these projects gets a full education in the subject, it is true that hands-on work with texts, which necessitates some level of intellectual engagement with issues of textuality, is happening broadly, and with many, especially faculty and upper-level graduate students, it is happening deeply. The act of marking up a text in Text Encoding Initiative (TEI) conformant XML requires the encoder to decide what features of the text need markup and to provide an accurate transcription. In my interactions with students who are collaborating with me on my projects, we regularly converse about such matters as proper name regularization, placement of annotation references, and identification of structural markers in nineteenth-century newspapers. I can say with certainty—and there are failed grant applications to prove it—that digitization and the cultural cache that came with it made those conversations possible. Without the draw of the digital, my students and I would not be engaged with the same editorial issues. The enthusiasm engendered by the promise of new digital models of scholarship is what drew the students and resources to these projects. Tanselle counters this enthusiasm for digital technology with a crucial reminder of what it is we are doing when we engage with texts in the digital medium: “We should be enthusiastic about the electronic future, for it will be a great boon to all who are interested in texts; but we do not lay the best groundwork for it, or welcome it in the most constructive way, if we fail to think clearly about just what it will, and what it will not, change. Procedures and routines will be different; concepts and issues will not. . . . We will be spared some drudgery and inconvenience, but we still have to confront the same issues that editors have struggled with for twenty-five hundred years.” Tanselle articulates an important point: the trend toward digital humanities is a boon for textual work, but it is not a fundamental remaking of it.

However, even if the fundamental intellectual issues are the same, the details are markedly different in the digital age. For an edition I’m working on,
the first complete, annotated edition of Willa Cather's extensive journalism, digital technology was not selected just to make it tenable in the current academic marketplace. Digital technology was selected because it made the edition better and more effective at communicating its content. These texts, for the most part, appeared once in Cather's lifetime, and that original publication exists only in the newspaper microfilm reels of the Nebraska State Historical Society. Additionally, these texts, though vibrant and highly readable to a modern audience, are choked with references to late nineteenth-century theater and popular culture, people and titles so well-known in 1894 that mere mention of the name was rhetorically adequate. With our digital edition, Kari Ronning and I can present the full texts of each of the 600 articles in an easily readable and searchable diplomatic transcription; we can provide a high quality page image of the original publication, which provides an authoritative image of the text and a glimpse into the fascinating context of the page; and we can provide thousands of annotations complete with images and, potentially, other media. The content of our edition of Cather's journalism could not exist in a print volume.

The edition of Willa Cather's journalism is only a part of the bigger digital project which I edit, the Willa Cather Archive (http://cather.unl.edu). This project is not exactly, or only, an edition. It is, more formally, what Carole Palmer calls "a new genre of scholarly production," a thematic research collection. Thematic research collections are, in Palmer's words, "digital aggregations of primary sources and related materials that support research on a theme" and are made because "[s]cholars have recognized that information technologies open up new possibilities for re-creating the basic resources of research and that computing tools can advance and transform work with those resources." It contains not just texts, but image galleries, interactive tools, and initiatives to organize communication among the community of Cather scholars. It is a project without a defined ending point that depends on collaborations with a wide range of people: undergraduates, graduate students, technical specialists, administrators, and scholars around the country. The thematic research collection is, in its most ambitious form, an attempt to digitally gather all the basic materials for one subject together in one place, to provide every reader, student, and scholar access to materials that traditionally have only been available to the privileged few that could afford to travel to archives around the world and carefully examine physically dispersed materials. Digitization can allow anyone with a web browser to see the documents only the elite have been able to see in the past.

This coexistence of a formal scholarly edition with other digitized materials under the same URL does perhaps blur for some the important distinction between “digitization” and “edition.” The popularity of mass digitizing initiatives, from library-driven digital library projects to Google Books, have proliferated shabbily edited texts in electronic form, and this also suggests a possible threat to
the careful work of the editor. For example, textual scholar Wesley Raabe has tracked the way digital versions of Stowe's *Uncle Tom's Cabin* have transmitted inadequate versions of the texts, primarily by basing the transcribed, digital text on faulty reprint editions. And the digital versions have life beyond the screen, for the easy accessibility of digital editions appears to have made them the go-to texts for new print editions. As Raabe argues, "Print and digital traditions have become intermingled, and the status accorded to print editions in citation, when compared to the suspicion toward digital texts, is to misunderstand our contemporary textual condition" (Raabe 2008). Raabe's research provides an example of textual transmission concerns with a big text-digitization operation, one without particular concern for the specific content but instead interested in generating lots of electronic texts. The failure of mass digitization projects to provide excellent texts is unsurprising, and we understand that the motivation for the digitization—the "mass"—precludes rigorous copyediting.

But for other content-focused projects, the thematic research collections, the blur between digitization and editions is more complex and subtle. For many texts on the *Willa Cather Archive*, we make no claim to scholarly edition, nor do we even use the word "edition" to describe those materials. For other parts of the site, however, we are doing a full-on scholarly edition with full apparatus. This means that users are given different reading experiences for different texts: sometimes only a digital transcription is presented, more often users get a digital transcription combined with full-color page images of the original publication, and in one section users get the transcription, the page images, and extensive annotations.

This variety may trouble some, but the *Cather Archive*, though it is based largely on a collection of texts, does not consider itself at heart to be a big "scholarly edition." Instead, it contains such editions within a broader thematic research collection. It is meant to be a meaningful site for students and scholars studying Willa Cather, and the needs of those users—and the wide variety of multimedia materials available—means that, for some materials, a scholarly edition is required, but for other materials, it is more important that we provide access to forms not readily available (for example, our collection of Cather short fiction texts is made up predominantly of digital forms of her original periodical publications, complete with the accompanying illustrations which most readers of Cather have never encountered before.) I provide this description to reflect the way digital technology is allowing an edition to coexist with other materials not traditionally wedded so closely to it. Though to some the thematic research collection appears to be new world, in many ways this profusion of forms under one URL—images, sounds, video, interactive visual tools, and texts—is simply a multiformat extension of the drive behind documentary editions. The *Cather Archive*, though it may exist in different forms, is only trying to bring the pri-
mary materials important to its subject before as many people as it can in the most intellectually responsible and appropriate way possible.

In his opening paragraph of his essay on documentary editing in the *Electronic Textual Editing* volume, Bob Rosenberg is unequivocal about the connections between digital editions and their print forebears:

The most important point to be made about any digital documentary edition is that the editors' fundamental intellectual work is unchanged. Editors must devote the profession's characteristic, meticulous attention to selection, transcription, and annotation if the resulting electronic publication is to deserve the respect given to modern microfilm and print publications. At the same time, it is abundantly clear that a digital edition presents opportunities well beyond the possibilities of film and paper.

I want to end today with some brief thoughts about what kinds of opportunities I can see with digitization, some of which will be entirely familiar, and others of which might be more unusual, but all of which I believe emerge out of the same concerns and desires that brought documentary editing into existence in the first place.

One of the most obvious benefits of digitization is the elimination of certain kinds of boundaries inherent in print volumes. In the digital environment, editors need not be so selective, but instead can contain all the texts they have the resources and moxie to produce, and they can present those texts as both searchable transcriptions and high-quality color images. In the presentation of texts, editors can choose multiple interfaces instead of just one: for example, if the text is encoded properly, one can alternate between a revision-ridden diplomatic transcription and a critical clear reading text with a click of a button. Or, one can allow users to browse edited documents chronologically or alphabetically or by any other arrangement that makes sense to the material being edited. The dynamism of the interface gives editors the chance to rid themselves of the tortured symbolic systems used in print to indicate various elements of the manuscript page and variants in different readings. Though rendering complex textual relationships is rarely straightforward, the digital environment's accessibility to color, animation, photographs, and space expands options considerably and allows us to dream of intuitive reading interfaces for our editions.

Once the texts are created, digital technology also allows readers to do more than just read them. Textual analysis gives users access to quantifiable data about the texts, information about word usage, phrase patterns, and grammatical choices. Willa Cather's readers can go to the *Cather Archive* and, thanks to Brian Pytlik Zillig's TokenX text analysis tool, gather unprecedented information
about the complete corpus of her fiction. They can see, for example, that she used the words “edit,” “document,” and “text” less than 20 times in her fiction, but used “book” or “books” hundreds of times (426 to be exact), or they can locate the most commonly used words and phrases used in sample texts. The value of these numbers will, of course, be determined by the value of the searches made and the interpretation of the numbers provided; the information does not replace interpretation, but gives the interpreters another piece of evidence to evaluate. One day, we hope to allow users to use increasingly sophisticated versions of this tool to track her language usage across time and across genres, to compare her language usage to her contemporaries, and to introduce part-of-speech analysis.

All of this, though, is simply an extension of an old motivating force: we want to give as many people as possible reliable and contextualized access to quality materials we consider important to the study of our subjects. Even the cutting-edge text analysis, though perhaps confounding for some modern literary scholars, would be recognizable to medieval monks who toiled on the first biblical concordance. In fact, the afternoon my colleague Brian Pylik Zillig showed me a recently generated list of all of the words Cather used in her fiction, I remarked, “Congratulations, Brian. You’ve just accomplished in a few minutes what some scholars used to take their entire careers to do.” The tools we now use may be more complex and sophisticated than tools used in the past, but they are still at the service of the same basic scholarly challenges.

This paper was presented at the 2008 ADE Annual Meeting in Tucson, Arizona.

Works Cited
Encoding Initiative website: http://www.tei-c.org/About/Archive_new/ETE/Preview/roenberg.xml

Preamble and Statement of Purpose

This white paper is intended as a tool for thinking through a set of pressing questions for the MLA’s Committee on Scholarly Editions (CSE) and as a contextualized expression of our current responses to those questions. These questions at bottom amount to “What is a (digital) scholarly edition?” and “How can the CSE, through its practices and guidelines, encourage excellence in (digital) scholarly editing?” These questions are of course not new, but the committee has not yet addressed them directly and formally in the context of digital editorial practice. This white paper is an attempt to do so.

The audience for this document is threefold. First, it is intended for the present and future members of the CSE, for whom it will hopefully serve as a record of current ideas and rationales and help in anchoring policy or explaining decisions to later committee members. Second, it is intended for scholarly editors who are interested in using the CSE’s guidelines and would like to know more about their intellectual background. And finally, it is intended for a wider audience that includes those who are interested in scholarly editing and editions and are curious about the evolution of the committee’s thinking.
The main focus of this document is definitional: we examine in turn the three major elements of the “digital scholarly edition” and explore their significance for the ways editions are read, used, and evaluated. In particular we consider a set of crucial features that we take to be fundamental to scholarly editing: transparency, accuracy, appropriateness of method, clear and responsible documentation, and the exercise of critical judgment in representing a full account of the textual situation at stake. We conclude by reflecting on the CSE’s own responsibilities in the light of these definitions. However, it’s worth posing at the outset a more fundamental question: What is the point of scholarly editions, as we currently understand them, in an era of mass data? One way to answer this question (anticipating some of the discussion further on in the document) is to note that a key trend in scholarly editing itself is toward the creation of an edition as a single perspective on a much-larger-scale text archive. The edition of Goethe’s *Faust* being developed at the University of Würzburg, for instance, includes a documentary archive of all *Faust*-related materials by Goethe and a critical edition that draws its data from that archive. Even standing alone, these archives constitute a resource that supports large-scale textual research; if aggregated through a mechanism like *HathiTrust* or *TAPAS*, they constitute a much larger body of material that could be used for broader cultural analysis. The scholarly edition, in other words, is in some cases being rethought in a way that involves both the micro and the macro scales.

It is also important to recognize that although some kinds of large-scale cultural research can be conducted on informationally undifferentiated resources like *Google Books*, more nuanced research (and more powerful scholarly argumentation) requires data that are more representationally detailed. For instance, scholars studying the changing revision habits of generations of American novelists would need access to data that capture revision as an explicit informational component. The preparation of this kind of data—whether we call it scholarly editing or something else—draws on the same levels of expertise and care, and the same kind of attention to the specificities of texts, as the traditional scholarly edition, albeit applied
toward new ends; this is a particularly fertile area for future collaboration among those in scholarly editing, bibliography, information studies, and the digital humanities more broadly construed.

Finally, along with the trend toward scale we are now also seeing a concomitant acknowledgment of the interdependence of micro- and macroanalysis, and we are seeing increased emphasis on approaches that enable scholars to move effectively between the two. The scholarly edition as we see it emerging here is well adapted to both kinds of work: it offers a detailed account of the data scholars need in order to make sense of a specific textual landscape, but it does so in a way that is formalized and programmatic and hence can support computational analysis at any scale. Through the use of standards like the TEI Guidelines, editions can also be studied in groups (though clearly this requires careful coordination of efforts to make the data commensurable across editions).

**Issues**

We proceed by first considering how to define the central terms (*edition*, *scholarly*, and *digital*) for purposes of this work and then considering the minimal qualities a “digital scholarly edition” ought to have, how we can enable these qualities, and what further research questions arise that might further the field.

**Edition**

Our definition of an edition begins with the idea that all editions are mediations of some kind: they are a medium through which we encounter some text or document and through which we can study it. In this sense an edition is a re-presentation, a representational apparatus, and as such it carries the responsibility not only to achieve that mediation but also to explain it: to make the apparatus visible and accessible to criticism.

To unpack this further: an edition is a systematic account of a text guided by a specific theory of what such an account should be (e.g., one that is
concerned with the genesis of a literary work, one that is concerned with the social ecology of the text, one that is concerned with contextualizing a single unpublished manuscript document). An edition is thus also a model, in the sense that it serves as an analytic surrogate for the textual landscape it describes, one that can be manipulated and queried to yield insight into its details. Although this definition sounds as if it might apply chiefly to digital editions, in fact it is also true of print: there the manipulation in question may happen through the creation of multiple views of the same data (for instance, indexes, bibliographies, concordances). In print these are necessarily represented as distinct entities, but they constitute the same kind of approach to the data.

There are a few other aspects of a potential edition that are more controversial. If we consider the facsimile edition as a kind of limit case—one that takes to one logical conclusion the idea of putting curated textual materials before the reader within a framework that permits analysis and interpretation—then we need to consider what kinds and degrees of curation constitute editing for purposes of our definition. Among other things this thought experiment encourages us to consider the role of the edition as a model of a textual space that makes its contents tractable to analysis rather than an aggregation that minimizes its mediation of those contents. Another issue is that of comprehensiveness and whether an edition needs to represent and curate all extant texts and documents. While comprehensiveness is desirable to the extent that it puts the reader in possession of a maximum amount of relevant information, there are certainly legitimate editorial situations where a focus on a single document, or on a limited subset of available documents, may be appropriate. In addition, in the digital medium an edition may in fact be a specific view of a larger set of materials.

**Scholarly Edition**

Differentiated from other types of editions, a *scholarly* edition is one that follows scholarly method and purpose, that is undertaken with professional
critical judgment and the fullest possible understanding of the relevant primary materials, and that provides clear documentary evidence of the relations and contexts of those primary materials. It is transparent and explicit in demonstrating an attention to the methods of its creation pertinent to the textual situation of its contents and evolving scholarly practice, in documenting the processes by which it was created, and in attending to the concerns of its medium or media. It is typically prepared with an audience of scholars and students in mind, although it may in fact serve a much broader audience, and it may also have pedagogical aims related to how it presents information and supports learning. The rigor of its preparation is assured through qualitative review, with attention to the application of, or critically constructive relation to, best practices; demonstrated historical knowledge and editorial method; completeness and accuracy of textual account and resultant text or texts; pertinence and utility of textual apparatus and paratext; and other factors relating to its scholarly reliability and usefulness.

A scholarly edition is clear about its commitments, and it keeps its promises. It is motivated to support further scholarship through its attention to these principles and their clear exposition, and it is understood to be part of larger scholarly enterprise, ultimately taking its place alongside and possibly in combination with similar works and allowing forms of analysis and engagement beyond those of its editorial intention, supporting further (re)mediation, (re)construction, and (re)mix in the advancement of scholarship in acts that allow, for example, the construction of other editions that may explore alternative hypotheses or challenge notions of authorial intention and editorial authority.

Digital Edition

The digital modes in which the scholarly edition of the twenty-first century is so often expressed are deeply significant, but in many cases they serve more to realize potential already inherent in our traditional understanding of the scholarly edition than to overturn that understanding. Although the
theme of innovation is common in discussions of digital scholarly editing, it is important to frame that innovation within the context of the goals and overall mission of the editorial enterprise. The digital is neither inherently a site of innovation nor a necessarily useful innovation in itself. In proposing approaches to the assessment and design of effective digital scholarly editions, this white paper therefore takes the position that the use of digital methods needs to be carefully thought through, motivated, and explained and that specific digital features need to be consistent with the scholarly goals of the edition (as articulated in the edition’s statement of method) instead of serving solely as decoration.

There are specific digital modalities that seem to us to offer particular value for scholarly editions. First, when digital editions are designed so that their textual data is captured using standards like TEI, this opens up important opportunities for alternative deployments of the data: as the basis for other editions, as transcriptions that can be compared using collation tools, as data that can be contributed to a digital repository or aggregated into a text corpus that might support quite different types of analysis. Second, the addressability of digital information (through linked data) makes it possible to support features such as user annotation, commentary, citation, and the creation of additional layers of editorial information. Third, the emphasis on writeability, which is so important to modern digital interfaces, also extends to theories of the digital edition: edition interfaces can serve as environments for manipulation and exploration of the edition’s textual space and also as environments within which the user can occupy the role of a contingent editor, examining less-traveled editorial paths and their interpretive consequences. This becomes especially important as one considers the alignment of emerging social computing principles and practices with those traditionally associated with scholarly editing, impacting traditional editorial authority through an emphasis on ongoing open editorial procedure and facilitation.

From another perspective, the digital offers not only additional ways of designing and building scholarly editions but also additional contexts for
their use and ways of understanding their pedagogical and cultural importance. Digital communication in general requires changed ideas about literacy, entailing new skills, abilities, and dispositions in front of the activities of reading, writing, and interpreting. Textual scholarship and the study of what N. Katherine Hayles and Jessica Pressman have called “comparative textual media” (vii) lie at the heart of these new literacies, which extend across an ever-expanding variety of textual, visual, and aural media.1

Supporting the Scholarly Edition in the Digital World: (Re)Considering CSE’s Mandate

Implied in the above, a key pragmatic issue for the CSE, as well as for scholarship more generally in the area, is how we choose to define the terms relating to the past and present of the digital scholarly edition and, indeed, how we choose to view extant and emerging digital scholarly editions. Many strategies of edition definition follow traditional models and understandings of edition-oriented typology, rooted in work such as that documented and exemplified in Greetham’s Textual Scholarship: An Introduction (in “Appendix II: Some Types of Scholarly Edition”). These strategies are also evinced in the elements of the CSE’s annotated bibliography of key works in the theory of textual editing (Van Hulle). Some definitions focus specifically on the digital, attempting to extend earlier traditions and to typologize digital scholarly editing trends of the past several decades in the context of current and future work (for example, Siemens et al.; Siemens). Other approaches and examples abound, some listed among materials mentioned in this document and others in and among those scholarly editions in digital form that have been submitted to the CSE for consideration toward the award of the committee’s seal that signifies an edition’s excellence.

The scope of this white paper and indeed the considerations it offers are necessarily framed by the historical moment and the place of the CSE, itself founded with the goal of “improving the state of scholarly editing and
... encouraging and identifying reliable textual work,” a mission that dovetails with larger initiatives within the profession to establish scholarly editing as an authoritative basis for scholarship (“Professional Notes” 274). The current CSE is seeking to further those larger aims at a time when our understanding of terms like scholarship and editing is under revision. This white paper thus serves as an attempt to articulate the CSE’s position in relation to that revision process. Pragmatically, we must ask ourselves what we need to know to ensure that the CSE best respond to changes in the field within the scope of its mandate. In brief: What is the frame of address to allow the CSE to best position itself, through the criteria associated with the award of its seal and its potential revision, so as to ensure CSE’s continued pertinent function in the scholarly editing community? While digital scholarly editing is an area actively engaged by scholarship at the moment, and it will likely be some years until there is relatively stable agreement in the field about this, there is still much available for us to consider now in relation to the CSE mandate.

In the first instance, to enable us to evaluate editions appropriately we will need to be able to define categories of editions in nonlimiting ways that can be embraced by the CSE and its processes as well, which should reflect the ways in which these considerations are emerging in our community. A key question for us in this regard is: How can we acknowledge the plurality and evolving nature of scholarly editions while nonetheless retaining the ability to recognize excellence and failure, at the same time that we rightly distinguish failure from improper categorization or the limitations of our CSE evaluative model? A clear answer here is that, as per our guidelines, the edition needs to include a statement of purpose that the reviewer can measure against: how appropriate were the methods? how effectively were they carried out?

A further pertinent question, however, is how well our current guidelines, guiding questions for reviewers, and other supporting resources reflect the current and anticipated needs of the future. It has been contemplated that these documents would require revision to accommodate this and other
issues. The CSE recognizes that the category of “edition” is extremely broad and sees its own mandate as encouraging and cultivating standards for excellence within that domain. However, at its inception and in its early documents the CSE adopted a fairly specific definition of the kinds of editions it would cultivate and endorse: critical editions in the tradition of Fredson Bowers. This definition tended to exclude editions of other sorts: for instance, documentary editions. The more recent CSE discussions have emphasized the need to broaden the scope of the CSE’s attention to include different editorial modalities. These discussions, however, have highlighted the need for a set of standards of excellence that can generalize well across different types of editions.

To this end, we have arrived tentatively at an initial set of minimal conditions that mark an edition—in our terms—as a scholarly edition now, extant across modalities that could not possibly have been anticipated at CSE’s inception or in some cases even a decade ago:

- it must account completely and responsibly for the textual landscape it represents;
- it must fully describe and justify its editorial methods;
- it should reveal the processes by which it was created and disseminated (including data, data structures and constraints, and algorithmic or dynamic processes), and it should include a record of changes and updates made to the edition over time, which otherwise tend to remain invisible in the digital environment;
- it should reveal the judgment and scholarship, the editorial rationales and processes, on which the edition is based;
- it should evince a rigorous standard of accuracy and consistency in applying a particular editorial approach, set of theoretical premises, or method;
- it should demonstrate the appropriate fit between stated
methodology, stated goals of the edition (e.g., reconstructing
authorial intent, reconstructing the social text, etc.), and the nature
of the existing textual witnesses;

- it should contain a detailed textual introduction or editorial policy
  statement, as distinguished from a critical introduction, that
  outlines these aspects; and

- it should include consideration of how the edition can circulate and
  function as a scholarly resource over time.

Further conditions that apply specifically to a digital scholarly edition
include, but are not limited to, the following:

- it must note its technological choices and be aware of their
  implications, ideally using technologies appropriate to the goals of
  the edition (see fit between methods and goals, above), in
  recognition of the fact that technologies and methods are
  interrelated in that no technical decisions are innocent of
  methodological implications and vice versa;

- it should be created and presented in ways ensuring greatest chance
  of longevity — a challenge the address of which involves
  infrastructural, financial, and data representation issues (such as the
  use of widely accepted, open standards);

- it should readily respond to the challenge of maintaining the
  scholarly ability to be referenced in view of the ways that interfaces
  change over time; and

- where possible, it should attend to possibilities of sampling, reuse,
  and remix, supporting approaches to the formation and curation of
  the edition such as reconstructing and documenting instances of
  texts and textual change over time, like algorithmic construction
  and reconstruction (with possible extensibility, including external
  data); in doing so, it should attempt to balance considerations for
intellectual property and labor with the goals of achieving open access and reusability.

Additional criteria may emerge with further discussion and consideration.

In closing, we wish to reiterate that the CSE remains open to and encourages the practice of a wide variety of editorial approaches, as these relate to both print and digital editions. As its primary mission, the CSE seeks to encourage excellence in scholarly editing, by which we mean above all:

- transparency with respect to data and methods
- clear articulation of motives
- persuasive rationale for the editorial approach taken
- thoroughness and accuracy
- attention to issues of usability of the edition, including questions of audience and of long-term usability

These criteria should be understood as applying equally to print and digital editions. If we take printed books to be “machines of simulation,” as Jerome McGann has recently put it (93), this vantage point may help throw into relief the extent to which both types of scholarly edition share overlapping motives, processes, and outcomes, even while we bear in mind their clear and inevitable differences in conception and execution.

For the foreseeable future, at least, there will continue to be editions that exist only or primarily in print as well as those that exist only or primarily in digital form, with the choice of editorial format responding to pragmatic, theoretical, and (in the case of editions affiliated with a university or commercial press) marketing considerations. Print editions will benefit from established practices of marketing and publicity, quality control overseen in part by a press, and proven means of distribution and long-
term preservation. At the same time, print editions lack the ability to incorporate seamlessly new discoveries after work has been finished, to make corrections, and to take advantage of the many other features of digital editions discussed in this white paper. Print editors would do well to think about the affordances of the digital, including data sharing. Significantly, even print editions now have a digital workflow. But there are also important editorial and interpretive aims that are much more feasible in print than in digital formats, as for instance McGann demonstrates in his discussion of J.C.C. Mays’s three-volume *Collected Works of Samuel Taylor Coleridge* (25–30). Building on Hans Walter Gabler’s edition of *Ulysses*, Mays’s *Coleridge* enables readers to toggle back and forth between “Reading Text” and “Variorum Text,” privileging neither through the editorial presentation or apparatus but instead encouraging readers to perceive the poems “in a permanent state of multiple vision” (115). Some materials from print editions might also be made available for nonconsumptive use by scholars in the digital humanities, in ways that would allow for text analysis and data mining, for instance. High-quality editions thus continue to be produced in both print and digital forms. The CSE will remain committed to encouraging and discerning the best practices for multiple types of editors and with multiple categories of readers in mind.

Note

1. Revised in September 2015. The original sentence read, “Textual scholarship and what N. Katherine Hayles and Jessica Pressman have called ‘comparative media studies’ (vii) lie at the heart of this kind of new literacy.”

Works Cited


Hayles, N. Katherine, and Jessica Pressman. “Introduction: Making,


*The Committee on Scholarly Editions*

*June 2015, revised September 2015*
Scholarly Editing in a Web 2.0 World

Scholarly Editing in a Web 2.0 World

Presidential Address, Association for Documentary Editing, October 16, 2009, Springfield, Ill. (Published in Documentary Editing, Vol. 31, 2010, pp. 92-103.)

Cathy Moran Hajo

How many of you remember when the World Wide Web was new? I remember being thrilled by the things I could do, the information that I could find quickly, and the ability to spread the word about our work. I also remember being unsure how the Web would change the practice of editing. Lately, the design advances and the use of Web technology often described in shorthand as Web 2.0 have made me feel that way again. I am excited about the possibilities, but uncertain about some of the underlying premises of Web 2.0 and what it might mean to the practice of scholarly editing. I am not sure that we have agreed upon the best model for digital editions in the Web 1.0 world, but editors and other scholars are already being pressed to move ahead to the next generation of tools if they want to create fundable and cutting-edge work. How will documentary editing fit into a Web 2.0 world? Will we be able to adapt our practices to the changing technology? Before we go too far along, we need to take the time to stop and think broadly about how we want to interact with the Web, with our documents, and with the public. You will hear more questions than answers here—questions that I hope get you thinking because we need to answer them, not only for ourselves and our projects, but for our profession.

What is Web 2.0?

Before we can figure out how it will impact editing, we need to grasp what Web 2.0 means. The term was coined in 2004 as a marketing pitch for a conference about the Web. Tim O’Reilly used it to sell the idea that new Web-based tools had caused a major shift in the way that people used the internet. Initially used to discuss the shift from desktop-based applications to Web-based ones, it has come to mean much more. One easy way of thinking about the difference between the old and the new is that Web 1.0 was about the consumption of both information and products, while Web 2.0 is about participation, the creation of communities, conversations, and information. Web 2.0 has come to stand for broader ideas about democratizing the Web and increasing user participation in Web sites. We all know of some of its more popular applications—Twitter, Facebook, Flickr, YouTube, and Wikipedia, of the ubiquity of blogs and wikis, all lumped under the term social networking. For many, Web 2.0 tools are “game-changers,” that challenge the way we organize, produce, and access information. They have decentralized and democratized access to media: one no longer needs a publisher, a news bureau, or a record deal to get one’s ideas, viewpoints, and creativity out to a large and growing audience. Through the Web we can interact directly with people across the block or across the world, and form vibrant
communities around common interests that would have been impossible just ten years ago. But it has its critics, among them Andrew Keen, who blames these tools for a cult of digital narcissism that places undue value on the amateurish opinions that he calls “an endless digital forest of mediocrity.” But we need to remember that some of this criticism hides a fear or reluctance to engage with a technology that many find difficult.

As I read about Web 2.0 and its possibilities, I have mixed reactions, ranging from skepticism to amazement. When it works, it harnesses the power of an engaged public to build in-depth knowledge at an incredible rate. When it does not work, it makes me feel as if I am eavesdropping on conversations in a high school bathroom, as masses of poorly spelled, self-absorbed ruminations threaten to overwhelm whatever good might be out there. How can we benefit from the best of the Web 2.0 technologies while avoiding the worst? Should documentary editing be influenced by these new tools and ways of organizing information? What happens to us if we do not get on board? I think that if we do not experiment with the underlying challenges of Web 2.0 tools, we may be left on the sidelines.

About 1.6 billion people use the Web. That is what is at stake for editions and for historical and literary documents themselves. With access to numbers that stagger the mind, any of our editions ought to be able to attract a far larger and more varied audience than we can reach with our print editions. If we fail to engage with this throng, will our work end up in a print “ghetto,” as each increasingly Web-savvy generation relies more and more heavily on online content?

Longevity vs. Accessibility

The Web 2.0 world is constructed on our computer displays as we open our personalized pages in browsers. The content of my Facebook page is very different from yours; it changes every day based upon the “friends” that I select, the interests that I acknowledge, and what those friends post on their pages. Facebook, a Web 2.0 technology, does not provide content; instead it provides a platform for its users to interact. Like other Web 2.0 tools, it is ephemeral. It may be popular for a few years but then be eclipsed by something newer, better, or just a bit cooler. So the idea of using a tool like this for our editions makes me uncomfortable. Editors come from a long tradition of preparing resources that will last for generations. We cannot conceive of publishing our documents using anything as short-lived as a blog or as ephemeral as a wiki, because there is no guarantee that these tools will be around in five years, never mind fifty. When we turned to digital publications, we sought to create the digital equivalent of the same lasting quality as our print editions.

From the start, with the formation of the Model Editions Partnership in 1995, editors have been advised to take the route that best preserves our digital editions, even though that road might be hard to travel. At the start it required expensive software to encode and display SGML files, and even with the conversion to XML it demands a familiarity with text encoding that comes naturally to very few editors. The results have been mixed, especially for history-based editions, as the complexity of the work slowed the creation of digital editions. Ten years ago, I thought that by this time there would be far more digital editions than there are. They are hard to secure funding for, difficult to produce, and oftentimes difficult for users to access. But many are well done and are as certain to migrate to the next generation of digital texts as any texts produced today. Most succeed in the effort to capture the detailed attention to texts that we value and promote.
I would not argue that we should not use XML and the Text Encoding Initiative’s descriptive schema for our editions, but I question whether this the only way that we can create digital editions. A case worth looking at is the online diary of seventeenth-century civil servant, Samuel Pepys. Using blogging software, the site is at heart a daily dose of Pepys’ life. Instead of publishing the entire diary, as a traditional edition might do, the Pepys site gives us one entry at a time, as if Pepys was blogging about his life. It started on January 1, 2003, with the entry for January 1, 1660. Pepys’ diary entry for October 16, 1666 was posted on October 16, 2009. Pepys kept his diary until 1669, which means that the site will be posting new entries until 2012. But that is not all the site provides. There are popup annotations of important people, some illustrated with portraits. There are identifications of buildings and organizations mentioned in the diary, all gathered into a searchable encyclopedia. There are also what they call “in-depth articles,” contributed by readers on more complex topics, such as the Great Fire of London. A sidebar provides additional information, providing the weather in central England, and links to Parliamentary journal entries, to letters, and to other primary sources that were created on that day. Samuel Pepys is also the first seventeenth-century tweeter. The editor summarizes each diary entry in the ubiquitous 140 characters and posts them on Twitter. You can follow Pepys on Twitter, receive e-mailed updates as his blogs are posted, or subscribe to an RSS feed to receive daily updates. The site fosters discussion groups, populated by experts and novices, where conversations about Pepys and his times flourish. Users make comments about the diary entries, showing a clear engagement with the texts and the historical period.

Simply put, the site is fun to visit. It offers a richly annotated text, and gives us a daily dose of a life very different from our own. It is easy to use and very accessible to the non-expert. The day-by-day release creates a dramatic tension difficult to reproduce in print when one can just skip ahead to the next page to see what happens. As Pepys recounts the Great Fire of London, his efforts to protect his property, tales of loss and bravery, and rumors of a French plot behind it all, the event becomes quite real and personal, doing what documentary editions do best.

The Pepys editor, Phil Gylford, is a Web site designer and developer. He is not a member of our Association and has no training as a scholarly editor. What he has done, and done well, is to take a resource well known to scholars, and re-purpose it, drawing attention to it in a way that has attracted a large following, far larger I imagine, than would have attended yet another published edition of the diary. But where did he get the text? Well, we find that Gylford used a published edition, in this case, the 1893 version edited by Henry B. Wheatley. Wheatley's edition was in the public domain, and Gylford did not even need to transcribe the edition himself—someone had already done so and posted the text on Project Gutenberg. I am sure that some of you would ask, “Why did Gylford not use the scholarly edition of the diary edited by Robert Latham and William Matthews?” We will return to that later.

Gylford linked a number of annotations to the entries. Those published in the Wheatley edition take the form of popups that you can access by mousing over the text. That is not necessarily Web 2.0. But Gylford also enabled comments on the entries, both to give himself a space to add notes or corrections, and to let his readers comment as well. Called annotations on the site, they include speculation, reactions to the story and scenes described, and research added by readers of varying expertise levels.

So, what can we learn from the Pepys Diary that we could adapt to our editions? It starts with the text. Gylford would not have been able to create his site were it not for the heavy lifting already done by Wheatley. Obviously
he chose the Wheatley edition over the more complete Latham and Mathews edition because he did not have to secure copyright or permissions from its editors and publishers. The digitized text was also freely available because it was mounted on Project Gutenberg. Gylford acknowledges that the Latham and Matthews edition is more complete, suggesting that readers wanting more should consult it. Gylford's site is not the only one on the Web using the same text; what sets his apart is his creativity in matching the software tool, the blog in this case, to the source material that made the diary come alive. He was not concerned about making his blog last forever. The Pepys diary has been edited and published many times, and his goal was to make it available. One would think that Latham and Matthews had the same goal, and it is disheartening when we find that the acknowledged best version of the text languishes in footnotes and in a select set of libraries, while an older text, less ably created and lacking in completeness, gets far more use. I do not know if there are any plans to digitize the Latham and Matthews edition, which is still in print, but it would seem that the best solution for all might have been a collaboration between Gylford and the scholarly editing team. While editors are undoubtedly the experts when it comes to creating print editions, we are not always the ones best suited for presenting them on the Web.

There seems to be a growing split in the digital humanities, with one branch preferring the creation of highly complicated digital texts, usually using XML to record the intricate details of the text’s creation and meaning. These texts are usually designed for scholars and advanced students to stand as digital versions of our print editions. The other branch is less interested in preparing such complex documents and more interested in producing digital texts more quickly and encouraging more people to use them. The Center for History and New Media at George Mason University offers an example of this kind of work. They create tools like the Omeka content management system to help organize and publish primary sources as Web-based archives and exhibits without becoming bogged down in detailed XML encoding. Which way should editions go? Both branches have their good points, and while scholarly editors generally favor the first, I think we should try our hand at both, even if that means that some of the digital products we create may be ephemeral. The creation of the digital transcription of an historical document, whether as a Word document, a blog post, an HTML page, an XML encoded text, or a simple ascii file, is the work we are best qualified to do. Making that base text as good as it can be by proofreading it and researching is the work that takes the most skill and time. Once we have that, there is no reason to have to settle on only one form of digital publication. Yes, there might be issues with migrating editions published as a blog or a wiki, but if the content is valuable enough, people will find a way to do it.

The Power of 1.6 Billion People

One point six billion people actively use the World Wide Web around the world, with more logging on each year. As editors, our goal has always been to preserve and disseminate important historical and literary documents, using the most appropriate tools available. For most of our lives, that tool was the book, and in some cases microform. It simply is no longer the case in 2009. We count ourselves fortunate to sell 1,000 copies of one of our volumes, but over 20 million people have seen just one 1987 video clip of pop star Rick Astley singing “Never Gonna Give You Up” on YouTube. If you have not been “Rickrolled,” you do not know nearly enough twelve year olds! If you have seen it, it probably means that you were tricked into clicking a link that you thought was one thing, but was instead a cheesy MTV video. Something is wrong here. I am not advocating trying to trick people to view our editions, but do we not need at least to try to get a tiny portion of this audience? They are out there—and they clearly have nothing better to do!
Most of our volumes are bought by research libraries, where serious scholars consult them. We do not know how often they are used and have only anecdotal feedback on how they are used. Did we get reviewed, and if so, was it favorable? What libraries purchased the volume? Do scholars use our books in their footnotes? Who contacts the project’s Web site? While this kind of feedback can help us tweak our editions, it rarely causes us to revise our editorial principles or fine-tune selection policies. It is feedback that is slow in coming, and because of the long lead time for publishing volumes, it is equally slow for changes to appear in print. Unless a second edition is published, we cannot even correct the errors found in our volumes. Don’t get me wrong, I like a hardcover book as well as the next editor. My pride in holding our first volume, never mind the sudden interest and excitement of friends and family, was so much greater than when a carton of microfilm was shipped to our offices. We have emotional attachments to books; we respect them more than articles, Web sites, or sheaves of microfiche. But if we acknowledge that our main purpose is to bring our documents to the greater public, the book can no longer be our pre-eminent form of publication. It is not the means by which we can reach a billion people.

We are in the midst of a transition over the control of media and publication and we do not yet know how things will play out. But can we wait around until things are hashed out, until publishers, especially the university presses that publish our editions, figure out how they want to deal with the Web? Are there other options to traditional scholarly publication? How will editors deal with the push for open access, the Web 2.0 imperative to make all materials free and accessible on the Web? As Robert Darnton wrote this past February, “To digitize collections and sell the product in ways that fail to guarantee wide access would . . . turn the Internet into an instrument for privatizing knowledge that belongs in the public sphere.” Yet, what about our publishers? What about royalties for the work that we produce? An esteemed colleague characterized these calls for free access as coming from a “kumbaya generation,” who generally work from secure posts at well-endowed institutions. They rarely address the costs of producing and maintaining these works. Should we give away our editions in order to encourage greater use? Some would argue that we can and should, and I tend to favor that camp. But others argue, equally cogently, that it is not possible to break even doing this.

It is a dilemma. We want high-quality editions, which require effort and attention, but we also want to reach the broadest audience. That audience is fickle, lazy, and cheap. Materials that are not free or are not digitized are often not used. Materials of lesser quality that are out of copyright are often preferred over higher-quality sources that are not free and easy to use. The online encyclopedia Wikipedia offers a good example of this principle. In its initial form, known as Nupedia, the idea was to create a monitored wiki-based encyclopedia, where scholars and experts would manage various subjects, selecting topics, writing articles, and vetting them before they were posted. Have you ever heard of Nupedia? It didn’t work. Participation was sluggish, and the few pages that were created were not so much better than what was already available. But the creators of Nupedia took a risk with a new version, called Wikipedia, that allowed anyone to post or edit an article, with no gatekeepers or scholarly supervision. They rely on Wikipedia’s users themselves to check, improve, and police the work. Who would have thought this would work? Sometimes trying something—even if you do not know what will happen—can result in unexpected and even runaway success. There is a cadre of people who dedicate huge amounts of time to crafting Wikipedia with no payment or even a byline; people who have become intensely protective of the site and are willing to work hard to keep it going. A far larger group of people—myself included—have contributed a single article, or made a few corrections here and there. What makes this kind of participation so powerful is the number of people who access Wikipedia’s pages. When you have so many
millions visiting, you only need a tiny percentage to participate in order to secure an army of editors who have built a three-million-page up-to-date encyclopedia in less than nine years. Wikipedia is among the top ten Web sites in terms of use, friendly and inviting to its contributors, and a successful example of Web 2.0 technology changing how we gather and report information.

But there is a flip side. People use Wikipedia because it is free and because it is easy, even if they admit that it might contain errors, some honest, others more malicious. Web 2.0 culture has embraced the idea that accessibility is king, that it is more important than accuracy. Listen to Paul Graham, a programmer and essayist:

Experts have given Wikipedia middling reviews, but they miss the critical point: it’s good enough. And it’s free, which means people actually read it. On the Web, articles you have to pay for might as well not exist. Even if you were willing to pay to read them yourself, you can’t link to them. They’re not part of the conversation.

The phrase “good enough,” I am sure stiffens your back. It tightens my jaw. Good enough is not what editors do. We will craft new editions of previously edited works, like the Wheatley edition of Pepys, specifically because it was not “good enough.” Whether it has inaccurate transcriptions, incomplete or subjective selection policies, or poor annotation, editors demand quality, not only for the scholars that are their main audiences, but for everyone who reads an historical document. Little errors matter and can build into misinterpretations and greater error. We stand as authorities on our subjects and take very seriously the work that we do. But authority is one thing that the Web 2.0 challenges when it states that “good enough” is good enough. We cannot let this notion go unchallenged, but it will take some doing to prove to this generation of Internet users that just because it appears on the Internet seventeen hundred times, does not mean it is correct or, more seriously, that it is not malicious and vicious. This is the risk of keeping quality sources behind subscription-based portals.

If we could build a Web 2.0 edition, with a different relationship to our readers, what might it look like? What could we do with the participation of a hundred thousand people? A traditional Web 1.0 edition might measure every time a page was opened and perhaps include an e-mail address where the reader might send feedback. We could run Google Analytics to get a sense of what pages were opened most and where geographically our users came from. Even those small advances provide us more than we know about the uses of our paper publications. But how would we deal with the “holy grail” of Web 2.0, user-created content? That is something that we have shied away from. Part of this is due to our reluctance to yield our role as experts. Do we want to turn our readers into collaborators? How would we do it? We do not know how it would work, or if it would work, but for a minute let us try not to scoff at the idea that by involving our readers we might actually change the way that we edit documents. I am sure it will be challenging, I will bet it will be nerve-wracking at times. Could it be that we fear learning how much, or how little our readers think about our editions?

I do not for a moment believe that there is someone in the vast digital wasteland who can read the letters of Thomas Jefferson better than Barbara Oberg and Jeff Looney’s teams; nor do I think that there are armies of armchair historians who can interpret Thomas Edison’s scrawled diagrams better than the Edison Papers staff. I certainly do not believe that document selection or interpretation of Margaret Sanger’s writings on eugenics should be left to activists engaged in the highly charged abortion debate. We have worked hard to gain the expertise and insights needed to edit our documents, trained as scholarly editors, and our immersion into the lives and times of our subjects takes time and talent. But that does not mean that we cannot find roles for our
users that could enhance our documents and editions.

The roles that our readers could take would vary from project to project and might include data gathering. We might start by making available some of our research files. If I mounted a version of the chronology database that we use to track Margaret Sanger’s life and allowed people to edit and contribute to it, what do you think would happen? I could envision a hot mess of Sanger haters foaming at the mouth in all the commenting areas, as they do on a number of anti-abortion blogs. But say that we manage, somehow, to keep most of the crazies out and enforce some rudimentary decorum. People are interested in Sanger and her role in the reproductive rights movement and they are also interested in local history. If we could tap into that, perhaps focusing on students of history, women’s history, or public policy, we could encourage readers to scour their local archives and libraries for additional documentation of Sanger’s travels around the country and the world. We would not ask them to build it from scratch—we would provide the dates, as best we have them, and what we know already. For example, based on our chronology database, I know that Sanger was in Illinois at least 53 days between 1916 and 1957. Could we persuade people to survey newspapers, check for photographs and ephemera in local archives, or even identify some of the places where Sanger stayed and spoke and plot them on Google maps? I do not know if they would come, but the interest and enthusiasm that we find among students, interns, and archivists when we tell them about Sanger’s involvement in their own city suggests that some might. It could also expand beyond Sanger. If readers were interested in posting events and timelines of local birth-control activism, the site could become a powerful resource bringing together materials in ways that could help us better understand the birth control movement.

Another way to use social networking is to have users rank or tag the contents of the Web site, and to use the information created to build better searches and to learn about the users. A digital edition might enable a ranking system for documents, much as we might do for Netflix movies or products that we purchased at Amazon.com. As more people provided rankings, software could use them to produce more useful search results. A five-star document would come to the top of the list, while a one star document would remain at the bottom. Users could add subject terms to documents, especially those that have been digitized in image format only, where text searching is not an option. Not only would this kind of social tagging help to search the edition, but it would help editors to understand our audiences better. We could learn about what people are looking for when they use our editions, which documents appeal most to them. I have always wondered whether users prefer to see the image of a document or a transcription, and wondered about whether advanced scholars actually do look at the image when conducting research. Are documents with difficult handwriting less likely to be used than those that are read more easily? How often do people consult the transcription guidelines when using an edition? Do they tend to prefer “sound-bite” documents that contain a short, strongly worded quote over longer more reasoned treatments of a subject? In which subjects are they most interested? Which document in our edition is most popular? By seeking feedback from our readers, we could learn far more about how they see the documents and annotations, in ways that are not possible with books. We could fine tune this data by gathering demographic information about our users as well. Armed with this knowledge, I am certain that we would learn important things. With such knowledge we might change the way we plan and create editions.

What, you might ask, makes people volunteer their time to contribute to such experimental sites? I believe it is the same thing that drives us: the appeal of our subjects, of working with these rich historical sources, and the
feeling that one is contributing to something bigger. We study fascinating people who did interesting and important work, and the lure of participating in some way, whether as an intern, volunteer, or user on a Web site, will attract people if we invite them in. The students and interns who work on our projects get as much as they give, in terms of seeing the past in a different light. This immersion in the past personalizes history and historical actors in a way that other treatments, even biography, do not approach. You can tell from the kinds of questions our students ask that they are connecting to the subject in a different manner than they would by reading a textbook or watching a documentary.

**Getting Dirty**

One thing that the Internet teaches is that trying something new, even if it fails, is in most cases better than waiting for the perfect opportunity to come along. Every time we try something, we learn—whether we succeed or fail—and in many ways the risk of trying something new on the Web is smaller than it is using traditional publishing media. Yes, we will trip up sometimes, and yes, we might get dirty. But if we do not take chances and try out new and interesting ways of publishing documents, we will not be able to re-invent the edition for a new generation. In general, editors, myself included, have been too cautious about digitization. We are so concerned about selecting the perfect system for publishing our documents online, capturing all of their complexity, enabling other editions to be searched along with ours, and employing the best standards and chances of longevity, that we are almost afraid to act. Our funders are caught in the same trap. Both the Model Editions Partnership and the University of Virginia Press’s Rotunda digital imprint have been designed with the best of these goals, but while we have been trying to solve the questions asked of digital publishing in the 1990s and early 2000s, new questions have come up. In many senses those systems replicate the experience of using our editions in print format, with some nice searching added in. But they do not address the challenges of Web 2.0 technologies. They do not invite our users in or make their experiences with the documents a part of our edition. Taking care is important, but sometimes we move too slowly and miss out on the opportunities that new and changing media offer.

Is this something that we want to do? We will not know if we do not try it. Looking at Pepys again, I would say that there is much to admire there. If it had failed, if no one had read it, no one had commented on it or forwarded the link to their friends and colleagues, how much would Phil Gylford have lost? Some time, but that is about it. I am not suggesting that we abandon the work that we have traditionally published as print volumes, but I am suggesting that we can do more things with that work and with the research files that we all have in our offices. We publish only a tiny part of the knowledge and expertise that we gather in our work and it is time to take some chances, to try new things, and to risk some investment of time, against the chance that we can make a connection with the biggest audience that any of us will address. Because we do not have to get it right every time, but if we wait and wait to create the perfect digital edition that can meet every standard out there, we may find ourselves passed by.

We may fail, but even in that we will have learned more about how people use documents online. If even one crazy idea succeeds, it could change the playing field for scholarly editions.

---

1I would like to thank Esther Katz for her comments on an early draft of this speech, and for arguing with me
about some of these ideas; it helped me enormously to sharpen this version. I would also like to thank Amanda French, the Digital Curriculum Specialist at NYU’s Archives and Public History Program, whose knowledge of and enthusiasm for these new technologies first started me thinking, and whose comments also helped enrich my speech. Lastly, thanks to my brother Mike Moran, author of several Web marketing books, including Do It Wrong Quickly: How the Web Changes the Old Marketing Rules (New York: IBM Press, 2008), who found a surprising number of parallels between the situation facing documentary editors and those faced by businesses in adapting to new forms of technology.


Andrew Keen, The Cult of the Amateur: how blogs, MySpace, YouTube, and the rest of today’s user-generated media are destroying our economy, our culture, and our values (New York: Doubleday, 2008), p. 3.


The site can be found at http://www.pepysdiary.com/. The Twitter site is: http://twitter.com/SamuelPepys.

See the entry for Wednesday, 5 Sept. 1666 (http://www.pepysdiary.com/archive/1666/09/05/).

For the Project Gutenberg text, see http://digital.library.upenn.edu/webbin/gutbook/lookup?num=4200.


Thanks to Esther Katz for introducing me to this Internet phenomenon known as rickrolling. For those who have not seen the clip (and want to): http://www.youtube.com/watch?v=Yu_moia-oVI.


For a sleepless night, read Shirky’s *Here Comes Everybody: The Power of Organizing Without Organizations*, in which he describes the anti-expert quality of Web 2.0 as “about what happens when people are given the tools to do things together, without needing traditional organizational structures.” See his blog as well: http://www.herecomeseverybody.org/.

For an interesting take on seeking user collaboration in museum settings, see Nina Simon’s *Museum 2.0* blog entry, “Self Expression is Overrated: Better Constraints Make Better Participatory Experiences,” Mar. 16, 2009. (http://museumtwo.blogspot.com/2009/03/self-expression-is-over-rated-better.html). Simon argues that only a small percentage of museum goers are interested in reacting to exhibits by writing open-ended comments; far more participate when the collaboration is clearly defined and explained.
The Sustainability of the Scholarly Edition in a Digital World

Scholarly editors have long been invested in the creation of long-lasting and sustainable publications. Whether they create complex multi-year projects that rely on cooperative teamwork or develop short-term solo projects, editors understand that their work will be consulted for years to come. The expense of locating, selecting, transcribing, annotating and publishing historical documents could not be maintained if these editions were not built to last. Editors have developed practices and policies to ensure that their readers can confidently rely upon their versions of important historical manuscripts. This care has always extended to the stable publication formats editors chose, whether in letterpress or microform. When editors turn to digital publication, sustainability remains of critical importance.

Scholarly editions are also committed to bringing primary sources to broader audiences. Editors take Thomas Jefferson to heart when he wrote: “...let us save what remains: not by vaults and locks which fence them from the public eye and use, in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.” By publishing edited works, we take fragile and unique archival manuscripts and make them available in research libraries where scholars and students can access them. They don’t reach everyone. These editions are expensive to purchase and most libraries do not carry complete sets of all of them, but given the technology of the time, printed volumes certainly served to disseminate and preserve historically important materials. The advent of digital publication challenges editors to expand their reach, to move beyond the ivory towers of research libraries to high schools, town libraries and even to the comfort of private homes. This monumental expansion of our audience forces us to rethink how we edit documents and how a truly accessible edition should behave.

Editors have always had to balance costs against their ability to preserve and disseminate documents. Never has this been a more difficult task than at present, when the cost of creating long-term quality digital editions often prohibits editors from offering them freely to the public. Digital publishing permits unprecedented access possibilities, but it is a fragile medium that is susceptible to obsolescence. Should we create an edition that is as
sustainable as digital text can be, but might not be widely accessible, or should we create a widely accessible edition that might not last a long time? Neither option is acceptable. The reality of funding for scholarly editing in the 21st century is this: it is difficult enough to raise the funds to create the content of these editions. Adding the technical specialization needed to render these texts in well-formed XML is beyond the capabilities of many editing projects. Some editors, too, are reluctant to embrace the notion of providing free and open access to their editions and support the idea of subscription-based for digital editions. Arguing that their print volumes were never free of charge, many editors seek to generate income to help offset project costs or seek royalties for their intellectual work. Federal agencies prefer, or in some instances demand, that the editions they fund be produced using XML, but they do not provide sufficient guidance or tools to help editors comply. After fifteen years of being told that markup is the gold standard for digital publications, we get it. We know that we need to use XML on our digital texts, but I would say that the best adjective to describe our adoption of XML might be “reluctant.” This is especially true of historical editors as opposed to literary editors. Few historical editors have embraced digital publication’s promise, and most have treated digital publication as an add-on to their central work of publishing volumes.

Aside from our goals of accessibility and sustainability, digital publication questions some of editing’s guiding premises. I don’t believe that many editors, even those working with digital text, have really explored or thought through these issues. To date, most digital editions still closely resemble the books they are based upon. As we finish digitizing older publications and begin to construct more and more born-digital editions, we have an opportunity to redefine the edition, the editor and the editing project. While none of us can peer into a crystal ball and know what is yet to come, I expect that the edition of the 21st century will differ substantially from the editions of the 20th century.

The transcription is central to the work of the scholarly editor. Faced with the problem of how to make primary sources useable and accessible in a print world, editors decided to transcribe them. Publishing paper-based facsimiles was too expensive and while microfilm editions offered images, they could not easily include annotation or much editorial intervention. To render all the complexities of their texts, editors developed typographical mechanisms to record changes of hand or pen, additions, insertions, deletions and margin notes. Their familiarity with the texts enabled them to read and transcribe difficult handwriting, account for variants of a text and provide historical context that makes the manuscript come alive. The transcription is the text for most editions, once it is proofread or verified as accurate, it is the center around which the project orbits and if transcriptions are questioned or found wanting, the edition is quickly discredited and disused. As the capability to provide high quality images over the internet has increased dramatically, editors need to grapple with the idea that transcription might have a different role as we move into the future.

Another agreed upon principle of scholarly editing is that the editor’s work should be as unobtrusive and objective as possible, which in part contributes to the edition’s long shelf-life. Despite the fact that selecting texts and drafting annotation can be intensely subjective tasks, editors try to keep interpretation and historical arguments to a minimum. Our role is to present the most important documents with factual annotation that enables the reader to understand the text and reach his or her own conclusions. More often than not, the editor who feels compelled to weigh in on current historiography or the controversial issues of the day does so outside the edition, in a journal article, monograph or biography. At projects that employ more than one editor, the work of each is subsumed into one consistent editorial voice. We credit editors on the title page, not each
portion of the edition that the scholar created. In this way editions are truly collaborative and a rare example of
team-based humanities research. As digital scholarship allows up to build larger and more far-flung
collaborations, editors need to decide whether that current mode of attribution should be continued.

Finally, editors agree that because their work is so labor intensive, it must be done well. It will be a long time
before anyone has the opportunity to do the work again. The existence of a previous published edition on a topic
forces the editor to explain why the existing publication is so flawed that the time and money needs to be
expended to redo it. One of the promises of the digital age is the ease of re-purposing objects once they have
been digitized. When editors can continually revise and enhance their editions will these projects ever truly end?
Will we develop mechanisms by which we can allow other scholars to try different approaches on our texts? Will
they need our permission?

Editions are not created in a vacuum. We rely on more than sixty years of tradition and previous work to guide
our steps. But not all our inspiration comes from the print editions of the past. We are not just content creators,
but users of digital resources. We have seen the work done by archivists, public historians, and digital historians
to digitize primary sources and to bring them to new and larger audiences. We also see new ways of researching
and collaborating with both experts and the public and we want to try these out in our editions. Editors need to
decide how many of our the tried and true methods still remain valuable, and what aspects of the new
technology would be best incorporated into our editions. Digital historians seek to do more than just enable
searches of electronically rendered texts, they want to encourage research and collaboration, to use data mining,
visualizations, and other computer-aided tools to analyze texts on a much larger scale than once possible.
Sometimes I am not sure that we recognize the power of conducting a simple Google Book search for a string of
text—we are searching more than ten million books in a few seconds, a task unthinkable before digital
publication. This kind of computing power can fundamentally change the way we do research as well as how we
formulate a feasible research question. The problem that historians will face from this time onward is how to
deal with an abundance of sources, not how to overcome a dearth. So any method that we can develop to help to
dice up our editions into smaller useful portions will be valuable. While digital historians are only small
minority of the profession at present, in just a generation or two, all historians will be digital historians. Editors
need to keep an eye on the trends in this emerging field, to educate themselves in the technology, and learn to
adapt their editions to meet the needs of this important group of stakeholders.

Scholarly editors have been hearing about the benefits of markup XML for fifteen years but that does not mean
that most are comfortable with the idea. Again, I am talking primarily about editors of historical documents,
rather than those working on literature. English departments and linguistic scholars adopted markup languages
and computer aided text analysis early and have led the way in the development of XML, especially the Text
Encoding Initiative guidelines that focuses on humanities texts. Digital editions like the Women Writers Project,
the Whitman Archive, Willa Cather Archive and the William Blake Archive have benefitted from close
associations with humanities computing centers to create sites that rely on robust encoding of complex
manuscript material.

Historians have not been as quick to embrace markup language and I am not entirely certain why. We are more
reluctant to put the time into mastering XML or the programming languages needed to search and display XML
texts effectively. We balk at the amount of technology we need to learn, seeing time spent on it as time spent
away from our documents and our research. It may also be that XML provides greater immediate benefit for literary research than it does for historical research. The Willa Cather Archive offers text visualization and analysis tools that enable readers to search for word frequencies, create word clouds and concordances. These are useful tools, to be sure, but not the first ones that a historian might apply to a set of documents. We might prefer to locate all the mentions of a person, organization, place, subject or idea. We want to find the appropriate documents and study them through close reading. I think that at some level we resist the idea that a computer could ape the way that we attack documents. If our print editions are any clue, we invest the most time in creating detailed subject indexes that organize the documents into important categories. While historians are interested in the process of document creation, tracing variants and versions over time, they are more interested in assigning content to the text. Encoding these contextual relationships are a sort of cross between annotation and indexing, something that might be done slightly differently by each project depending on the interests of an editor or the specific documents being published. Believing that each of their project’s needs and processes to be unique, historical editors have not really united to develop the digital tools that would lead to new ways of looking at texts, either within or across editions.

We don’t have a good model in use today for a sustainable XML edition with which we could develop a shared conception of digital editing. There are a lot of silo projects that have been developed for specific sets of documents and that are not broad-minded enough to serve all editions. I think that a lot of these idiosyncratic editions have good ideas, but often do not have sufficient infrastructure to ensure longevity.

The XML-based edition we are building at the Margaret Sanger Papers is an example of such a specific application. The Project’s focus is on Margaret Sanger, the 20th century American birth control activist. Our entire archive consists of slightly under 100,000 documents, published on microfilm. Only about one percent of these documents were selected, transcribed and annotated for our four-volume print edition. For our digital edition, we did not want to repeat the work that we did with our book edition. Instead we wanted to explore the capabilities of XML encoding, and text searching. We selected Sanger’s articles and speeches, six hundred texts in all, as the best test for digital publication. Few of these documents were included in our book edition because of their length and the fact that they were somewhat repetitive—the vast majority dealing with some aspect of birth control. Even though all of them were accessible within the microfilm, they were not searchable in that format. We believed that these documents would best benefit from the ability to search text and the addition of metadata. We began this project in 2003, but without dedicated funding for it, progress has had to be slow. The beta site contains about three hundred documents now, searchable by text, date, title, format, publication venue, and by subject. We worked with staff at New York University’s Humanities Computing Group to set up the initial search and display and as we complete the edition we will add additional searches for the names of organizations, people, places, titles of books, and text quoted by Sanger. All these are already tagged in the texts. Our XML encoding is not as complex as many others I have seen, but we feel that the combined search capabilities of text and metadata offers something different than either our book or microfilm editions. It does not offer annotation in its classic form, but allows in-depth research of Sanger’s ideas, audiences, and changes in her arguments over time.

Ours is a concise digital edition drawn from a closely connected group of texts that share a common format and purpose. The content encoding that we have done required an understanding of the texts, knowledge of the references made, and the ability to construct detailed subject entries that provide meaningful intellectual
divisions between six hundred documents that might all fall under a handful of Library of Congress Subject Headings. If we wanted to expand this edition to include correspondence or other kinds of documents, we would have to revise our tagging scheme as well as our display interface. We could not use a system like this to digitize all one hundred thousand documents in our microfilm edition because it would be too time consuming. We do not have the staff to carry out transcription and content encoding and index entries for such a large number of documents. It is not even easy for us to make small adjustments to the edition’s design because the editorial staff did not create the programs that search and display the texts.

Our situation is similar to that of many other projects who don’t have access to digital text experts at a humanities text center. Our original project was encoded using a slightly amended version of the Model Editions Partnership DTD for TEI P4, created with the help of Matthew Zimmerman of NYU’s Humanities Computing Group. Sustainability has become a concern for us. Since our edition began, the TEI has introduced P5 and NYU dissolved its Humanities Computing group. What we have right now works, but we need to decide whether to spend the time and money that it will take to update our encoding to comply with TEI P5 or stay with the older version. If we do choose to migrate to P5, we would need to redevelop our encoding policies, resolve the differences in the texts already encoded, and recreate the search and display interfaces to work with the new encoding. While these tasks might not be difficult for those who work with TEI, XSLT and PHP every day, for us it will require either raising the funds to hire a programmer, or spending many, many nights laboriously working our way through web tutorials and a small library of the Complete Idiot Guides!

These kinds of predicaments tempt editors like us, pressed for time, to consider farming out XML encoding to a consultant or their publisher. There can be a danger in going that way because I don’t think that anyone knows better than we do how people can use our texts. If we don’t master XML encoding we won’t be able to participate fully in decisions made on how the texts should be tagged, nor will we be able to fully explore the possibilities of digital editing. And I think that we need to always be thinking about how to make better editions, even if it means breaking with some of our traditions. For example, I think that when we convert a print edition to digital form, we should be describing the manuscripts, not trying to replicate the organization or structure of a specific volume. When we cling to our older formats, I think that we limit the possibilities for redefining the way we do things with an eye to the capabilities of digital publishing.

A case in point is the University of Virginia’s Rotunda digital imprint. Rotunda is the fastest growing source for digital historical editions, best known for its ambitious project to digitize the massive multi-volume Founding Era projects, including both volumes previously published in paper and those still to be published. The Founding Fathers problem offered a “perfect storm” of a test for digital publication. Their shared geographical, chronological, and subject focus provides a strenuous test of how XML can integrate research not only across volumes, but across editions as well. The Founders had hundreds of print volumes, only some of which were in any kind of digital form, which meant embarking on a large-scale legacy conversion project, while at the same time creating a workable platform for the continued production of volumes. Finally, it offered the challenge of developing a way to combine the editor’s main access tool—the index—across large collections of volumes and editions. To date, the combined American Founding Era collection includes six editing projects, almost 90,000 discrete documents and almost 850,000 index references.

Putting the American Founding Era Collection online is a massive undertaking, but one that doesn’t really serve
as a model for developing new digital editions. Perhaps because it was led by an academic press, the Collection seems wedded to the idea that readers want to see a digital version of the old print edition. Yes, it provides text searching across the entire collection, but the organizing principle is not the document, but edition and the volume. Granted, merging the work of so many different editorial projects is no simple task, as even slight differences in editorial approach or transcription styles can result in unexpected cross-edition search results. Despite its lush appearance and useful hyperlinks to texts mentioned within each edition, this digital publication still feels very much like using a book, or a series of books, rather than an integrated digital collection. Some examples:

- Each edition is organized by the original print series and volume. Documents retain the original print volume’s page breaks, rather than that of the manuscripts they describe.
- Because of the topical overlap in editions, the same document can appear in more than one edition — for example a letter written by George Washington to James Madison will appear both in the Washington Papers and the Madison edition. Though each edition includes its own internal hyperlinks to related texts, there is no link to take the readers between the two versions of an identical text.
- The text searches that tie together the six editions are rudimentary. In the body of the texts, XML encoding has been used sparingly, no doubt because of the cost of converting all those back volumes. When, as often is the case in scholarly editions, a portion of the text is bracketed to indicate the editor's regularization or uncertainty, the brackets are not removed from the text searches. Thus, if one edition used brackets in a phrase and the other did not, the text search would not find all instances of that document. Searches also do not always locate variant spellings of the same word, though the Collection does employ stemming. Most documentary editions use a literal transcription policy that seeks to capture the text as written, with misspellings and abbreviations rendered as is. This isn’t usually confusing when read by a human, but it becomes problematic when we rely on a computer to read the text.\textsuperscript{11}
- Consolidated the indexes to multi-volume editions that were published for more than fifty years is a daunting proposition. Indexing styles change over time and with historiographical trends. Rotunda has made available consolidated indexes for the Adams, Washington and Jefferson editions thus far. The indexes are created by coding in hyperlinks from the index to the volume and page number, hence the decision to retain the original pagination. Right now you can only search one index at a time and it is not clear if there will be an effort to merge them.

Most of the questions raised by the American Founding Era Collection’s digitization are ones that every editor would have to address when trying to make the conversion to digital form. It is made that much more complicated by the number of projects and legacy print volumes. The decision to maintain the organizational structure of the volumes limits the functionality that new volumes of the Founding Era or other editions can have if they want to join Rotunda. Of the six editions represented in the American Founding Era, only one was “born digital.” The Dolley Madison Digital Edition, when used as a standalone product presents a more website-like document display. To the left of the text are links to biographical identifications of names mentioned, keywords assigned to the text, and places mentioned. Each of these are linked to fuller annotations, short biographical studies and description of the places. A summary of the document is used on search results pages to help the reader decide which documents to open. This digital edition doesn’t offer the indexing depth found in the Washington, Adams or Madison indexes, but it does provide a more useable and flexible digital text. When the Dolley Madison Edition is searched as part of the American Founding Era Collection, however,
I don’t mean to seem overly critical of Rotunda’s *American Founding Era* project. Rotunda is the only organization even making the attempt to tackle the issue of digitizing legacy editions. I don’t think that Rotunda is claiming to be developing the next generation of digital editions, but because they are actively seeking new projects, their approach has the possibility of becoming a de facto standard for digital editions. I worry that its very literal approach to these print volumes may inhibit the development of more ambitious digital editions. Rotunda’s light encoding of the texts and its limited search options do not maximize the capabilities of XML encoding. Rotunda editions are likely to always resemble print publications. Neither does it have the capacity to include social media tools that digital humanists and web-savvy readers seek to foster collaboration and reader engagement. And in order to cover its costs, Rotunda charges a subscription fee for access to the collection. To be fair, our small, intensive, and understaffed digital edition doesn’t provide a good model either, with little in the way of support staff or capability for migration to newer data formats.

So where do we go from here? We need to do some hard thinking about how digital editions should look how we can sustain them. Digital media presents a number of challenges to the way that editors think about their texts and how they prepare them for the public. Thinking and talking about some of these and trying to see how XML might affect our decisions may help developers to anticipate future needs.

- Images of manuscripts are now easier to digitize and serve over the web. If editors can link their transcriptions to images of the manuscript, will that change the role of the transcription? How many people really want to see the image? How many readers really want to see all those strikeouts, additions, false starts and other complications of a text? While we could provide two different views of a transcription using stylesheets, do we need to do it? Could most of the people interested in the complexity of the original be served by looking at a good digital image? Could we then default to a regularized transcription that would be easier to read and more accurate as the base of text searches? Perhaps we can encode a links to the document image in places where the user might want to consult it. I am sure that some editors and some users would not feel comfortable doing this, but I think that it is an option worthy of a trial. It might not be appropriate for complex literary editions, but for many historical editions, it might serve well.

- The idea that once an edition was done it was unlikely to be done again is not a product of the digital age. Once text is digitized, particularly when using markup like XML, it becomes far easier to re-purpose it, run it through text analysis tools, add new levels of encoding, and open up the possibility that other scholars might find new uses for our old editions. At some point, a scholar can go back to the American Founding Era Collection and encode those variant spellings, or create a version that ignores brackets when searching. Someone might even want to try to tackle creating a comprehensive index. The chances of this happening go directly to the question of sustainability; because these texts were encoded in XML, they should be useable, so long as the scholars are allowed to use them.

- How will the editor’s job as annotator change as more and more materials are made available on the web? In days past, the editors’ subject specialization and familiarity with hard-to-find primary and secondary sources ensured high quality annotation. No single scholar could dedicate the hours and years that long-term editing projects do to their subjects. But now, the availability of more and more web-based resources means that
many once hard-to-find sources are readily available to the average reader. Should the editor still summarize a book when he can link directly to it on Google Books? Do we need to provide a short biographical identification when we can add a link to an entry in the online *American National Biography*, the individual’s obituary in the *New York Times*, or heaven forbid, his entry in Wikipedia? I don’t actually think that links can replace all kinds of annotation, but with many kinds of facts easier to find every day, editors should question how they annotate documents. As the digital edition reaches further outside its boundaries for annotation, it may start to resemble a web site more than a book. With the ability to use the rest of the World Wide Web as linkable resources, will editions begin to resemble an ever expanding “life and times” of the subject, limited only by the questions asked by the researcher, and the paths they choose to take?

- Following up on the idea that annotation may change, it strikes me that content encoding might replace at least some annotation and indexing tasks. If instead of spending time annotating, we can use our expertise to encode links to other documents, to names, organizations, and topics, and spend more time creating in-depth indexing entries, we may be able to provide as good a service to the readers as we now do by conducting annotation research. All annotation will not go away, but the editor will be freed to focus on the difficult concepts that the average reader might not be able to find for herself.

- One of the promises of digital publication is that it will make collaboration easier. One can see that an editing team in the 21st century might not need to reside in the same city or same continent. Figuring out how we can use cloud computing to construct digital editions, and looking into how we might credit contributions may help attract collaborators that have a skill or specialization we lack. One of the main problems with digital scholarship, especially of a collaborative nature, is the inability to easily cite the works on vitae or resumes. This can dissuade some academics from participating in team-based research as they build tenures portfolios. Can we develop new systems where portions of the edition are credited, such as translations, annotations and metadata?

- How will social media networks affect editing and XML? Web 2.0 tools are increasing in sophistication and enabling large amounts of people from all walks of life to participate in the creation of editions. One could conceive of a digital edition constructed as a wiki by volunteers who locate, digitize, transcribe, research, and proofread historical texts. Such a wikidition could grow either incrementally or exponentially depending on its ease of use, general interest, and word of mouth. If it should become even a hundredth as popular as Wikipedia has, one could see a large and diverse collection of materials taking shape outside of the control of editors and scholars. Blogging software has been used to present diaries, like that of Samuel Pepys, a site that encourages readers to comment on individual entries or provide more formal annotation in a companion digital encyclopedia. Investigations of the feasibility of using crowd sourcing for transcription and annotation are currently underway at the Papers of the War Department, an image-based digital edition sponsored by George Mason University. If any of these experiments take off, how will we preserve the digital editions they create? Can one export a wiki or a blog post into an XML format for long-term preservation? Can we develop XML-capable wikis and blogs that retain their ease of use?

- Will editors eventually include their project research files and databases in their editions? Only a small amount of the research conducted by editing projects makes it into the footnotes of their published editions. Should we share these research files, open up project chronologies, genealogies, image files, and name databases? Should we blog about research queries that come into our offices and about out own research undertaken for the edition? Should we somehow provide our readers with the experience of working in our editorial offices, where libraries, vertical files, word processing files and databases are all pressed to the service of one topic? If we do, can we easily convert these kinds of work files to XML, or should we be seeking
the development of an XML Office Suite that can handle our ongoing needs and also make them sustainable and accessible in the long haul?

Will we develop new ways of thinking about documents if we look outside of our traditions and perhaps if we look beyond what XML offers at present? Looking at other applications like GIS–Geographic Imaging System–for inspiration on how to organize information might be instructive. It strikes me that the way that a GIS map is constructed, made of discreet layers and kinds of data that can be selected in any combination by the user makes for an interesting model for digital editions. If the transcription and linked image served as the “map,” with interpretation, annotation, and metadata organized as stand-off encoding many people could share the transcription but be free to add their own interpretative layer. A letter written by radical anarchist Emma Goldman to Margaret Sanger while Goldman was on a speaking tour in Portland in 1915 would interest both the Sanger and Goldman editing projects. The Goldman project might focus its annotation on Goldman’s doings and ideas, whereas the Sanger editors might see the letter as evidence of Goldman’s mentoring role in these years. Other interested parties could also use the letter, adding their own comments, contextualization, and interpretation to it. For example a staff member at an Oregon historical society could use the letter in an online exhibit on the radicalism of Portland at the time. He could link the places mentioned to maps or historical photographs of the city. A genealogist might simply comment on a passing reference to her great-grandfather, adding a link to her web-based genealogy. Each of these users of the text would be adding to its meaning in different ways, all of which could enrich a reader’s experience of the letter. How can we create this kind of document in a way that allows any user to their annotation, and also allow users to choose to see any combination of these annotations simply by selecting them with a mouse click?

I don’t know if XML can do this, but what I am getting at is that if we don’t keep thinking creatively about how we might present these documents, we will end up replicating digital versions of old book editions. If we don’t continue to evolve and improve, we run the risk that other kinds of digital publication, perhaps those that are not as long-lasting, will become more popular because they have better functionality. XML was created to render in digital form the publications and scholarship that we were already producing in print. So it will be good at doing that, and maybe not as good at representing less structured organizations of texts. That doesn’t mean that we shouldn’t figure out ways to make XML do what we need it to do, it just means that it might be harder.

Being that we don’t know–we can’t know–where digital history and digital editing will go in the years to come, how can we ensure that the work we are doing now will last five years never mind fifty? Sustainability is the capacity to endure change and if we can say one thing about technology, it is that it is constantly changing. There are a few simple ways to ensure the longest life for our work. One is to make high quality content. If the legacy volumes of the Founding Father projects were not seen as a valuable resource, there would be no great effort made to preserve them. Because they have lasting value, efforts were made and money spent to keep them viable and accessible. The other best practice is to do what you can to make it easy for the next generation to preserve your work. I don’t think we have to promise that it will always last, just so that it will last until the next generation of technology comes along. At that point, if the value of the work isn’t there, it won’t be preserved. If it is, it should not be that difficult to migrate it. So don’t develop your own markup language, unless you are truly a genius, and if you do make sure to share it with the world. Pay attention to what the digital humanists of the day are using and advising. If we stick with the educated pack when it comes to data formats, we can
reasonably expect that the tools will be there to preserve our work. In short, that means we should use XML.

But we are not the only ones with a responsibility for making our tests last. A big part of determining whether or not a format is sustainable is whether it achieves buy-in from those it seeks to serve. As I said, most editors know that they need to use XML to create their digital editions, but that doesn’t mean they really want to. We need better tools and encoding environments to win over editors and other content providers. We need increased and sustained educational offerings and practical examples and templates that can help the most numbers of content providers, whether they be editors, archivists, scholars or students, to put up XML encoded manuscript material, and we need to make available the programs and stylesheets that will make these texts display clearly and will take advantage of the encoding to generate valuable searches. It behooves us to master XML encoding in order to take a creative part in the development of digital editions. If we hand XML encoding over to consultants or to or publishers, we are unlikely to get the kind of rich encoding that can substitute for annotation or indexing.

XML encoding is expensive. Even well-funded providers like Rotunda do not have the manpower to create in-depth encoding. It is no surprise that some of the best digital editions are coming out of universities that have dedicated digital humanities centers. Virginia, Nebraska, and Brown have built expertise and tools in XML encoding that benefits their affiliated projects. George Mason University’s Center for History and New Media has taken a different approach, fostering image-based projects that use their open-source Omeka software that relies upon a form-based creation of Dublin Core metadata for each object. The costs of adopting XML for your edition at an institution that is not engaged with digital humanities is high. Educational opportunities on the web as well as through in-person workshops are wonderful resources, but they don’t replace the place of the kind of extended aid one can get from experts or consultants. Lacking funding to build up a national program of educational resources on XML, we need to foster more communication among XML users working with digital texts, archives, literature, and museums. We can learn from each other and share with them the special experience that we have in publishing historical texts in easy to read forms.

Ultimately, the best thing that we can do to ensure the long-term sustainability of our editions is to engage more fully with XML developers and our colleagues working with similar materials. Those of us that use XML ought to encourage our peers to become more involved and truly master the capabilities and limitations that the format has to offer. We need to pursue joint projects and consortia to fund this development. If we can build some simple tools that can get editors started on encoding their manuscript material and displaying it on the web, we will have come a long way towards ensuring that both the XML format and the work of historical editors will be around for the long haul.

Notes


11. To see this, consult George Washington to James Madison, June 12, 1784. In the first paragraph, a sentence begins “Must the merits...” The Madison Papers transcribed the phrase as “Mus[t] the merits” while the Washington Papers regularized the case silently. When searching for “Must the merits” only the Washington Papers version of the document is returned. This will likely compromise the accuracy of searching throughout the collection. ([http://rotunda.uppress.virginia.edu/founders/default.xqy?keys=FGEA-info-search](http://rotunda.uppress.virginia.edu/founders/default.xqy?keys=FGEA-info-search) [Accessed 29 July 2010].)
12. See, for example, Dolley Payne Todd Madison to Anna Payne Cutts, 18 May 1804 (http://rotunda.upress.virginia.edu/founders/DYMN-01-02-02-0026 and http://rotunda.upress.virginia.edu/dmde/DPM0077 [accessed 29 July 2010].)


14. For a short description of this effort see http://www.neh.gov/ODH/Default.aspx?tabid=111&id=154; for more on the Papers of the War Department see http://wardepartmentpapers.org/ (Accessed 29 July 2010). This project is jointly funded by the National Endowment for the Humanities and the National Historical Publications and Records Commission.

15. As this paper was given, the Center for History and New Media announced the creation of Anthologize, built during its One Week, One Tool. Anthologize is a WordPress plugin that creates eBooks out of websites that can be saved in TEI format. For more on this development, see http://www.anthologize.org.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.
Chapter Three

Identifying Source Texts

<table>
<thead>
<tr>
<th>I. Authentication and Attribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>II. Types of Documentary Materials</td>
</tr>
<tr>
<td>A. Inscribed Sources: Handwritten or Typewritten Materials</td>
</tr>
<tr>
<td>1. Letters</td>
</tr>
<tr>
<td>2. Papers</td>
</tr>
<tr>
<td>3. Diaries and Journals</td>
</tr>
<tr>
<td>4. E-mail and Other Personal Electronic Records</td>
</tr>
<tr>
<td>B. Printed Works</td>
</tr>
<tr>
<td>C. Orally Generated Texts</td>
</tr>
<tr>
<td>1. Lectures, Speeches, and Sermons</td>
</tr>
<tr>
<td>2. Conversations and Interviews</td>
</tr>
<tr>
<td>3. Oral History Memoirs and Other Recorded Interviews</td>
</tr>
<tr>
<td>4. Audio-Recorded Speeches</td>
</tr>
<tr>
<td>III. Horrible Exceptions to These Simple Rules</td>
</tr>
<tr>
<td>A. Reconstructing Source Texts</td>
</tr>
<tr>
<td>B. Source Texts Demanding Translation</td>
</tr>
<tr>
<td>1. Foreign-Language Materials</td>
</tr>
<tr>
<td>2. Shorthand</td>
</tr>
<tr>
<td>3. Codes and Ciphers</td>
</tr>
<tr>
<td>C. Transmission of Documents in the Ages of Electricity and Electronics</td>
</tr>
<tr>
<td>1. Telegrams</td>
</tr>
<tr>
<td>2. Electronic Records</td>
</tr>
<tr>
<td>IV. Conclusion</td>
</tr>
<tr>
<td>Suggested Readings</td>
</tr>
</tbody>
</table>

Documentary editing is most clearly distinguished from traditional critical textual editing by its customary reliance on a single handwritten, typed, printed, drawn, or otherwise recorded document (the source text) for each editorial text. Although documentary editors occasionally use such tools of textual editing as conflation to establish an ideal text of some kind, their general rule is that one source text, whether an original or its reproduction, will be the basis of one editorial text. Identifying these source texts is the next phase once the work of assembling documents is under way.

Almost every editor will face the challenge of choosing among two or more candidates to find the source text for a document. Even editors of archival editions often find that their group of documents includes variants of the same item. Even though such variants between the source text and other versions will be recorded in editorial notes, it is the words, phrases, and punctuation of a single source that should be readily and conveniently available to the reading audience. Even in an electronic edition, in which the reader can use links to move from transcriptions of one stage of a document’s evolution to another, the editor must choose one of these versions as the basic text from which the links will generate. Thus, for any editor, choice of appropriate source texts is a crucial matter.

A variety of factors will determine an edition’s criteria for source texts. Some early decisions about an edition’s organization and scope may play a role. Knowledge gained in collecting and cataloging will add more considerations. However, it is the nature of the individual documents or groups of documents themselves that finally dictates the best approach.

I. Authentication and Attribution

Perhaps the most important criterion to be met by any document published in a scholarly edition is that it is, indeed, the intellectual product of the person or organization purported to be its author. Editors must rule out forgeries and misattributions while simultaneously making every effort to identify documents not previously recognized as the work of their subjects.

The papers of well-known individuals, like the works of great artists, are most likely to become the targets of forgers. The commercial market for letters and other manuscripts penned by heads of state, military and naval
leaders, famed scientists and inventors, and literary lights is centuries old in the Western world, and forgery is almost as ancient as that trade. Some editors must be more conscious of the danger of forgeries than others. The editors of the Papers of Abraham Lincoln, for instance, maintain an extensive file of known Lincoln forgeries. 

Joe Nickell’s Pen, Ink, and Evidence: A Study of Writing and Writing Materials for the Penman, Collector, and Document Detective and its bibliography provide a useful introduction to the history of documentary forgers. Experienced dealers or auction house experts familiar with an edition’s subject can also offer hard-earned knowledge of such fakes.

Almost as common is the problem of ruling out published or unpublished works that, while historically genuine, have been erroneously credited to an edition’s subject. Perhaps the best-known recent discovery in this field came when editors of the two-volume Political Correspondence and Public Papers of Aaron Burr (1983) revealed that the cipher letter from Burr to James Wilkinson that led to Burr’s denunciation as a traitor was not, in fact, written by Burr but by his associate Jonathan Dayton. Less well publicized but more historically significant was the Madison Papers editors’ decision to omit from volume 17 of their series the 1799 Virginia Resolutions and Address of the General Assembly. On the basis of both contextual evidence and a statistical analysis of word frequency from a model in Frederick Mosteller and David Wallace, Inference and Disputed Authorship: The Federalist, they became convinced that Madison was not the author of these statements.

Traditionally, scholars have shown more concern for the attribution of works to famous authors than to renowned politicians or generals, and attributions of anonymous and pseudonymous literature to novelists and poets are more likely to be subjected to exhaustive investigation. A notable exception is Benjamin Franklin, whose newspaper contributions were the subject of J. A. Leo Lemay’s monumental Canon of Benjamin Franklin, 1722–1776: New Attributions and Reconsiderations. For any published author, computer analysis of usage may help corroborate or disprove attributions.

In most cases, the editor will not be the first scholar to examine the subject in question, and questions of attribution will be addressed throughout the editorial project, from the time of collection until the point when informational footnotes are added to the editorial text. The editors of the Papers of John Adams were able to credit Adams with drafting a March 1773 message of the Massachusetts House of Representatives only after they had reviewed all of Adams’s correspondence concerning the edition (1:312–13).

The decision to omit from or include in an edition unattributed verbal documents must be weighed carefully. It involves a combination of factors including writing style and word usage and circumstantial and contextual evidence—all of which presuppose that the editor is familiar with a large cross-section of the documents. It means that verbal documents have to be scrutinized as carefully as graphic works have been when experts authenticate drawings and paintings.

It is not only published works that may raise questions of attribution. Scholars who deal with the papers of any figure who commanded a large civilian or military staff must establish guidelines that define when a document may be considered that figure’s work and when it should be credited to a junior associate who drafted the document, which may have been dispatched without a moment’s review by the person whose signature it bears. The editors of the papers of Dwight D. Eisenhower and George C. Marshall admit that their task here was made easier by the clear chain of command both generals demanded. They were sticklers not only for precisely established office procedures but also for putting those procedures on record. This meticulousness enabled scholars to understand patterns of documentation and to recognize which set of initials or stamps identify documents by Marshall or Eisenhower.

Even those editors, however, faced a decision requiring Solomon-like judgment when they confronted a radio message to Douglas MacArthur of 7 February 1942. As a member of Marshall’s staff, Eisenhower drafted the dispatch, and a typescript copy of that draft was submitted to Marshall, who then covered margins and interlinear spaces with his additions and changes. So extensive were Marshall’s revisions that a complete retyping was necessary before the message could be coded and dispatched. Here the two sets of editors saw two documents with two different authors. The Eisenhower edition (1:101–3) printed the typescript portions of the draft, noting Marshall’s revisions in footnotes. The text in the Marshall edition (3:100–102) was based on the typescript that incorporated those changes, the version that reflected Marshall’s wishes.

Authentication and attribution are only the beginning for the editor who must choose source texts. Only a few candidates for source texts will be excluded here, and the process of evaluation goes on.

II. Types of Documentary Materials

The methods used to create the documents at hand are likely to dictate criteria for evaluating source texts. These rules for selection, like all editorial procedures, should be stated clearly in an edition’s introduction. For modern and Anglo-American documents, the criteria generally involve little more than common sense, and they relate to how a document came into being.

A. Inscribed Sources: Handwritten or Typewritten Materials
1. The manuscript or a reliable photocopy or scanned image is to be preferred over any later scribal copies or transcriptions as the source text.

2. If the original has been lost or destroyed, contemporary copies are preferred over later ones unless evidence demonstrates that later copyists both had access to the original and were more accurate than the earlier scribes.

3. In general, the most nearly final version of a document is the preferred source text. Editors can take comfort in the thought that variants can and should be noted in the editorial apparatus. By choosing one version over another, editors do not deny readers access to significant differences among versions of a document. The source text is simply the one that serves as the best working basis for the edition and most closely meets the needs of the edition’s audience.

Other criteria for establishing priorities among such materials are listed below.

1. **Letters**

The first order of preference is given to a version of the original letter (preferably signed to denote authorial approval) known to have been received by its addressee. The best evidence for identifying a recipient’s copy of a letter may be an attached address leaf (for letters before the mid-nineteenth century) or a mailing envelope (for later correspondence), the recipient’s endorsement, or the location of the letter in the addressee’s papers. Even when no address leaf or envelope survives to prove that a copy of a letter passed through the mails, fold patterns and other physical evidence often mark a particular document as the recipient’s copy. The editor familiar with a subject’s letter-writing habits will generally be able to identify “finished” versions with little trouble.

A special problem may arise with the correspondence of statesmen of earlier centuries who were assigned to foreign posts in times of national emergency. At such periods, diplomats frequently sent duplicate or even triplicate or quadruplicate signed copies of their dispatches to friends or government offices at home. When two or more such multiple addressees’ copies survive, the editor has to decide which version to use as a source text. The decision may be based on the completeness of the variant copies or the care that the letter’s author or secretary used in copying the dispatch. A triplicate letter in which words are carelessly omitted or in which the handwriting reflects haste would be a less desirable text than the quadruplicate of the same letter in which the text is complete and the manuscript shows that it was copied more accurately than the triplicate. Nontextual considerations, however, may be decisive here—in particular, evidence showing which copy of the letter was received first or received at all. A carelessly copied duplicate version of a letter from abroad that reached the State Department in June and was the basis for a foreign-policy decision would be a better source than an elegantly inscribed first copy that arrived several weeks or months later.

A second order of preference is given to copies of the letter in the hand of the author or someone under his or her direction that were designed to be retained, not transmitted. Such versions fall into four categories: drafts, letterbook copies, copies that present fairly exact facsimiles of the recipient’s version (such as letterpress or carbon copies), and independently prepared loose file copies.

Preliminary draft versions on separately inscribed sheets are easily identified, but some of the other categories are not mutually exclusive. In the eighteenth and early nineteenth centuries, bound volumes of blank pages were used to record outgoing letters. Many authors used such letterbooks as a convenient place in which to draft correspondence, recopying the final version on a fresh sheet for mailing. Others copied texts of a letter’s final version into the letterbook. Thus, some letterbooks represent volumes of drafts while others are assemblages of file copies of polished versions of outgoing correspondence. More difficult still are public figures who did not leave well enough alone. George Washington could not resist tinkering with youthful letterbooks once he achieved middle age, correcting spelling and awkward usage. A clerk then set to work recopying the emended results, creating a textual nightmare for present-day editors (Abbot, “An Uncommon Wareness of Self”).

Documents that represent contemporary facsimile versions include letterpress copies, carbon copies, and modern photocopies and computer files. The efforts of American letter writers before the era of carbon paper to create files of outgoing correspondence without recopying the letters are a tribute to their ingenuity. In the late eighteenth century, Thomas Jefferson and others enthusiastically experimented with every invention that could ease their burden. Stylographic pens and pencils, which produced duplicate images of holographic materials, were employed. Jefferson and Charles Willson Peale purchased polygraphs, devices that made two pens move simultaneously to produce two copies of each letter, one for transmittal and one for filing.

Last, but not least, there was the letterpress. In the earliest of these devices, a thin, nearly transparent sheet of paper was moistened and pressed against the inscribed surface of handwritten material. Ideally, enough ink was transferred from the original to the back of the blank sheet so that the handwritten words showed through and could be read from the front. In practice, most letterpress copies were fragile, smudged horrors until the mid-nineteenth century, when more sophisticated techniques produced more satisfactory copies.

The invention of the typewriter and the use of carbon paper added still another form of simultaneously created facsimile copy. Although carbon copies of typed materials were part of American life and usage for more than a century, editors have seldom written of the problems of rendering intelligible versions from uncorrected carbon
copies. They've been even less eager to comment on copies created by photocopying or printed from computer files. (A notable exception was Fredson Bowers’s “Multiple Authority: New Problems and Concepts of Copy-Text.”) Until more such literature appears, editors can only rely on their own common sense and the experience of other editions.

A third order of preference is given to transcriptions and printed copies. Editors commonly designate as “transcriptions” copies made substantially later than a document’s composition and executed by someone acting without the authority or assistance of the author. As source texts, such typed or handwritten transcriptions and printed versions of letters rank far below an existing contemporary copy. To complicate the matter, the families of many well-known figures like Thomas Jefferson began copying an author’s letters in his or her lifetime and continued doing so in the years after the celebrated author had died. It may be impossible to determine which copies are contemporary and which posthumous. The editors of Jefferson’s papers are fortunate, since the original documents, in Jefferson’s hand, usually survive for use as source texts, while the copies and transcriptions need only be cited in notes. Other editors aren’t so lucky.

Editors must often choose among several transcribed copies of the same document to determine which is the best source text, the one closest to a vanished original. For modern materials, nontextual evidence may identify the best transcription: patterns of handwriting that indicate age, results of chemical tests on paper, or a comparison of typefaces in printed sources. When nothing useful can be learned from such historical evidence, editors use more advanced methods of classical textual filiation, whose commonsense rules apply to modern as well as ancient materials. These rules require that the editor (a) attempt to learn something about the copyist for each scribal version and (b) identify patterns of common and unique errors in the copies to determine the sequence in which the variant transcripts came into existence. Transcripts that contain the same errors are likely to be part of the same family, one descended from the other in order of recopying.

Like handwritten transcriptions, previously printed versions of a letter can reflect the interpretation and even the style of a later copyist, not the author’s intention. The problem addressed here is not that of establishing the preferred text of a work written for publication and printed during an author’s lifetime, but of letters or other private materials for which the original has disappeared and only a version printed in some earlier edition survives. When more than one such printed text survives, the editor must decide which has the greatest claim to authenticity and accuracy. Some of the techniques for choosing among handwritten or typewritten transcriptions apply. Obviously, the editor has an advantage in establishing chronological patterns among printed texts, which usually bear a date of publication. But even this can be misleading when earlier editors prepared their printed versions of a source text by referring independently to the same archetype. The eye of the editor who published a text in 1830 need not have been more reliable than the one who went to work in 1880. When evidence indicates that two printed versions or two transcripts of an original were both based on the same source text, the editor relies on such internal evidence as faithfulness to known patterns of authorial spelling, punctuation, and capitalization, as well as the accuracy of the copyists’ readings of proper nouns. Many printed versions dating from the nineteenth and early twentieth centuries appeared without statements of editorial policy, and modern readers must resort to their own critical judgment in selecting a source text.

As Herman Saatkamp pointed out in “Private Rights vs. Public Needs,” editors may also have to deal with letters that were intended to be private but were published nonetheless. He gives a hypothetical example of an unknown author whose letter became a rallying point for political revolution and was published and republished, with significant variants. The documentary editor’s decision whether to publish the earliest or the last version is a choice between authorial intention and the letter’s context and significance. While the author’s intention would be represented best in the earliest published version, the political movement’s use of the letter would be more evident in the last published version. A volume of the individual’s writings might well include the earliest published version of the letter, while an edition of the records of the political movement would demand use of the last published form. Here, the “context is not the letters of an individual, but the letters of the revolution” (92).

2. Papers

In general, the rules that apply to an individual’s personal or professional correspondence also apply to selecting source texts from variant versions of his or her papers, records of professional activities or public life that cannot be defined as any form of letter or diary. This category includes the legislative reports of a lawmaker, the technical notes of a scientist, the general orders of a military commander, or the lecture notes of a reformer or educator.

Special considerations sometimes may make the best source text something other than the most nearly final version that survives. These are very practical issues relating to the availability of reliable editions of these final versions and the absence of such accessible editions for preliminary versions that may display significant variants. The problem arises frequently in the papers of women and men who held public life (p.95) life. Official government publications often provide adequate, accurate versions of legislative committee reports and formally adopted statutes. Given their limited resources in terms of money or size of publication, modern documentary editors may hesitate to republish an already available printed version of the same materials. If the search for the edition uncovers manuscript or typewritten draft versions of such reports or pieces of legislation, these little-known
documents might better serve readers as source texts. Their notes can direct scholars to the location of the printed final versions, which can be used to create parallel texts of the evolutionary stages of the same document.

3. Diaries and Journals

Personal diaries, like some letterbooks or pocket notebooks, have been subject to subsequent revision and improvement by their authors or later editors. When the resulting earlier printed versions have achieved wide circulation and popularity, a new edition of the source, restoring its value as documentary evidence, can be one of the greatest challenges in documentary or textual editing. C. Vann Woodward recounted his efforts to deal with this problem in “Mary Chesnutt in Search of Her Genre,” and Joel Myerson and Daniel Shealy discuss their more conservative methods of dealing with another nineteenth-century woman’s journals in “Editing Louisa May Alcott’s Journals.”

Usually, however, only one version of a journal or diary survives. While its pages may be difficult to transcribe and its references may be baffling to annotate, there are seldom rivals for the title of source text. If more than one does survive, the one closest to the original inscription has the most documentary value.

4. E-mail and Other Personal Electronic Records

“Inscribing” one’s correspondence or daily reflections on a keyboard connected to a computer will not necessarily make life easier for your future editors. Electronic forms of correspondence and diary keeping can create variant versions as easily as those created by pen or pencil or typewriter. Even though those versions may be more legible than scrawled handwriting, they don’t save an editor from decision making. Not only may an author’s files contain drafts of a message, that message may have been sent to multiple parties. Authors who maintain personal journals based on bits and bytes rather than ink and paper are still free to rethink and revise these entries.

These convenient electronic writing systems create their own challenges as well. E-mail makes a distinction between “cc” (carbon copy) and “bcc” (blind carbon copy) transmissions. Editors who catalog e-mail files must decide early on whether to retain the original e-mail address and domain and server names tracing the route of transmission. Beyond individual e-mail exchanges, there are the treacherous reaches of the “Listserv.” As this Guide focuses on references to existing editorial practice rather than speculative advice, we will do little but recite the special problems of this new form of text. Readers will be reassured to know that the issues have already been considered in a recent presidential address to the ADE, Beth Luzy’s “Editing Bill Gates’s E-Mail.” Luzy employs humor and insight to analyze the problems of editing a large corpus of the papers of a figure who spans the twentieth and twenty-first centuries.

B. Printed Works

The literary works of any writer—prose and poetry composed for publication and issued in print more or less as intended by the author—present special problems for the documentary editor. Of these, the first is the choice of which version of the published work will be the basis for the text in the new edition. The editor who aspires to a CSE emblem will naturally refer to the guidelines of that organization before choosing a version for transcription. The choice there may be the most appropriate copy-text, not a source text, and the criteria for its identification lie outside the boundaries of this volume.

A copy-text may be the foundation of an emended critical text whose aim is the representation of the author’s final intentions, intentions that may or may not appear in all respects in any single surviving copy of the work in question. The critical edition will then be a new document itself, recording the best judgment of the editor, not the words or punctuation of any single version published in the author’s lifetime or available to any earlier reader. The documentary edition of such a work, however, is normally a noncritical one based on a single printed version that was actually read by the audience for which it was originally intended.

The differences between the goals of a textual and a documentary editor may extend to the choice of the version that will be the basis for the editor’s work. Traditionally, textual editors of literary works prefer the printed version that reflects the author’s most fully realized intentions. They assume that most of their readers are concerned with the development of an author’s literary craftsmanship, which is often reflected in her or his correction of earlier printed errors. Modern textual scholarship has questioned these assumptions, but a focus on authorial intention is still an overriding consideration for many. Documentary editors must also weigh factors such as the historical impact of specific editions or printings of the same book or essay. Many considerations can make something other than the final version in a set of printed variants far more important in documentary, evidentiary terms than the last, most polished edition.

As an example, textual and documentary editors might make very different decisions were they to choose between two versions of a syndicated column published in a 1940 newspaper and then revised by its author for inclusion in a book-length anthology of such pieces twenty years later. An editor concerned with the journalist’s literary intentions might conflate the elements of the two into a critical text, but a documentary editor would weigh
the public influence of each version. If the 1940 column was an important factor in that year’s presidential race, while the 1960 revision languished in books that quickly went to remainder tables, the documentary editor would choose the first. The pamphlet’s significance for an audience of historians of politics or journalism is its impact in 1940. The author’s later improvements could be recorded as variants in notes, but it would be nonsensical to use the 1960 revision as a source text, forcing readers to reconstruct the 1940 column through laborious reference to footnotes or tables of emendations.

For printed material, as for any source text, both the source and the projected audience will influence editorial decisions. Whether literary or historical values are the primary reason for republishing a printed document, the choice of source text or copy-text demands a thorough knowledge of the document and its creator and the needs of those likely to consult the new edition. An editor must assume the role of historical bibliographer to master the story of a work’s composition and publication before deciding on the appropriate basis for a new edition. Knowledge of the technology of publication methods must be part of this skill.

Newspapers present a special challenge, for pieces published in that medium are peculiarly affected by changes in the mechanics of dissemination. Until the mid-nineteenth century, the newspaper that first printed an author’s essay or articles was usually the one to which the writer had submitted a manuscript, and its columns were likely to carry a version that had received some authorial review. Later newspaper {p.98} printings in this era customarily drew on the first, and the variations that appeared were more likely to be typographical errors that could be noted in the editorial apparatus. Tracing authorial contributions to the press in this age can be challenging, but textual problems are comparatively straightforward. (See, for instance, John M. Robson, “Practice, Not Theory: Editing J. S. Mill’s Newspaper Writings.”)

This situation changed with the development of newspaper syndication and the telegraphic transmission of text in the late nineteenth century. If an author’s article or story was published in syndicated form, the editor’s research must expand to the methods used by the news syndicate in question. Here the methods of filiation may again be helpful, for the original can often be reconstructed by analyzing the variants in second- and third-generation copies.

Simply put, editors preparing noncritical documentary texts of printed sources cannot be uncritical in their methods. They may not ignore tools of textual scholarship or other disciplines that can helpfully be adopted. Many printed items can be viewed as documents: official government publications, pamphlets, essays, and books. Within each category, the editor may need to become familiar with all the tools of bibliography and textual criticism, even though the editorial product itself will not be a critical one. To edit printed sources as documents, one must investigate their printing history with all the care employed by an editor hoping for a CSE emblem. Various—and variant—printings of the document may need to be collated as rigorously as for any MLA series, and the results of that historical research and the mechanical collation should be analyzed scrupulously.

Neglecting the tools of sophisticated bibliography for documentary editions can have embarrassing results. The editors of the Laurens Papers admit openly, if not happily, to such a blunder in the fifth volume of their series, where they presented a scrupulously printed facsimile of Laurens’s 1767 pamphlet A Representation of Facts. Variant readings from all surviving copies of the pamphlet are recorded in textual notes, but the introduction to the document states that the editors were unable to discover the precise order in which the last two versions of the pamphlet were issued. Had the editors used the tools of modern bibliography to compare the variants for such clues as broken type, they might have determined the sequence in which the versions of the pamphlet were run.

In documentary editing, the compositor’s type font should be regarded as an element of inscription as important as a scribe’s copying {p.99} practices, the configuration of the keyboard of an 1882 typewriter, or Woodrow Wilson’s use of Graham shorthand. Editors of printed documents cannot assume that their task is easier because their sources already exist in typeset form. Instead, they must make themselves experts in this method of documentary inscription. Fredson Bowers’s Principles of Bibliographical Description will disabuse such editors of the notion that editing printed sources is simple, and the members of the CSE are ready to offer advice and encouragement to novices in their fields of specialization.

For a sharp reminder of the ways in which a disregard for bibliography and printing history can invalidate an edition of printed documents, see Jack Warren’s “The Counter-Revolutionary Career of Peter Porcupine.” Assessing modern reprints of pamphlet literature that pretend to scholarly quality, Warren remarks:

A pamphlet edition ought to present complete and authoritative versions of clearly defined source texts, sufficient to make it unnecessary for scholars pursuing most lines of inquiry to refer to the originals; the edition as a whole should be based on well-defined selection criteria relevant to the contemporary importance of the writings involved; the notes should provide as much information as possible about the publishing history of the texts and should point out any significant variations between contemporary editions. Such basic standards may seem obvious to most documentary editors, but they are often neglected by those editing pamphlets. (93)

C. Orally Generated Texts
Until the development of practical sound recording at the end of the nineteenth century, speeches, conversations, sermons, and interviews could survive only as inscribed records, handwritten, typed, or printed texts. In the last century, phonography—sound recording and reproduction by machines—has provided facsimiles of spoken archetypes to complicate and enrich the life of scholarly editors. First, we’ll examine the choice of source texts for the prephonographic era—a wide and varied menu, indeed.

1. Lectures, Speeches, and Sermons

It is not uncommon for such oral presentations to survive only in the handwritten or typed version from which their author read aloud, with no records made by witnesses to the event. Such a “reading text” or “pre-text” is the only surviving text, and the editor is spared further work in identifying the source text.

However, popular clergymen and other speakers were often not the only persons likely to leave records of these events. The editors of the Frederick Douglass Papers found such an embarrassment of documentary records that they devised useful categories into which they grouped the records left by witnesses to his speeches: summary, narrative, extract, and stenographic text. The first three are usually brief paraphrases of the words actually spoken, and the distinctions among them refer to the completeness of the paraphrased record left behind. When one of these forms is the only record of an oral text, it is transcribed and reproduced in a documentary edition as literally as possible. Editorial notes can explain the nature of the source, and readers can judge for themselves how accurately the reporter has mentioned, narrated, or summarized the speech or conversation involved.

Through the early nineteenth century, various mentions, narrations, and summaries of the same speech or conversation often survived in textually irreconcilable versions. In the absence of systematic shorthand, no two reporters left accounts of the same spoken words that could be viewed as variants of the original. Verbatim extracts and stenographic reports of speeches are more common in documents inscribed after the mid-nineteenth century, when systematized methods of shorthand became increasingly common.

The editor may be unable to identify one source text here, choosing instead to give readers access to all conflicting reports. Should one record of the speech be a ten-page narration, while the others are paragraph-long summaries of the same words, editors usually transcribe and print the longest record as the editorial text, reporting transcriptions of the shorter versions in notes or as separate documents.

Lectures and political speeches are only one variety of orally transmitted texts to have received attention from modern documentary editors. Sermons have also received considerable attention, and many American divines have left modern editors both manuscripts used as reading texts in the pulpit and manuscript or printed versions modified later for the reading public. For a valuable discussion of the differences between such sermons as oral events and as literary documents, see Wilson Kimnach’s “Realities of the Sermon: Some Considerations for Editors.”

2. Conversations and Interviews

Other difficulties in editing oral communication were faced by a group of editors of conversations with well-known writers that were published in the Victorian era. In any series of published interviews, the personality of the interrogator may be as important to the reading public as the words of the person being interviewed. When more than one version of an interview survives, the expected audience of the modern edition may dictate the choice of source text. An earlier, uncorrected version would serve those interested in a biography of the interviewee; while the published, final versions better serve the study of that same figure’s public reception or reputation (Patrick Scott and William Thesing, “Conversations with Victorian Writers: Some Editorial Questions”).

3. Oral History Memoirs and Other Recorded Interviews

An even more complicated form of orally communicated record arose from the oral history movement initiated nearly sixty years ago by Alan Nevins at Columbia University. In earlier years, oral history projects tended to interview well-known figures, but today interview subjects represent all walks of life. Indeed, it is now recognized that the records of such interviews may provide scholars with unique evidence from men and women who would otherwise have no voice in the public forum. Oral histories have special value now precisely because they can record the memories of people unlikely to keep extensive diaries, maintain large personal archives, publish memoirs, or attract the notice of contemporary media.

Whether the subject is well known or comparatively obscure, oral histories are firsthand spoken narratives recorded with high-quality sound or video equipment in a carefully structured interview setting. Such interviews are quite distinct from recorded speeches or reminiscences in that interviewers do their best to elicit detailed responses to open-ended questions. The very setting of the oral history interview introduces an intellectual challenge. As Ronald Grele pointed out, unlike written diaries and letters, “oral history interviews are constructed, for better or worse, by the active intervention of the historian” (“Movement without Aim: Methodological and Theoretical Problems in Oral History”).
The process of transcribing the interview can never capture all its nuances, but methods for representing the
false starts, breaks in thought, (p.102) and other nonverbal signals are established in numerous oral history
resources. Less standardized are the methods oral history interviewers or projects use to control the “authorship” of
those transcriptions. Most oral history programs ask interviewees to review initial transcriptions of their interviews,
encouraging them to correct errors and allowing them to delete passages they regret or to place restrictions on the
use of other sections. Some programs provide interviewees with a chance to review even these “corrected” versions
as well. Quite simply, this means that the choice of a source text for an oral history interview may be determined by
factors beyond the editor’s control. If the oral history project’s guidelines make the original audio recordings or
uncorrected transcriptions unavailable, there will be no choice but to use transcriptions reviewed and emended by
the interview.

The Margaret Sanger edition found that even a complete and apparently straightforward source text for an
interview had its pitfalls. A videotape of Sanger’s television interview by newsman Mike Wallace had survived in
excellent condition, and the network had created a transcription of the Sanger-Wallace conversation at the time.
The surviving transcription was so incomplete and inaccurate that Sanger’s editors had to view the videotape again
and again, correcting the original transcript to produce a truly accurate text for their own edition. The supposedly
authoritative contemporary transcription became merely a reference to be mentioned in the source note; the videorecording itself was the source text.

4. Audio-Recorded Speeches

Some of the methods adopted for oral history interviews can apply here. The luckiest editors of publicly recorded
speeches are those with the option of using sound recordings as their source texts for transcriptions and even of
making these recordings available to the public. Here, too, issues of a special kind will appear.

Recordings of arguments before the U.S. Supreme Court since 1955 have proven so popular that they are now the
basis for competing editions, some free and some fee based. The first was Peter Irons’s 1997 May It Please the
Court: The First Amendment: [Live Recordings and] Transcripts of the Oral Arguments Made before the Supreme
Court in Sixteen Key First Amendment Cases. This set consisted of four cassettes and a 262-page book containing
explanatory notes and transcriptions of the materials on the tapes. It was followed two years later by Jerry
(p.103) Goldman’s edition, The Supreme Court’s Greatest Hits, a CD-ROM containing recordings of arguments
before the Court and editorial annotation based on Goldman’s classroom use of the materials. The Greatest Hits
appeared in a second edition three years later, this time expanded to two CD-ROMs. In addition, Goldman’s online
“Oyez Project” makes much of the material free over the Internet in MP files.

A more conventionally scholarly example of such publication is the Presidential Recordings Project at the
University of Virginia. This project transcribes, edits, and publishes selections from recordings made secretly in the
White House, 1940–1973. The audio files include meetings and telephone conversations of presidents from
Roosevelt through Nixon. This historical record is more than 5,000 hours long, and neither the complete audio files
nor their transcriptions will be mounted on the Web site nor published in the sets of volumes in the Thematic Series
or the Reference Series (both available from W. W. Norton). To provide access to the larger archive, the project
maintains its Presidential Recordings Program on the Web with a host of information on the recordings—including
their provenance and articles on the generation of tapes in different administrations—along with downloadable files
of the complete unedited tapes as provided by the National Archives. When available, there are flash modules for
the recordings, linked to transcriptions.

III. Horrible Exceptions to These Simple Rules

A. Reconstructing Source Texts

Occasionally, editors will have to apply scholarly skill and knowledge simply to reunite the physical components of a
source text. André De Tienne recounts the heroic measures demanded of the editors of Charles Peirce’s manuscripts
in “The Peirce Papers: How to Pick Up Manuscripts That Fell to the Floor.” In the decades after his death, Peirce’s
papers were disorganized and reorganized by a series of well-meaning students and admirers. The problem was
exacerbated by Peirce’s disinclination to supply page numbers or dates in any consistent fashion. Today his editors
must use a “hanging page display” of photocopies that offers “all the advantages of a recycled laundry line” to bring
order to the chaos. To aid them, the project maintains separate databases for manuscript organization, publication
history, and a calendar.

While few editors will face a task like that of the Peirce scholars, many will encounter the phenomenon of a
document that (p.104) was simply torn in two. For the detective work needed to resolve such a dilemma, see John Lupton’s “Putting the Pieces Together,” the tale of a Lincoln legal document ripped in half, with
each segment finding its way to a different repository. The reconstruction of Lincoln’s professional papers for a
comprehensive edition enabled editors to reunite the pieces of the whole.
B. Source Texts Demanding Translation

Some documents demand more than traditional textual methods of transcription and emendation to make them intelligible to their intended modern audience. These are the texts that must be translated from a foreign language or from authorial shorthand or cryptography of some sort.

1. Foreign-Language Materials

Editorial policies on the translation of foreign-language materials in an edition are discussed below, in chapter 6. Final decisions in this area may depend both on the number of such texts and on the intended audience. Whatever the choice, it will have an impact on the choice of source texts.

In past decades, American documentary editors have followed a variety of paths. A few editions publish all documents in their original languages, but this is rare. Some leave foreign-language materials untranslated if they are in a language the recipient was able to read. Others translate all non-English materials. Of these, some publish both the foreign-language original and its translation, while others publish the translations alone. Some publish contemporary translations prepared for the recipient of a letter or document, whether or not these translations are entirely accurate. Others commission new English translations. We offer some examples to make clear their rationales.

A single volume for Woodrow Wilson’s presidential years may include materials in French, German, Spanish, and Japanese as well as English. The foreign-language materials that qualified for inclusion in the Wilson edition formed but a small percentage of the whole, and the editors assumed that diplomatic historians interested in specific documents were already masters of the languages in which these letters and state papers were inscribed.

Other editions have tailored—and sometimes altered—their policies on foreign-language sources more flexibly. Whenever possible, the Washington Papers edition publishes the translations actually used by Washington, those prepared by his secretaries and aides and filed with the foreign-language originals. Originally, the Thomas Jefferson edition did not translate foreign-language materials if they were written in languages Jefferson himself could read easily. More recently, however, the editors of both Jefferson editions provide readers with translations of all foreign-language materials.

The Lafayette editors followed still another course. Lafayette’s papers for the Revolutionary era include so many documents in French that only bilingual readers could have read the series had all documents appeared in their original tongues. The editors chose to translate all materials into English, the language of the audience to whom the letters were primarily addressed. These translations were clearly labeled as such, and they appeared in the annotated, chronological series of editorial texts in the Lafayette volumes. An appendix to each volume printed transcriptions of the French originals.

The editors of Albert Einstein’s papers based their decisions on sources whose audience would be substantially different from the readers of the Lafayette edition. There was never a question of presenting English translations of Einstein’s writings in German as editorial texts. Unlike Lafayette, who learned English as a young man and who maintained a substantial correspondence with English-speaking friends throughout his long life, Einstein thought and wrote in German—only rarely did he compose in English or French. Many argued for parallel texts of non-English documents and their facing-page translations in the printed volumes, but Einstein’s life and intellectual contributions made him a figure of world interest. Americans might have been well served by English translations in the annotated volumes, but the books would then have been of little use for the cosmopolitan audience they deserved. It was finally determined to publish all documents in German, English, and French in their original languages; translations appeared for the small number of sources inscribed in other tongues. A simultaneous microform of unannotated English translations for German and French texts appeared with the first printed volume, but interest in these first translations proved so great that a paperback edition was issued as well, and this format continues throughout the series.

John Stachel, the Einstein edition’s first director, explained his rationale: “Placing the translations in a separate and clearly subsidiary position has an additional advantage: it reminds readers that German was Einstein’s language, which he habitually used when he wanted to express himself precisely; hence, a struggle with the German text, using the translation as a pony if need be, is highly preferable to assuming that any translation has adequately conveyed the full import and all the nuances of Einstein’s deceptively simple German” (Einstein from “B” to “Z,” 64).

2. Shorthand

Almost everyone adopts some system of abbreviations or symbols to save time and effort when taking notes or drafting a letter or other document. This may be no more sophisticated than using “&” instead of “and” and scribbling Arabic numerals instead of writing out “one,” “two,” “three.” A few figures who have become the subjects of editorial analysis were trained in formal shorthand systems, and many more have employed secretaries with
these skills. All of these forms will eventually have to be translated if an edition is to be intelligible.

Among American political leaders, Woodrow Wilson ranks first as a shorthand writer. Unfortunately, Wilson learned the rather obscure Graham system instead of the far more common Pitman shorthand technique, and he used the handy method throughout his life. His editors at Princeton counted themselves blessed when they found an expert who could still translate the archaic symbols. Here the only possible source texts were the sheets Wilson had covered in Graham symbols. Until they were translated by the Graham expert, they were meaningless to anyone else on the staff.

3. Codes and Ciphers

Coded and enciphered communications present textual problems analogous to those of standardized shorthand. In both cases the editor should make every effort to translate symbols into verbal equivalents that the modern reader can understand. Systematic codes and ciphers are customarily used to ensure confidentiality in a writer's communications with a second party. Such cryptic passages usually appear in communications of considerable historical significance, such as diplomatic dispatches or private correspondence between political leaders. These documents are in the form of communications between two parties. Not only should the translated clear text of codes enable the reader to see just which sections were entered in code and cipher, but the text or the accompanying notes should also record what the editor has been able to determine about the recipient's success in mastering the ciphered passages. (p.107) Indicating which words, phrases, or sentences were significant enough to deserve encoding allows the reader to see exactly which information in the letter was judged confidential and which facts the writer felt free to leave open to prying eyes. Noting both the author's skill in encoding his or her own words and the correspondent's accuracy in using the key to the code is critical in showing the effectiveness of the transmission of the ciphered information.

Diplomatic records after the mid-nineteenth century usually contain exhaustive files of every version of an encoded message, and their editors are seldom left in doubt about what message could be read by whom, but editors of the writings of statesmen of the American Revolution frequently confront greater challenges. Many of these leaders established personal codes or ciphers for use in private correspondence during their public service in wartime, and the same men often received diplomatic appointments that required them to use official government ciphers for their correspondence.

This means that possible source texts may survive in several forms. In a draft letter, symbols for passages the author entered in cipher may be written above their verbal equivalents. The author's file copy may carry interlined cipher equivalents, or it may be copied directly from the final recipient's version, with only numerical symbols or hieroglyphics for the coded passages. Among the recipient's papers, the editor may find a virgin copy, with the encrypted text deciphered on a separate sheet. Some luckless editors can find no deciphered version of a text. If they can find the code or key to the system employed, they go to work as their own decoders, working from the most authentic version of the ciphered original as the source text. The source text for an encoded document from this era must be chosen with special care to ensure that it represents the most significant version of the message. As the editor may not be able to make a final decision here until different copies of the coded document have been translated into clear text, there is no disgrace in choosing several candidates for decoding and transcription.

The editors of modern political and military records seldom need to decode enciphered materials. Unlike eighteenth-century diplomats, who were responsible for encoding and decoding their own official correspondence, twentieth-century statesmen and generals seldom saw the ciphered version of confidential dispatches. Instead, aides or technicians unraveled the mysteries of increasingly sophisticated systems of encryption. (p.108) The messages' addressees customarily saw only deciphered texts or summaries of them prepared by these aides. Editors of documents in this category try to identify the version read and acted on by the recipient for use as a source text.

C. Transmission of Documents in the Ages of Electricity and Electronics

Along with visible agents of inscription such as pens, pencils, crayons, and typewriters, the currents that made possible electrical telegraphy and the computer systems that have given us word processing and other electronic records are responsible for their own challenges and complications in choosing source texts.

1. Telegrams

At first glance, editors with the text of received telegraphic messages deal with nothing more than a ready-made translation of Morse code signals. But the words of any telegram existed at one time or another in at least three different versions: the text the sender submitted at the telegraph office, the Morse code signals transmitted by the telegrapher while looking at the sender's text, and the translation of those signals as inscribed on the telegraph form finally delivered to the addressee. While it's rare for a record of the second coded form to survive, the tripartite nature of telegraphic transmission must be remembered when dealing with cabled communications.
The editor generally has at hand either the sender’s manuscript or the recipient’s translation of Morse code. While the first represents an author’s final intentions, the second is evidence of what was actually communicated to the recipient. The puzzle is compounded in coded telegraph transmissions, where the possibilities for corruption reach extraordinary levels. The files of World War II military leaders’ headquarters are filled with requests for retransmission of garbled coded materials. The George Marshall and Dwight Eisenhower editions generally contain paraphrased texts of such messages as they reached the recipient’s office, because these, at least, are evidence of what each commander learned from the cables.

In general, most editors will choose among the possible source texts from the standpoint of their own editions. When draft or final versions of outgoing cables survive in the hand of the edition’s subject, they will be chosen as the paraphrased texts of such messages as they reached the recipient’s office, because these, at least, are evidence of retransmission of garbled coded materials. The George Marshall and Dwight Eisenhower editions generally contain requests for extraordinary levels. The files of World War II military leaders’ headquarters are filled with requests for retransmission of garbled coded materials. The George Marshall and Dwight Eisenhower editions generally contain paraphrased texts of such messages as they reached the recipient’s office, because these, at least, are evidence of what each commander learned from the cables.

In general, most editors will choose among the possible source texts from the standpoint of their own editions. When draft or final versions of outgoing cables survive in the hand of the edition’s subject, they will be chosen as the paraphrased texts of such messages as they reached the recipient’s office, because these, at least, are evidence of retransmission of garbled coded materials. The George Marshall and Dwight Eisenhower editions generally contain requests for extraordinary levels. The files of World War II military leaders’ headquarters are filled with requests for retransmission of garbled coded materials. The George Marshall and Dwight Eisenhower editions generally contain paraphrased texts of such messages as they reached the recipient’s office, because these, at least, are evidence of what each commander learned from the cables.

2. Electronic Records

This volume is confined to the experience and comments of American scholarly editors, who have not yet addressed the special problems created by electronic documentary records. While we have suggested some ways that e-mail and other products of a personal computer or laptop may further complicate an editor’s life (see pp. 95–96, above), we hesitate to go further. Instead, we will repeat the commonsense words of G. Thomas Tanselle in “Textual Criticism and Literary Sociology”: “Computerization is simply the latest chapter in the long story of facilitating the reproduction and alteration of texts; what remains is the inseparability of recorded language from the technology that produced it and makes it accessible” (88).

IV. Conclusion

Choosing source texts is only the first step in a long series of critical judgments an editor makes on the road to publishing an edition. Once that text has been identified, an editor must cautiously move from the “original,” the document that holds evidentiary value, toward some accessible version that will serve the needs of a wide audience. Next comes transcription, the conversion of those source texts into physical forms that can serve the purposes of a modern editorial project.

Suggested Readings

The proper selection of the source text has largely been ignored in monographic literature by editors, although reviewers of documentary editions have not been so negligent. See, for instance, the reviews of the Naval Documents of the American Revolution series in William and Mary Quarterly 22 (October 1965): 660–63, and American Historical Review 77 (June 1972): 831. For useful discussions of the techniques of classical scholars in applying filiation to ancient texts, see Paul Maas, Textual Criticism; and Paul Oskar Kristeller, “The Lachmann Method: Merits and Limitations.”

Although nearly thirty years old, Leonard Rapport’s “Fakes and Facsimiles: Problems of Identification” is still useful.


For a history of one man’s experiments with labor-saving devices for correspondence, see Silvio A. Bedini, Thomas Jefferson and His Copying Machines; for a broader study, see Barbara J. Rhodes and William W. Streeter, Before Photocopying: Art and History of Mechanical Copying, 1780–1938.

Useful essays on the relationship between bibliography and editorial problems appear regularly in Studies in Bibliography. Vol. 3 (1950–51), for instance, includes Fredson Bowers’s “Some Relations of Bibliography to Editorial Problems.” The same scholar’s “The Function of Bibliography” remains an able introduction to the topic. G. Thomas Tanselle ably discusses bibliographical problems raised by nineteenth-century authors in “Bibliographical Problems in Melville.” Jennifer Tebb’s “Print and American Culture” is a good introduction for novices.

The state of the art of print bibliography for government records is discussed in review essays by Ted Samore and Stewart P. Schneider in Government Publications Review, and in Martin Clausen’s review essay “Revisiting America’s State Papers, 1789–1861: A Clinical Examination and Prognosis.” The most pointed critique is Edwin Wolf’s “Evidence Indicating the Need for Some Bibliographical Analysis of American Printed Historical Works.” For analysis of modern editions of government records in the United Kingdom, see Christopher Kitching, “Record Publication in England and Wales, 1957–1982.”

The reading lists and handbooks of the CEAA and CSE, of course, should be consulted for any edition that presents the text of a published work.

For special problems of diaries and journals, Leonard N. Neufeldt discusses the matter on a theoretical level in Neopragmatism and Convention in Textual Editing, with Examples from the Editing of Thoreau’s Autograph Journal," as does Klaus Hurlebusch in “ ‘Relic’ and ‘Tradition’: Some Aspects of Editing Diaries.” For a rare account of one well-known

Eugene A. Nida offers general reflections on the problems of foreign-language translation and editing in “Editing Translated Texts.” In the realm of practical examples, the quinquicentennial of Columbus’s voyage inspired a spurt of translations of significant works from Spanish with editorial apparatus that borrows from traditions of documentary and textual editing. See, for example, Charles Hudson and Paul Hoffman’s edition of *The Juan Pardo Expeditions* and the review of this work by David Henige in *Documentary Editing*.

Ralph E. Weber surveys important elements of cryptography in *United States Diplomatic Codes and Ciphers, 1775–1938*. The notes in *Adams Family Correspondence*, 4:viii–ix, 393–99, comment on problems raised by the cryptographs employed by one group of correspondents. This *Guide’s* Web site carries generous samples of coded documents and the evolution of their clear text translations.

For a discussion of the range of problems created by oral history methodologies, see J. A. Prögler, “Choices in Editing Oral History.”

Morris Fishbein offers useful general reflections in “The Evidential Value of Nontextual Records: An Early Precedent.”
Chapter Four

Transcribing the Source Text

I. Who Does the Transcribing?
II. General Rules for Transcription
III. Special Transcription Methods
IV. Transcription Forms and Control
V. Factors Affecting Transcription Methods
   A. Methods of Recording a Text
      1. Handwritten Source Texts
      2. Typewritten Documents
      3. Printed Source Texts
      4. Nonverbal Documents or Document Elements
      5. Maps
   B. Transcribing Different Types of Records
      1. Correspondence
      2. Business and Financial Records
      3. Professional and Technical Records
      4. Government Records
      5. An Author’s Works
      6. Journals and Diaries
      7. Records of Oral Communications
   C. Transcribing an Earlier Transcription
VI. Conclusion

Suggested Readings

In the literature of documentary editing there is often confusion over what is involved in transcription. It’s sometimes assumed that an editor’s transcription practices—standards for rendering the source text’s words, phrases, inscribed symbols, or sounds into a format that can be used in editing—are reflected precisely in the editorial text that is finally printed. This is seldom the case. The process described here is merely the initial conversion of the document’s contents to a rough but accurate transcription that will become editorial working copy. This may bear little resemblance to the editorial text that is finally published. Sources should be transcribed as literally as possible, even when the editorial text will be a heavily emended clear text. Unless that original transcription is literal, it will be difficult to provide a clear record of the changes made by the editors—changes that may need to be recorded in the edition’s editorial apparatus.

Moreover, literal transcription is far easier and more efficient for the transcriber. Most American editorial experience is with documents expressed in words, and that will be our emphasis here. Translating handwritten or typed pages or recorded speech to typescript is challenge enough without the additional burden of mastering lists of preferred editorial emendations and acceptable corrections. Modern word-processing equipment eliminates arguments that literal transcriptions demand costly rekeyboarding once emended for publication. Word processing allows the editor to enter such changes when editing is complete and to produce new, clean copy as needed.

Almost inevitably, the initial transcripts will reflect some policies of emendation. Before transcribing begins, most editors reach decisions on retaining or suppressing such details as the position of the date and place lines in letters; the treatment of salutations, closings, signatures, and paragraph indentation; the standardization of formal headings in public papers; and the treatment of addresses and endorsements. \{p.113\} Transcribers can be instructed in these matters, but the instructions must be recorded on paper or in computer files, and they must make perfectly clear to the transcriber what forms are to be expanded, which types of misspellings are to be corrected, and so on. The editor without the time or inclination to prepare such a transcription manual should ask transcribers to type what they see or hear in the sources.

Theoretical as well as practical considerations argue for a careful record of transcription methods. Even solo editors responsible for their own transcribing are well advised to keep such a log, for transcribing sources is a learning process. As the editor-transcriber moves through the collection, he or she will inevitably learn to recognize meaning in patterns of inscription that earlier seemed meaningless or baffling. Only by keeping track of their hard-won knowledge of what matters and how it is to be translated can editors hope to be consistent or accurate.
I. Who Does the Transcribing?

In earlier decades, it was assumed that the editor of an edition or staff members directly under her or his direction would keyboard or otherwise transcribe source texts to produce working editorial texts. As in every other area, twenty-first-century methods challenge this assumption.

In the early 1990s, grants from the Packard Humanities Foundation made it possible for editors of the “Founding Fathers” group (the Papers of Franklin, Washington, Jefferson, Adams, and Madison) to mount an effort to complete the transcriptions of documents earmarked for their editions, and to ensure that the machine-readable transcriptions would be available to the projects in searchable form. Some of these projects assigned initial transcription to off-site workers, who received photocopied images of the original documents along with detailed instructions, but who keyboarded the new texts without further direct consultation with the editors themselves. These transcriptions, of course, received careful verification from the editors. Editors were pleasantly surprised at the results, and today several projects use service bureaus for preparation of draft transcriptions. Some of the work is done not only off-site but offshore.

This method demands, of course, even more careful planning of transcription policies and procedures. The Jefferson Papers Retirement Series, for instance, created a basic template, which included rudimentary tags for transcribers working for a vendor whom they employ.

Similarly, OCR (optical character reader) scanning rather than rekeying may be used to prepare a machine-readable transcription of a printed source text. Should the source text present an unusual or archaic typeface, an entirely new transcription may prove more economical. In either case, the transcription must be viewed as a very rough one, demanding serious verification and proofreading by senior staff.

The editors of the Papers of Eleanor Roosevelt have introduced a novel and useful assignment of transcription responsibilities for their edition. Each document is transcribed independently by two different editors. Computer comparison of the two products identifies letters, words, or phrases that demand special attention. Once necessary corrections reconcile the two versions, the transcription is proofread against the source text by still another editor.

Whenever possible, foreign-language materials should be transcribed by someone familiar with the language in which the texts are written. Better still, they should be transcribed by someone familiar with usage of the historical era in which the documents were created. Orthography, spelling, and usage in eighteenth-century Spanish, French, and Dutch are quite different from modern forms of these languages, just as they are for English. While proofreading transcripts against the source texts will catch errors, it’s easier on all concerned to have as reliable a transcription as possible at the beginning.

II. General Rules for Transcription

“Exact copying,” Edith Firth reminded her fellow editors, “involves a fair amount of hack work.” There is no way to make transcribing fun, but there are some simple rules that can make the process less painful and avoid duplication of effort and unnecessary rekeyboarding.
Whenever possible, editors should consult their publishers before transcribing begins. Modern computer typesetting makes the retention of details of inscription like raised letters or archaic symbols much easier, but different programs use different codes for these details. Employing the codes appropriate to the publisher’s system from the beginning reduces the number of corrections and changes to be made later in the editorial process.

When the edition is not a single-source project and has created a database to maintain control of cataloged materials, that same tool can aid in regulating and recording the transcription process. Fields in each document’s record can be used to enter the name of the transcriber and the dates on which transcription, emendation, and proofreading have occurred as the new translation of the source moves through the editorial process.

At the risk of insulting our readers’ intelligence, we offer these commonsense rules:

1. Happily, we can ignore the problems of bygone days created by typewritten carbon copies of transcriptions. With computer equipment, it is still wise to maintain a separate file of proofread but unemended transcriptions and to store backup disks carefully. The editor’s working copy may be a separate file with a different file name or a printout made from that archival file.

Computer disks have not been tested for longevity, and even mainframe computers can lose stored materials. The truly cautious editor may wish to print out two copies of each computer-created transcript, one for editorial use and the second as insurance should the computerized storage facility fail. As editorial work progresses, backup files should be made for each level of revision, and disks with these files should be stored securely and separately. Modern equipment makes the process far easier, and most editors can do automated backup of all their files, while servers can be backed up remotely. With a content management system, all versions are backed up every time a change is made, so that editors don’t even have to think about naming files and ferreting out the location where the most recent copy is saved. The biggest issue for projects may be a system of naming the most recent version of the file and recording where it is stored. Content management systems do this automatically.

This decision is determined by whether editors work on desktop PCs where they store all of their material or whether the project has a central server for storage and backup. When editors store backup materials on their individual computers, there is no secondary backup, and other staff members may not be able to access the most current version of these files.

2. Transcribers should be instructed to insert a special code in the transcript for material that they cannot read in the source. The editor can then do a computer search for this code in every transcription.

3. Transcribers should give their work a preliminary review before filing the copy.

4. The physical format of the printout should reflect its intended use as working copy for an editor. Like any such manuscript, it should be double-spaced, with generous margins on all four sides.

5. Many of a word processor’s convenient features must be turned off for the duration of the project. These include justified right-hand margin. Transcriptions should introduce no new punctuation in the form of end-of-line hyphenation, and ragged right-hand margins on the typescript are a small price for accuracy. More exotic features like AutoCorrect and automatic capitalization must also be disabled.

6. The transcriber and editors should not use a computer font like Times Roman that is proportionally spaced. A nonproportional or “fixed” font like Courier will make it easier to see spacing problems in the transcription.

If transcription can wait until the editorial collection is complete, it’s often more efficient to assign each transcriber a chronological sequence of source texts or a group of topically related sources. The transcriber can thus become expert in the problems peculiar to the time period or theme in question, as well as with the correspondents involved.

### III. Special Transcription Methods

The editor who aspires to the CSE emblem must adopt methods to facilitate the preparation of records of editorial emendations required by that agency. Literal transcription of words, of course, is a prerequisite, but the transcriber must also transcribe documents in facsimile fashion in certain areas:

1. The transcriber should preserve the end-of-line breaks of the original document so that the editor can create an accurate record of the author’s end-of-line hyphenation.

2. In transcribing paginated source materials such as notebooks or journals, it is useful to transcribe not only line for line but also page for page. A new page in the document will call for a new transcription sheet so that the editor will have a convenient record of such breaks.

3. The lines of transcription, as well as their pages, should be numbered so that the editor can prepare textual notes without the use of footnote numbers. Most word-processing software offers line numbering as a routine option.

Documents that are entirely in shorthand, codes, or ciphers are usually best left untranscribed until they can be translated. Keyboarding page after page of numerals or alphanumeric codes is an unreasonable and unconscionable trial for a transcriber. Transcribing shorthand symbols would be impossible. Proofreading the results to present to the outside expert or editor who will do that translating would be a ridiculous waste of time. A duplicate hard copy of the shorthand or coded pages will be working copy for the translator. Results of the translation can be...
entered interlinearly, and the results emended according to the edition’s rules for such passages.

**IV. Transcription Forms and Control**

In past years, specially designed transcription sheets were almost a necessity for projects that aimed for CEAA/CSE approval. Those sheets carried line numbers along one margin, and their standardized headings included spaces in which to record the completion of special procedures demanded by the CSE: verification, or perfection of the transcript against the original document; completion of requisite proofreadings; and so on. Computer-assisted systems are a special boon here, for such details of format can be entered automatically.

Even projects without expectations of CSE approval learned the value of establishing appropriate formats for their transcriptions, and computer equipment makes the generation of special headings simple. The transcription’s heading should indicate the document’s title, preferably in the form that will appear in the printed edition. The transcription should also bear the project’s code for the location of the original of the source text, as well as the identifying number assigned to that version. Each succeeding page of the printout should repeat the document’s date and identifying source information as well as the new page number. Word-processing equipment makes creation of such headers easy and routine.

Each transcription will require its own unique identifying number. For a project that has collected materials, this can be based on the accession number assigned the photocopy, image file, or recording. When an edition is based on an archival collection, and no identifying numbers have been assigned to the sources at the beginning, the word processor can be programmed to assign sequential numbers to each transcription. Sequentially assigned transcription numbers may also be called for when source texts are transcribed from microfilm reels.

Use of automated equipment for logging transcriptions has many advantages. A system that links data in the control file allows easy and accurate transfer of that information to the transcription record for the same document. Editors can design in advance any forms they or the transcribers will need and can store these formatted headings. When needed, they can be called up by a word-processing macro command. It will be easy to match transcriptions of variant versions of a text with their sources, and if several transcribers are at work simultaneously, each can access the central database to determine which sources need attention and which are already being keyboarded. Later reviews of the transcription can be indicated in fields for the initials of the editors and the duties performed, such as verification, annotation, final proofreading, rekeyboarding, and so on. Database logs of transcription and other editorial processes eliminate the need for hundreds of file drawers and notebooks in editorial projects.

Linking transcription files to the project’s control file system may be the point at which many editors rethink early decisions to use content management systems, relational databases, flat-file databases, or still less sophisticated computer methods. An edition with a few hundred transcriptions may be able to manage without a relational database, but larger ones may have to investigate one. Projects in which a number of editors and transcribers work simultaneously on transcriptions and notes may need networking equipment and a content management system so that each member of the staff can access information when necessary. While access should be open to all, there should be strict control over the entry of information in the record of a document’s progress through publication, to ensure consistency and currency of record keeping.

Nonarchival editorial projects with large staffs and no facilities for networking their computer workstations often find it convenient to file a copy of each transcription’s printout in the file that holds the photocopied source. This provides all concerned with easy, logical access to the transcriptions in progress. Naturally this is impossible when the source texts are original manuscripts that could be damaged by chemical reactions to the transcriptions’ paper or inks or when the source texts are on microfilm. Other projects simply store hard-copy printouts of transcriptions in notebooks, file drawers, or boxes in the chronological or topical order in which they will appear in the projected publication.

**V. Factors Affecting Transcription Methods**

The same factors that determine the choice of source texts dictate the appropriate method of their transcription. The first are the methods by which the source texts were originally inscribed, printed, or otherwise recorded. The second are the forms of documentary evidence that the sources represent. Only after editors have mastered the intricacies of all the methods used to inscribe the sources in their editorial collections can they begin to rule out textual practices that would distort the details of those sources. The form of documentary evidence represented by each source text—letter, diary, state paper, scientific treatise—will raise another set of questions. The editor should consider the method of communication that each embodies before reaching a final decision on its treatment in the new edition. Whether the editor is his or her own transcriber or delegates this task to a staff member, work will move more smoothly if transcription follows carefully thought out decisions that can be applied consistently rather than choosing a procedure so early that its rules must be changed over and over again.
A. Methods of Recording a Text

The means used to record documentary evidence are often the primary factor that an editor considers in determining standards of transcription and in settling on the final textual method. Certainly it is always the first factor to be weighed. Editors stress the need to master the inscriptive history of unprinted documentary sources, as well as the history of printed materials that will be the basis of a documentary edition. While choosing among competing source texts, the editor should learn as much as possible about the methods by which these variants were produced.

The knowledge may run the gamut from the practices of a group of clerks in the same office to the peculiarities of a particular eighteenth-century letterpress or the details of book production or newspaper composition. For comparable discussions of the inscriptive and organizational history of nonprint sources, see the Eisenhower Papers, “Essay on Primary Sources” (9:2259–73) and “Notes on Primary Sources” (11:1521–23, 13:1507–11).

Unfortunately, there are few secondary studies of modern methods of nontypeset inscription to which an editor can refer. In European medieval studies, there is the recognized academic discipline of “diplomatics,” which offers systematic studies of record-keeping methods of particular groups of clerks and administrators. There are no such formal courses from which to learn how the secretary of the American Continental Congress maintained his records in 1785 or how elementary students were taught to standardize the forms of personal correspondence in 1830. Each editor must learn these methods for him- or herself. Often the best introductions appear in existing scholarly editions, and the novice should consult one that drew on sources similar to the one he or she hopes to publish. Documentary editors who have survived the agonies of learning the intricacies of American inscriptive history frequently share this wisdom in their own edited series. The general rules to be drawn from those experiences follow here.

1. Handwritten Source Texts

Anyone who has transcribed handwritten materials or proofread the resulting copy knows that no typeface can reproduce all the subtle distinctions in the originals. Any typed or printed transcription of such a source is a critical one, because it silently incorporates dozens of editorial judgments and decisions. The editor cannot indicate every instance in which experience enables her or him to recognize a scrawled mark as a period instead of a comma or an m. Often the experienced editor exercises such judgment quite unconsciously. Knowing this, the editor should choose a textual policy whose conventions do not conceal additional subtleties in a source that is already being transformed from script to type.

Here we refer you to an exceptionally useful article in the 1999 number of Studies in Bibliography, “A System of Manuscript Transcription,” by David L. Vander Meulen and G. Thomas Tanselle. We thank them, first of all, for a sensible and practical definition of the process of transcribing handwriting: “The effort to report—insofar as typography allows—precisely what the textual inscription of a manuscript consists of” (201). Beyond that, the authors propose an easy-to-use system of recording details of a manuscript that includes neither complicated symbols nor exotic typographic coding. These descriptive codes, of course, can be changed later in the editorial process when other symbols need to be substituted.

Only a careful analysis of the sources at hand enables the editor to decide which conventions will least distort the source. Some writers may employ one method of punctuation in correspondence with close friends and another, more formal, system in letters to strangers and professional colleagues. Obviously, standardizing such marks would conceal important evidence of the author’s relationship with each correspondent. Still other writers vary the length of a standard mark of punctuation such as the hyphen for different functions. When such differences are identified as important, they are marked or coded in the keyboarded transcription to reproduce the author’s consistent, if idiosyncratic, patterns. Individual patterns of capitalization, the use of contractions, and care or careless in spelling often provide unexpected insights into an author’s state of mind. The editor cannot responsibly ignore such slips of the pen until certain that they do not represent significant patterns in the writer’s orthography.

For this reason, it’s useful as the project progresses to create a profile for each writer frequently represented in a collection of correspondence or other groups of materials by separate authors. The profile can (and probably should) include images of examples of each author’s idiosyncrasies. There should be just as careful a record of editorial decisions on how to deal with these peculiar styles of punctuation, capitalization, or formation of letters of the alphabet.

It is difficult enough to discover, in a single individual’s handwriting, patterns that change with a writer’s progressive education or with advancing age. The possibility of distinguishing and analyzing such patterns in a group of documents representing a number of authors is far more challenging. Editors who are tempted to impose arbitrary, often modern notions of punctuation or spelling to normalize the two sides of a handwritten correspondence—journals that may have been kept by a dozen scribes, or reports and papers from a hundred
contributors—should consider the consequences of such a practice. Such standardization can mean the loss of information about their subjects that readers of the edition might find informative.

Imagination and inventiveness are called for here. One of the thorniest transcription issues faced by editors of the Papers of Margaret Sanger was her patterns of capitalization. Sanger’s script makes every word that begins with *th* look as though it’s capitalized; she formed initial *c’s* larger than other letters of the alphabet; and she always capitalized the word “Doctor.” By comparing handwritten materials with Sanger’s use of capitals in typewritten correspondence, it was possible to uncover her true authorial intentions. Thus the edition capitalizes words beginning with *c* and *th* when appropriate in standard usage. “Doctor” remains capitalized as Sanger had written it; this practice is evidence of her personal style. When doubt persists, the editors follow standard usage, since Sanger generally did not capitalize words for emphasis, preferring double or triple underlining in such cases.

The process of translating handwritten text to print once almost necessitated standardization. Such alterations to details of inscription as lowering superscript letters were necessary to keep the costs of producing the print edition within bounds. Happily, there’s no longer any compelling reason to adopt such regularizing conventions in the text solely for the sake of economy. Modern computer-assisted technology, including word-processing programs and XML tagging, makes more faithful transcription and publication of handwritten sources in a print edition more affordable. Transcribers can record and reproduce superscripts and subscripts, dashes of various lengths, and other details, producing transcriptions of source texts as literally as their legibility allows. It will be up to an editor, of course, to decide which of these details will appear in the published edition; but computer composition of printed pages makes retention of such details a practical possibility, and traditional emendations like expanding the “tailed p” are no longer an economic necessity.

Printers seem more liberated than ever in their ability to tap into Unicode and import graphic representations for characters for which there is no known symbol, and the users of documentary editions have proven remarkably adaptable in digesting such details. Several well-established projects, such as the Papers of Thomas Jefferson, the Papers of Benjamin Franklin, and the Adams Papers, have reversed earlier decisions on standardization of inscriptive detail with no signs of mutiny from compositors, publishers, or readers. Online editions may need to show more caution: editors of electronic editions must remember that the limitations of HTML display will be significant factors owing to variations in the Web browser being used.

The age and condition of the manuscripts that bear the author’s script may make even a rough transcription of their contents from surrogate images difficult. In such cases, editors should verify transcriptions against the originals before even beginning an assessment of the importance of each detail of inscription. They may have to refer to those originals again and again during the period in which they labor to establish their texts for the edition, sometimes resorting to ultraviolet light to pick up faded ink. If access to the originals is impossible, the tricks of digital imaging available in programs like PhotoShop or Adobe Acrobat can help by enlarging or rotating images or even reversing the direction of text for the deciphering of ink-blots impressions of words.

### 2. Typewritten Documents

Only in recent years have documentary editors confronted the textual problems created by modern office machines. A recipient’s copy of a signed typewritten letter or other communication presents few problems: no matter what its flaws, it may be considered to represent the author’s final intentions. If the typescript carries the author’s handwritten corrections or revisions, these can be indicated by textual symbols, special typefaces, or annotation. Often the same methods that the edition adopts to meet the needs of handwritten sources can be modified for typed sources. In the absence of such symbols, the editor instructs the transcriber to key in such emendations in a special typeface reserved for handwritten additions to typed pages or to flag them with a special code. If such authorial revisions are rare, footnotes can explain these occasional details of inscription. Whatever the final decisions may be, the initial transcription should indicate all these details, so that they can be recorded as the formal textual apparatus dictates.

An uncorrected carbon copy creates new problems. Faced with file drawers full of such uncorrected copies of outgoing correspondence, the editors of the Woodrow Wilson Papers compared surviving carbon and ribbon copies of the same letters and found that Wilson himself had corrected the originals. They concluded that the ribbon copies of these letters had been corrected before dispatch, and the texts were emended accordingly. The Eisenhower edition took a similar approach. After analyzing Eisenhower’s patterns in proofreading outgoing typed correspondence, the editors adopted a twofold policy of emending errors in typed letters for which only the retained carbon copy survived. Mere typographical errors and errors of spelling, which Eisenhower routinely corrected, were emended silently, but editorial corrections of misspellings of proper nouns are accompanied by a footnote recording this emendation.

Editors of late nineteenth-century typescripts should exercise special caution. Any assumptions about “typographical” errors here must take into consideration the specific keyboard of the machine employed. Decades passed before these keyboards were standardized, and emendations can be made only with full knowledge of the particular instrument involved. Contemporary photographs of the subject’s office, bills, or correspondence ordering
office equipment help the editor to identify the machine employed. If the machine's manufacturer is no longer in business under the same name, the editor may have to embark on a new career as corporate historian to find a source for an image of the typewriter's keyboard and determine patterns of likely typographic error. If no such records survive, the editor may have to abandon plans to correct such errors wholesale, for the identification of their patterns may be well nigh impossible.

Transcription and textual policies for typed sources should reflect not only that medium of inscription but also the basic nature of each source. A typed draft letter with handwritten or typed revisions has requirements quite different from a nonauthorial typed transcription of a handwritten source. Editors who ignore this rule and lump together typed letters, retained carbon copies, typed transcriptions, and printed transcriptions of lost sources into one textual category compromise their product before the process of transcription even begins. Typewriting is merely an exceptionally legible form of inscription by an individual, and the appearance of a typed document should not mislead an editor into ignoring other factors that may dictate editorial treatment.

3. Printed Source Texts

The editor who establishes the text of a published work for the purposes of a critical edition may sometimes legitimately depart from the substantive and accidentals of any single surviving printed copy of that writing. But the documentary editor views print as merely another way to inscribe a document, and the surviving copies of a given source must be evaluated to determine which is the best document, the one that is a unique, authoritative source with evidentiary value. Once that source has been identified, it should be transcribed literally and published without emendation, conflation, or the other heavy artillery of textual editing.

4. Nonverbal Documents or Document Elements

The editing of completely nonverbal documents was not addressed until late in the twentieth century by editors in the NHPRC and CSE traditions. Earlier, many editors had to solve the problem of how to present scattered authorial illustrations and even doodles that were part of a more conventionally inscribed source. When these elements had recognized equivalents in type, the transcriber entered such ready-made symbols from the printer's font in good conscience. The editors of the Hawthorne Notebooks translated the author's hand-drawn marginal “fists” into \(\mathfrak{f}\), the existing standardized type unit for this symbol (called a “fist graphic”). The editors of the Emerson Journals employed the conventional typeset caret (\(^\wedge\)) for Emerson's mark for interlineation; and Mark Twain's editors translated Clemens's autograph proofreading symbols in his Notebooks and Letters into the equivalent printer's symbols.

Unfortunately, writers seldom confine themselves to such easily translated marks. Some authors, like William Blake, were artists as well, and neither critical nor documentary editions of their works would be complete without reproduction of the images that are an integral part of their texts (G. E. Bentley, "Blake's Works as Performances: Intentions and Inattentions"). With modern print publication techniques, the photoreproduction of sketches, drawings, doodles, or maps in a printed volume of documents is far less expensive than it once was, but the method is still limited to occasions when the literal presentation of such elements seems essential to understanding the documents. (See, for instance, the facsimile of a Jefferson diagram in Papers of Thomas Jefferson, 22:74.)

Among documentary editors, scholars in the field of the history of science and technology seem more ready to explore the transcription of nonverbal materials—and to report results to their peers. In the September 1984 issue of Documentary Editing, Reese Jenkins and Thomas Jeffrey of the Edison Papers pointed out that most editors had actually retreated from these challenges. Julian Boyd, they noted, relegated Jefferson's nonverbal documents such as architectural sketches to a separate series—one that has not even been started although the Jefferson project is well into the second half of its first century. The Franklin Papers excluded nonverbal materials altogether except as illustrations. And although the Emerson editors reproduced Emerson's easily translatable symbol for an insertion, they excluded all his pen and pencil sketches from the Journals and Miscellaneous Notebooks (Jenkins and Jeffrey, "Worth a Thousand Words: Nonverbal Documents in Editing," 7).

In happy contrast, a more recent examination by the editors of Charles Darwin's Correspondence shows the increasing ease with which they can deal with the phenomenon of “extra-text” objects such as tables, formulas, photos, and diagrams. In 2004 they reflected on the advances in print technology that allow them to abandon photographs of photographs or even tracings of photos in favor of high-quality digital scans of original illustrations for their editions (Duncan Porter and Alison Peran, “Editing Textual and Extra-Textual Materials in Charles Darwin's Correspondence”).

Still, as Jenkins and Jeffrey first noted more than twenty years ago, editors of papers in the arts and sciences have a better record for the treatment of nonverbal documents than editors of the papers of a statesman or a novelist. The graphic records of great artists and illustrators introduce a new dimension to the problem of editorial transcription. Reproduction of such records for author-artists must meet technical standards far higher
than those adequate for the comparatively simple line drawings of a scientist or the doodles of an author in a notebook. American documentary editors first addressed the problems of reproduction techniques in editions of the papers of figures like Benjamin Henry Latrobe and Charles Willson Peale. In these two examples, the costs of reproducing all sketches as a physical part of the printed pages would have been prohibitively expensive. Instead, separate microfiche supplements were published to provide students with a comprehensive reproduction of graphic materials, while the book editions reproduced selected drawings and sketches on plates inserted as illustrations to the printed text.

Web-based publication makes unnecessary many such agonizing decisions about inclusion and exclusion. The online archives of William Blake (http://www.blakearchive.org/blake/) and Dante Gabriel Rossetti (http://www.rozettiarchive.org/) provide easy worldwide access to an astounding variety of images. For most editors, image editions like these will be unnecessary. However, almost all will deal with an instance when the nature of a sketch or drawing or marginal doodle is essential to the reader’s understanding of a document’s verbal elements. Even when literal photoreproduction of the sketch is not required, some editorial action is necessary. Transcribers will often be the first to notice such symbols, and there should be a code that can be entered to bring such images to the attention of an editor. If they are comparatively rare and isolated, the editor may choose to describe them in a note adjacent to the text or simply import the graphic into the text. A recent example of this is the marginal reproduction of Thomas Jefferson’s sketch of a chair in Papers of Thomas Jefferson: The Retirement Series 1:44.

If such graphic materials are common to the group of documents being edited, the editor must find another course. Some may need to imitate the Latrobe and Peale editions by issuing facsimile image supplements where the reader can consult the unprinted and even unprintable elements of the documents. The diagrams and schematic drawings of great figures of science obviously fall into this category. Joseph Henry’s sketches and diagrams were conscientiously reproduced in the Henry Papers, as are such elements in laboratory notebooks and other technical sources in the Papers of Thomas A. Edison and Albert Einstein. Here the “transcription” of the source text is a photoreproduction, as much a translation of the original as a typescript is of a manuscript. Even when literal photoreproduction of the sketch is not required, some editorial action is necessary. Transcribers will often be the first to notice such symbols, and there should be a code that can be entered to bring such images to the attention of an editor. If they are comparatively rare and isolated, the editor may choose to describe them in a note adjacent to the text or simply import the graphic into the text. A recent example of this is the marginal reproduction of Thomas Jefferson’s sketch of a chair in Papers of Thomas Jefferson: The Retirement Series 1:44.

Much of the most useful literature on this subject comes from the editors of the Edison edition—Reese Jenkins’s “Words, Images, Artifacts and Sound: Documents for the History of Technology,” and Robert Rosenberg’s “Technological Artifacts as Historical Documents”—but the discussion has now enlarged significantly. Jenkins and Rosenberg examined the relationship between artifact documents and recorded sound. Jenkins conceded that this might seem a special concern for historians of technology: “To ‘read’ the nonverbal artifactual record is to read what the inventive enterprise was all about. In the case of the Edison papers, to have a documentary edition without the artifacts would be equivalent to having a marriage without the bride” (45). But Jenkins and his colleague Jeffrey argued persuasively that all editors must now confront the challenges of nonverbal documents: “While acknowledging that our Western historical tradition was rooted from the beginning in words, we must also realize that the documents of the past, and, increasingly, of the present and future beckon for recognition and challenge us with diverse languages expressed in symbols other than words” (Jenkins and Jeffrey, “Worth a Thousand Words: Nonverbal Documents in Editing,” 8).

5. Maps

While most documentary editors use maps only as helpful illustrative adjuncts to their texts, some have had to confront the question of editing maps as historical documents. Stephen E. Wiberley Jr. offers a detailed discussion on the cartographic equivalents of facsimile publication, emended transcriptions, and clear text in “Editing Maps: A Method for Historical Cartography.” Wiberley’s discussion centers on his experience with the Atlas of Early American History. A more strictly documentary approach to maps as evidentiary sources was pursued in the Atlas volume of The Journals of the Lewis and Clark Expedition.

B. Transcribing Different Types of Records

The method of transcription should reflect not only the physical format or history of the document but also the type of documentary record that each source represents. The traditional literary distinction between a writer’s public and private writings is of little use to the documentary editor. Unpublished letters have influenced the course of history.
State papers not set in type in their author’s lifetime have shaped the thinking of legislators and executives. Confidential technical reports have radically affected the history of science. Each of these is private under the literary scholar’s definition, but all were composed for an audience, and all derive their historical importance from the influence they exerted on those contemporary readers.

Thus documentary editors long ago abandoned any methodological division between public and private writings and, instead, examine how these source texts functioned as agents of communication. Any edition of such sources should strive to preserve the communicative intention or effect of the original. Recent scholarship in the humanities as a whole, of course, seconds this decision.

With documents, the editor may also have to modify notions of what constitutes the appropriate contents of an edited text. Elements in the source that could safely be ignored by the editor of an author’s published works may be an integral part of a source’s documentary contents. Such accidentals as capitalization, indentation, and spacing may perform important functions in the design of certain modes of communication. And for some documents, even nonauthorial contributions must be considered part of the source’s evidentiary contents.

1. Correspondence

The literature on editing correspondence once focused on the need to make the letters of an earlier age “readable”—even “enjoyable”—for a modern audience (see, for instance, Robert Halsband, “Editing the Letters of Letter-Writers”; and William Ferraro’s review of Mark De Wolfe Howe’s Home Letters of General Sherman). Today, discussions of the textual needs of correspondence are considerably more sophisticated, even if agreement on methods remains distant (see Robert Stephen Becker’s report on his survey of editions of nineteenth- and twentieth-century letters in “Challenges in Editing Modern Literary Correspondence: Transcription”).

The move toward clear text in CSE-approved editions of correspondence was justified, in part, by the claim that the literary critics who formed a large part of these volumes’ audience would not wish to be distracted by textual symbols or numbered footnotes pointing out details of the original. Even exponents of silently emended, expanded transcription concede that readers soon become used to archaic and idiosyncratic usage in a documentary edition, and Ernest W. Sullivan has pointed out that literary critics were ill served by the practice of standardizing texts in editions of personal correspondence (“The Problem of Text in Familiar Letters”). Steven Meats lectured the editors of the correspondence of literary figures: “As a general rule, the less editing (that is, the less editorial emending and altering) done to the text of letters, the better the job of editing. A letter is, after all, a primary historical document; one might even call it a ‘fact.’ In any case, silent emendation in the editing of letters should be severely restricted” (“The Editing of Harold Frederic’s Correspondence,” 38). As the editor’s compulsion to conventionalize, normalize, or otherwise emend the text of letters came under serious attack, editors of the private writings of American literary figures moved further and further away from the aim of a clear text.

It is not the business of documentary editors to introduce new readings into a documentary text for the sake of historically unrealized clarity. When letters survive in some preliminary form that reflects revision by the author, they may eventually require the application of textual methods appropriate to genetic elements (see pp. 188–91, below) that reflect the evolution of a text, but the initial transcription should be literal. When a letter survives in the form in which its author dispatched it to its addressee, it deserves the same treatment in transcription. Literal transcription of hastily scrawled and ill-proofread recipients’ copies of letters may be untrue to the author’s final intentions, but such transcription is true to the documents. Equally important in documentary terms, it represents the form of the letter that influenced its recipient. Unnecessary emendations will make the editorial text of any letter useless as evidence either of what its author wrote or of what its addressee read. Even editors who are their own transcribers should make initial transcriptions of such sources as literal as possible. Any policy of emendation for letters adopted later should be as conservative as their method of inscription allows. A detailed back-of-book textual record grants no license to violate this rule.

If any corrections are to be imposed on the author’s prose, they will come after the editor has established the edition’s final textual policies. And all editors should remember that most people do not write nonsense. Words that are incomprehensible to the transcriber may in fact be archaic words, creatively spelled words, or foreign terms. “Correcting” apparent mistakes will take the editor down the wrong path. A correction of such words should only take place after all the above possibilities have been eliminated. Editors should consult the Oxford English Dictionary (so easy in the online edition), and foreign-language dictionaries, and consider alternative spellings or similar pronunciations. (For a compelling discussion of the need to remember the effect of punctuation on oral patterns, see Kathryn Sutherland’s review of Chapman’s editions of Jane Austen’s Mansfield Park.)

Transcribers are generally authorized to incorporate certain conventions for standardizing formal elements of letters. Such standardization most commonly concerns the location of datelines, greetings, and closings. Even this degree of intervention may prove shortsighted if the letter writer employs individual formats for different types of correspondence. David Knight, editor of Sir Humphry Davy’s correspondence, pointed out that there was much to be learned from the openings and closings of these letters: “We can trace, for example, through the tops and tails of his letters, his relationship with Faraday as he progressed from amanuensis/valet to what in modern terms might be
research student,’ to ‘research assistant,’ and on to colleague” (“Background and Foreground: Getting Things in
Context,” 10). Should these variations show a significant pattern, the original format should be maintained.

When the writer used stationery with an imprinted letterhead, there is no reason to repeat this form at the
beginning of the transcription of every letter inscribed on such paper. The editor may reproduce the letterhead’s text
verbatim or simply import its image as an illustration at some point in the edition, and design an easily recognized
contracted form that can be used in the printed edition.

a. Source or Provenance Note Information

The letter’s transcriber is also responsible for recording information that may eventually find its way into the
document’s source note or {p.132} provenance note (see pp. 227–31, below) rather than into the editorial text. If
the letter’s envelope or address leaf survives, all authorial inscriptions should be transcribed verbatim. At the very
least, the transcriber should note postal markings. When they represent the basis for assigning a date to an undated
letter, they should be described in detail. Any notations by postal officials or others involved in forwarding a letter to
its intended recipient should also be noted. Once these elements have been transcribed, the editor will review them
and decide on their treatment in the edition.

The transcriber should also indicate the existence of any endorsement, that is, a notation made on the letter, its
envelope, or its address leaf after its receipt. Whenever such an endorsement indicates the date of the letter’s
arrival, summarizes the reader’s reaction, or otherwise supplies important documentary evidence, it should be
transcribed verbatim. Correspondence with governments or their agents often carry endorsements made after the
letter’s receipt by someone other than its addressee and other important documentary notations. They, too, must be
recorded in full.

These suggested practices apply only to sources that are truly letters, not essays or other short works written in
letter form with the intention of print publication. The problem of distinguishing between the two is discussed in
John A. Walker’s “Editing Zola’s Correspondence: When Is a Letter Not a Letter?”; John M. Robson’s “Practice, Not
Theory: Editing J. S. Mill’s Newspaper Writings”; and E. Grace Sherrill’s “ ‘The Daily Crucifixion of the Post’: Editing and Theorizing the Lowry Letters.”

2. Business and Financial Records

Documentary editors generally agree that documents such as business records, accounts, and others recorded in a	tabular manner should be printed in the most literally transcribed format possible. The reasons are twofold. First,
such documents make sense in visual terms only if the original arrangement of columns and indentations is
preserved. Within such a format, it is usually impossible to expand the author’s contractions or abbreviations to
make headings or entries fit into the spaces available. Second, there is no theoretical justification for emending such
records in the name of recovering the author’s literary intentions. Such sources have only evidentiary, documentary
value, and there is no reason to emend them to achieve easily appreciated literary value.

3. Professional and Technical Records

Editors who publish professional and technical records should not only master the technical terms in the
document’s verbal text but also become familiar with the special formats peculiar to such professions as law,
medicine, physics, or mathematics. They can assume that a large proportion of their readers will themselves be
specialists in the history of these fields. Such readers will be best served by a faithfully printed facsimile of the
source’s format, preserving the styled brackets of the attorney and the indentations and spacing of scientific
formulas. These elements need not be standardized in the name of readability, for they will already be familiar to
the edition’s audience, and their normalization would destroy important aspects of the source as it was
communicated to its original audience of judges, court clerks, or fellow scholars.

Yet with scientists and businessmen, just as with statesmen and authors, the line between public and private
writing will blur again and again to make the editor’s life more difficult. M. J. S. Hodge’s instructive review essay
discussing Charles Darwin’s Notebooks and Correspondence comments on this phenomenon: “It would be a
mistake to see Darwin as addressing himself in the notebooks, addressing single, known correspondents in his
letters, and only going openly to an indefinitely broader public in the books. For, even when Darwin wrote in his
notebooks or his letters, he was articulating thoughts that were formulated and refined in his mind as potential
contributions to the intrinsically public activity that natural history and science were explicitly committed to being”
(“Darwin: The Voyage, London, and Down”).

4. Government Records

Official records of any government or its agents cannot bear emendations in the name of clarity, readability, or
enjoyability. They are what they are what they are. Government financial records, of course, are subject to the rules
for tabular documents, but special respect is also due the formats and accidentals of nonfiscal records. Legislative journals, like financial records, are often intelligible only when their original patterns of spacing and indentation are retained. Greater respect is also usually shown the original formats of state papers outside the definition of journals, for these visual and inscriptive patterns were familiar to the lawyer-legislators and lawyer-clerks who inscribed them.

Transcribers should also copy the nonauthorial labels and notations on government records faithfully. Eventually they will be noted, described, or printed verbatim in the final edition. Because such documents are remarkably long-lived in terms of their documentary significance, an unprinted legislative report or treasurer's account from the 1790s may well have been consulted by congressional committees two decades later. This, in turn, may be indicated by the notations of the clerk responsible for filing and refiling the manuscript. Unless such dockets are transcribed or noted fully in the edition (either as part of the text or in a note), the source's documentary elements have been only partially reproduced.

5. An Author’s Works

An author’s works—essays, stories, plays, books, or other forms written with the intention of print publication—used to receive different treatments from the textual editor and the documentary editor. Textual editors who hoped to recover an author’s intended meaning sometimes emended one source text or conflated authorial portions of two or more sources into one new text. Textual scholars now look with more favor on the “archival” approach, that of documentary editors, who focus on a single source for their new edition, with the new noncritical editorial text reflecting the characteristics of that one source. If the source is a draft version, it should be transcribed according to the edition’s general rules for sources with genetic elements. If it is a fair copy or a printed version, the transcription should be a verbatim rendering of that single source, with later emendations indicated clearly by the editor and with variants recorded later in adjacent notes rather than incorporated into the new editorial text.

6. Journals and Diaries

The nature of a source’s original intended audience may influence transcription policies and textual method for literary works, public papers, and even letters, but this is not the case for an author’s journals or diaries. Here the intended audience is usually the author (although one wag remarked that in the case of the Adams family, the diarists’ audience was posterity). These intimate records, revealing so much of the inner life of a public figure, demand the most literal textual treatment their method of inscription permits.

Informal in nature and private in intent, diaries lose rather than gain by any attempt to impose excessive conventions of print publication. If a writer’s punctuation or spelling is less regular and correct in diaries than in correspondence, so be it. The very fact that authors allow themselves such lapses may be significant. If the author employs any form of shorthand, though, more liberal editorial intervention is not only permissible but also necessary.

The format of a diary or journal may also require special treatment. Diary entries inscribed in books whose pages carry no preprinted dates will demand additional editorial intervention in standardizing or expanding the dates of entry furnished by the author. The editor’s annotational format may play a part in this decision. When numbered informational notes follow each daily entry, an arbitrarily formalized heading, free of brackets or other typographical barbed wire, may be less distracting; with footnotes, the daily headings are less striking visually. Some standardization is often necessary even during transcription, and the most common method consists of placing each date flush to the left-hand margin. The editor will decide later whether to set the date in boldface or italic type or in full and small capital letters to give the reader easy reference. With such a run-on text of diary entries, it may be necessary to normalize the substantive elements of the date as well as its form. Expanding abbreviated forms in the dateline or even standardizing the author’s arrangement of day, month, and year may be necessary to give the reader consistently useful reference points.

Bound diaries and journals often require a more complete description of the original source than letters and other shorter manuscript sources. If the author has paginated the journal, or if the journal carries preprinted numbers on its pages, these should be noted by the transcriber for reproduction, in brackets, in the new print edition. If the diary’s pages are unnumbered, the editor may need to assign numbers for easy reference (see the explanation of such a plan in the introduction to the Irving Journals, vol. 1).

Editions in the CSE mode tend to carry more detailed descriptions of the physical appearance of bound journals and diaries, but even historical editors recognize the need to offer more descriptive information for this category of source text. As examples, see introductory notes for the John Adams, John Quincy Adams, and Charles Francis Adams Diaries. The reason for such explicit description is practical. The reader must be able to locate the original entry for comparison with the editorial text, and this is clearly a greater challenge within a bound volume of diary entries than in a separately cataloged one- or two-page letter. Thus, the transcriptions themselves should indicate breaks between volumes, as well as note any special problems in arrangement of entries in the original.
Editors transcribe records of the spoken word with special care, for any printed or even handwritten versions will be far removed from their originals. Multimedia electronic editions offering access to both the audio record of spoken words and their machine-readable transcriptions can address this problem fairly simply. (They find, of course, that users who can compare recorded words with their editorial transcriptions are even more impatient with inaccuracy.) Most editions do not have this luxury.

Transcriptions of oral communications present a visual record of words and thoughts intended to be spoken aloud. Only with the greatest skill can such a text aspire to be even second best in documentary terms. The pauses, tonal inflections, and accompanying physical gestures that once gave these words their authors’ full intended meaning are lost, and editors are often faced with conflicting inscribed versions of the same spoken words. Here, documentary editors resort, again, to the critical methods of textual editing to reconcile transcriptions of variant versions (see pp. 187–203, below).

a. Inscribed Records of Spoken Words

Editors of documents created before the development of phonography are fortunate when they find even an author’s draft, outline, or notes for a speech. In the absence of any other record, it is impossible to determine whether such a “pre-text” bears any relationship to the lecture or oration actually delivered.

Whenever a pre-text is the sole source, it is transcribed and edited as conservatively as its method of inscription allows. The question of audience here is a delicate one, for the pre-text itself was intended for its author in visual form, an aide-mémoire to the words and phrases that he or she intended to utter. When a pre-text was scribbled hastily or inscribed in some authorial shorthand, editors should consider expanding idiosyncratic contractions and translating shorthand symbols. The editor should be exceptionally restrained in emending or standardizing such formal elements as spacing, punctuation, and the like, for each device may have served the author as stage directions in speaking, and no useful purpose is served by destroying the clues that format provides. The modern edition of Jonathan Edwards’s History of the Works of Redemption, for instance, retains all of Edwards’s unique marks of punctuation to indicate pauses for emphasis in these sermons.

When the editor has access to records of sermons or lectures made by members of the speaker’s audience, it is usually wise to prepare literal transcriptions of any and all versions that survive—anything from a brief newspaper mention to a full stenographic text. Editorial notes can explain the nature of the source, and readers can judge for themselves how accurately the reporter has mentioned, narrated, or summarized the speech or conversation involved.

Fragmentary paraphrases can also occur in other documents, like a newspaper article on a matter irrelevant to the edition. In such cases, there is no need to transcribe the larger document. The editor customarily omits those portions that do not concern the oral text in question, always indicating omissions with ellipses or some other device reserved for that editorial intervention. (See, for instance, the treatment of “third-party” documents discussed in the introduction to vol. 1 of the Lafayette edition.)

Through the early nineteenth century, various mentions, narrations, and summaries of the same speech or conversation often survived in textually irreconcilable versions. In the absence of systematic shorthand, no two reporters left accounts of the same spoken words that could be viewed as variants of the original. They are usually so dissimilar that there is no question of conflating or combining them into one master record of the spoken words. Instead, each should be transcribed literally, for any normalization of format or punctuation could obscure the meaning of the original that the reporter tried to convey. The editor may be unable to identify one source text, choosing instead to give readers access to all conflicting reports. Should one record of the speech be a ten-page narration while the others are paragraph-long summaries of the same words, editors usually transcribe and print the longest record as the editorial text, reporting transcriptions of the shorter versions in notes or as separate documents.

b. Sound-Recorded Documents

Scholarly editors have just begun to address the textual problems of the newest documentary records, those created by the perfection of sound-recording equipment. When tapes or phonograph discs for actual speech or conversation survive, editors have at hand something close to the archetype for the document—their transcriptions will be the first imperfect witnesses in a long series. The skill or awkwardness with which documentary editors meet this challenge will take the measure of their scholarly specialty. At the moment, the editions paying closest attention to these problems are the Martin Luther King project and the Presidential Recordings series. For King, the editors have as source texts audio recordings of the sermons he delivered at various African American churches. The “text” of these recorded sermons includes not only the words King uttered but also the words of the congregation’s responses. The methods used to translate these elements for print publication are described by Peter Holloran in “Rediscovering Lost Values: Transcribing an African-American Sermon.”
The problems of transcribing oral history interviews have been addressed at length by experts in this field. In addition to books like the Handbook of Oral History and The Oral History Manual, Baylor University’s Institute for Oral History provides a free transcribing style guide on their Web site at http://www.baylor.edu/Oral_History/index.php?id=23607. The editor of oral history interviews should learn as much as possible about the peculiar inscriptive history of this form of document, for the practices of American oral history archives can camouflage pertinent facts. Customarily, both the interviewer and the subject review the typed transcriptions of such memoirs, the first correcting errors of transcription and the second making emendations for style and indicating passages that he or she wishes omitted from the final archival version of the transcript. Some oral history projects even destroy the original tapes, thus eliminating any chance for comparing the typed witness against its archetype. Most oral history projects now use word-processing equipment to emend transcriptions after review. This use of computer technology may ensure that only the final version of such interviews survives in the word processor’s storage, with earlier and fuller versions lost forever. As American documentary editors catch up with the products of phonographic technology, this form of source text will become a focus of lively debate.

C. Transcribing an Earlier Transcription

Sometimes earlier handwritten, typed, or printed versions of a lost archetype are all that survive. Whatever their format, such earlier transcriptions are to be regarded as scribal copies—or copies of copies—of those lost sources. They should be transcribed literally for the modern edition, and any later emendations should be as sparing and carefully considered as possible. Frequently, the editor will be able to second-guess the earlier transcriber, recognizing that the scribe has consistently read the author’s a for o or r for n. Such systematic mistranscription by a specific copyist can and should be emended, but only when the editor’s notes warn readers by explaining the reasons for these decisions and pinpointing the areas where emendation has occurred. Similarly, typographical errors in printed transcriptions (whether based on manuscript, typed transcriptions, or the lost originals) can be corrected sparingly if such emendation is needed to make the transcriptions intelligible. Here, too, later notes should explain editorial decisions in the case of any substantive emendations.

The transcriptions’ punctuation and other accidentals should usually stand without any attempt to make them conform to conventions established for original source texts in the new edition. Second-guessing an earlier transcriber does not extend to mind reading.

VI. Conclusion

While some rules of editing change with electronic technology, some do not. In “Some Unrevolutionary Aspects of Computer Editing,” Hoyt Duggan writes: “Simple accuracy is perhaps the hardest thing to achieve in an edition, and electronic technology scarcely affects the labor of transcribing and proofreading. Transcription on a keyboard, like writing on animal skins with a quill, still takes place character by character.”

Even the most basic rules for transcription of source texts raise the questions central to scholarly editing—the considerations that make possible the establishment of printed texts that reflect the editor’s experience and knowledge. A series of literal, verbatim transcriptions is almost always the sturdiest base for any documentary edition, but the edition’s textual standards as well as its organizational format may require a modification of this rule. The conventions by which American editors have presented documentary texts are analyzed in the next chapter. Well before they begin transcribing source texts, editors need to familiarize themselves with these modern textual conventions, lest their working transcriptions fail to serve the editor’s purposes.

(p.140) Suggested Readings

Anyone doubting the welcome impact of computerized methods on documentary transcription need only consult the “Directive” prepared a half century ago by Lyman Butterfield for transcribers of the Adams Papers. These instructions, models of clarity and detail, reveal how much responsibility had to be borne by junior staff members who generated hard-to-correct typewritten transcriptions for an edition of heavily emended texts.

Edward A. Levenston’s The Stuff of Literature: Physical Aspects of Texts and Their Relation to Literary Meaning is a timely reminder of the subtleties revealed and concealed by spelling, diacritics, and punctuation, typography, and layout.


Some of the essays in J. A. Dainard, ed., Editing Correspondence, touch on the special problems of transcribing letters. Good examples of editorial treatments of business and financial records can be found in the Hamilton Papers and the Morris
Papers. The editions of the legal papers of Hamilton, Adams, John Marshall, and Webster provide examples of skillful editing of this form of professional record, while the editions of Edison’s, Einstein’s, and Joseph Henry’s papers offer generous examples of the treatment of a variety of scientific materials spanning a century.

Although much of American historical editing is devoted to government records, editors have contributed little to the literature in this field. Students who wish to introduce themselves to this form of diplomatics should consult Christopher N. L. Brooke, “The Teaching of Diplomatics”; and Buford Rowland, “Recordkeeping Practices of the House of Representatives.”
Chapter Five

The Conventions of Textual Treatment

I. The Bases of Scholarly Editing: Standardizing, Recording, and Emending
II. Photographic and Typographic Facsimiles
III. Editorial Texts Requiring Symbols or Textual Annotation
   A. Textual Symbols
      1. Passages Deleted by the Author
      2. Unrecoverable Gaps in the Source Text
      3. Additions to the Original
      4. Underlining in the Source Text
      5. Authorial Symbols
      6. Line Breaks in the Source Text
      7. Editorial Supply
      8. Editorial Expansion of Abbreviations or Contractions
      9. Editorial Omissions
     10. Alternative Readings
     11. Editorial Interpolations
   B. Some Rules for Using Editorial Symbols
   C. Descriptive Textual Notes
IV. Diplomatic Transcriptions
V. The Middle Ground: Inclusive Texts and Expanded Transcriptions
   A. Inclusive Texts
   B. Expanded Transcription
   C. Textual Record
VI. Clear Text
   A. Textual Records
      1. Symbolic Method
      2. Descriptive Method
VII. Electronic Publication
VIII. Conclusion

Suggested Readings

We apologize for discussing the conventions of textual treatment in terms of verbal documents recorded on paper by hand or type. We don’t dismiss the needs of nonverbal or aurally recorded documents, we simply recognize the fact that most of these textual conventions arose from the problems of inscribed or printed documents and that the best examples of their use currently appear in editions of such texts.

This nostra culpa registered, we proceed to that least loved of creatures, a reference book’s glossary. It presents not only summaries of technical methods of confronting textual and nontextual problems but also tables of symbols and repetitive examples of the different results that can be obtained by applying varying methods to the same source. The problems and solutions discussed here are the most basic an editor will address. In documentary editions, the patterns of characters, words, phrases, and paragraphs offered to the reader are seldom the only ones that the edition’s source could have produced. Instead, they form but one text that the editor might have extrapolated from the handwritten, typed, spoken, or printed material that is the edition’s base.

It is difficult to make the choices necessary to establish an authoritative documentary edition without being familiar with earlier editorial traditions and gaining the knowledge necessary to invent new methods for novel materials created by new technologies. The editor of modern documents often has problems and goals different from those of analysts of classical texts or canonical literary works. The classicist’s aim is usually to recover a lost archetype, usually by carefully comparing the surviving witnesses to that archetype—copies made directly from it or even later transcriptions based on earlier scribal versions. Since it is impossible to hazard guesses about the formal accidentals (spelling, punctuation, or format) of that missing archetype, modern editors often standardize such elements of the text. Editors of literary works published in an author’s lifetime may have something more complicated than a recovery of the author’s original manuscript as their goal. Instead of trying to reconstruct a lost archetype, they may try to determine the author’s final intentions in an idealized form that combines elements of an incomplete authorial manuscript and subsequent printed editions of the work based on that manuscript. Frequently, such editors cannot point to a single source that represents all of the author’s careful proofreading or stylistic revisions. Instead, they must painstakingly collate, or compare, manuscripts and printed editions for their variants. They familiarize themselves with the work’s publishing history so that they can evaluate
responsibility for such changes and determine authorial patterns in the accidentals of punctuation and spelling as well as in the substantive elements of patterns of words. Thus, the editions prepared by the classicist or the literary critic can themselves be new works. Critical judgment and scholarly insight can give the reader the text of an archetype that no longer exists in a physical sense or of something that never existed, such as a critical edition of a novel that is more intellectually consistent and textually reliable than any published during an author’s lifetime.

Documentary editors of American historical materials have certain advantages over classical analysts and editors of literary works. In most cases, their source texts are themselves archetypes, and if they survive in transcribed form, such copies were usually made within decades, not centuries, of the original inscription. Thus, textual recovery is a comparatively rare concern. Unlike the editor of published literary works, the documentary editor seldom needs to compare dozens of variant versions hundreds of pages in length. There are seldom more than two or three copies with any claim to such consideration, and most such source texts are unarguably the final intention of the author—copies of letters dispatched to and received by their addressees, or public papers that the author signed and submitted to government bodies or other agencies.

This is not to say that the documentary editor’s task is easier than other scholarly editors’, merely different. He or she may, for instance, have at hand a sufficiently wide selection of holograph materials left by the subject to justify conclusions about that writer’s customary use in such accidentals as paragraph indentation, punctuation, and spelling; but the documentary editor does not have license to exploit this knowledge by emending (or correcting) the literal transcription of the source text to standardize its accidentals.

The printed versions of materials that have been edited as documents rather than reconstituted as idealized texts should themselves be usable documents—as evidence for factual research. The aim of such editorial texts is to present what was actually written or spoken, not what might have been inscribed had the author had the luxury of revising the materials for publication. It is the responsibility of documentary editors to translate handwritten, typescript, or printed source texts into a form that their readers can trust as an accurate representation of the specific original materials they represent. Even when they use traditional techniques of textual scholarship such as emendation and conflation, documentary editors stop short of making their texts too smooth, too finished in appearance. Their readers usually need to know when words or phrases in the reading text are the result of editorial judgment, not of clear evidence in the source text.

If every author of documentary material obliged posterity by inscribing her or his letters, journals, and other papers in a regular, immaculate hand (or, better still, by leaving behind impeccably proofread typescripts or word-processing files), documentary editors could discharge their duties by serving as little more than faithful scribes. Historical figures are seldom so considerate. Their records are filled with inconsistent and confusing usages, with symbols for which no equivalent exists in any printer’s font. Water, fire, insects, the ravages of time, and the scissors of autograph collectors may have defaced pages that were scarcely legible in the beginning. The array of physical details of the source texts range from authorial idiosyncrasies that are clearly pertinent to marks such as a cataloger’s notations, whose value to the edition’s readers may not be immediately apparent.

Thus, after collecting and cataloging materials, documentary editors must survey these source texts and their initial transcriptions to inventory their peculiarities. Next, they must devise ways to present an edited version of this collection that will serve the majority of users almost as well as would the archive itself. Finally, they must remember that each stage in establishing an edited text of these documents may take both editors and readers a step further from the source’s full meaning.

If the source collection consists of printed source texts, the editor’s task is comparatively easy. Textual decisions are limited to choosing a typeface that will accommodate any archaic characters in the old printing and determining the degree to which original formatting should be retained in terms of page and line breaks, line spacing, and paragraph indentation. The editor may need to do nothing more than devise a method of indicating corrections of obvious typographical errors.

But the editor of unprinted sources must make one agonizing decision after another while considering how to standardize details of inscription whose nuances might serve the purpose of some researcher. The very act of printing such source texts suppresses some of their details, for the informational content of an unprinted document can extend far beyond textual elements. The character of the handwriting or typewriting can offer clues to the author’s alertness or health. Careless penmanship in one recipient’s copy of a letter and painstaking inscription of a letter addressed to another can indicate different degrees of formality between the author and the two correspondents. The nature of the paper, ink, or pencil or typewriter ribbon can provide important clues to the time and place of the preparation of an undated letter or journal entry. Many of these important factors in the source’s documentary contents cannot be reproduced; they can only be described.

Even elements of the document added by persons other than the author must be considered part of the source text’s documentary contents. The recipient’s endorsement of a letter can be an exposition of his or her reaction to that communication, making the endorsement, in its turn, a separate document that is physically a part of the first. Notations by third parties can also be significant. Postal markings indicating a letter’s date of receipt; words, numbers, and codes entered by a clerk in the docket of a public document; similar notations made by a compositor on a manuscript that was the printer’s copy for a published essay or poem—these can be important parts of a source’s documentary evidence.

In short, the special problem of the modern documentary editor is more often an embarrassment of textual riches than the absence of an archetype or of some single manuscript that represents final authorial intentions. The
question that faces the documentary editor is how to share as much of this wealth as possible without making the printed pages of the new edition an incomprehensible mass.

I. The Bases of Scholarly Editing: Standardizing, Recording, and Emending

Documentary editors generally agree that standardization should be limited to certain formal elements of the texts, that the recording or description of physical details of the source should be limited to those that cannot be readily reproduced in print or on a computer screen, and that emendation of the source’s transcriptions should meet a few clearly defined standards. The boundaries between these categories of editorial intervention are not well defined, and one editor’s standardization may be another’s emendation.

In general, standardization concerns elements in the source’s physical format. An edition of correspondence may arbitrarily place all datelines for letters at the beginning of their texts, no matter where the date appears in the source text. Similarly, it is customary to standardize an author’s paragraph indentation in handwritten source texts so that all paragraphs in the print edition follow a consistent visual pattern.

Many details of inscription in a handwritten source text are not readily translated into printed symbols and so defy standardization. A writer may use different forms of capital letters or vary the length of dashes beyond the capacities of any typesetting system. In revising a draft, the author may employ a variety of methods to cancel rejected versions, sometimes using a line to cross out an earlier thought and sometimes simply writing over the preliminary version. The editor faced with such patterns must not only decide which patterns are worth recording in the editorial text or its accompanying notes but also find a system of symbols or abbreviations that will most conveniently communicate the detail in the source.

Beyond standardizing formats and recording significant patterns of inscription, editors must decide how lightly or heavily to emend, or correct, the transcriptions of the source texts and must choose the method of emendation most appropriate to the sources and the edition’s likely users. The term silent emendation describes changes made in a text that are not enumerated individually on the page of edited text. Silent emendation is the method traditionally used in critical editions of literary works and transferred by CEAA/CSE editors to documentary series: the editorial reading text itself contains no hint that emendations or alterations have occurred. Most but not all such emendations are reported in a back-of-book textual record. The regularization of capitalization and punctuation and corrections of misspellings are often unrecorded. Overt emendation, on the contrary, refers to changes indicated within the editorial text (usually enclosed in some form of bracket) or in notes immediately adjacent to the text.

In choosing methods for standardizing documentary formats, recording details of inscription, and emending an author’s spelling or punctuation, the editor should bear in mind the discouraging fact that the printed version of an unprinted source text—or even of a printed source with unique characteristics—will always be an inexact copy. One can hope to do nothing more than choose those conventions of print publication that best communicate the significant patterns of the source text at hand.

What is more, the editor must be aware that any documentary edition that encompasses more than one source text may have to use more than one textual method. Even different passages within a single lengthy source text may require varying approaches. The editor’s wisest course is to choose a textual method that will serve the bulk of the material being edited. When exceptional circumstances require a departure from the general rule, a note to the reader can explain the reasons for this variation, as well as the new method’s implications for the source text at hand.

Modern editors of American documents usually perform their tasks within five general methodological frameworks, and an infinite variety of results is possible within each general approach. This chapter examines the theoretical principles and practical results of these methods: facsimiles, diplomatic transcriptions, inclusive texts, expanded texts, and clear texts. Of these, facsimiles produce editorial texts closest to the source in physical appearance. At the other extreme, eclectically constructed clear texts of documents can suppress many elements of inscription, making recovery of the details of the original extraordinarily difficult. Diplomatic transcriptions, inclusive texts, and expanded texts fall between these two poles, and they are described in the order in which they move away from the literal facsimile’s presentation toward the systematic emendations sanctioned in a clear text.

A theoretical and historical discussion of each method will be followed by a demonstration of the results of applying that approach to the same source text, a handwritten draft of a note from one of the authors of this book, Mary-Jo Kline, to Richard K. Showman, chairman of the committee that supervised her work in the early 1980s. The advantages and failings of each method are apparent.

II. Photographic and Typographic Facsimiles

Facsimiles of documentary texts in an authoritative edition may be photoreproductions or digital images or typographic facsimiles on printed pages or computer screens. Traditionally, examples of photo facsimile editions have been limited to companion or supplementary microform series, particularly for NHPRC series in which select print editions were paired with comprehensive microforms. For editors of historical documents, the use of photo facsimiles in published volumes remains generally limited to reproductions of printed sources. A modern exception was Ralph Franklin’s two-volume edition of The Manuscript Books of Emily Dickinson.
(Cambridge, Mass., 1981), a facsimile publication of the “fascicles” or handmade booklets in which Dickinson drafted and copied out her poems. The same press has on its Web site a facsimile edition of Dickinson’s Herbarium, edited by Richard Sewall, a fascinating online example of the technique: http://www.hup.harvard.edu/catalog/DICEMI.html.

More typical is the Albert Einstein edition’s Works series, which reproduces facsimiles of the print pages bearing Einstein’s published essays and reports rather than resetting them in a modern typeface, and the modern edition of Thomas Jefferson’s Extracts from the Gospels, which provides a facsimile reproduction of the clipped Bible verses in Jefferson’s compilations.

Both theory and technology now make such editions more common and more intellectually demanding. Among literary scholars, the call for a document-based approach to sources customarily served by highly emended, printed editorial texts has created a demand for convenient access to images of source materials. Increasingly common are print editions in which a photo facsimile appears as part of a parallel text accompanying a printed editorial transcription. Digital scanning creates wider options for editors who wish to offer such photographic images in online or DVD-based editions, conveniently linked to machine-searchable transcriptions, accessed through automated indexes. The technical processes of creating such facsimile editions are described in chapter 7.

The typographic facsimile demands more discussion here, for its rationale and methods play a significant role in textual and documentary editions. This form attempts to duplicate exactly the appearance of the original source text as far as possible within the limits of modern typesetting technology. Naturally, a typographic facsimile reproduces the author’s spelling and punctuation without any correction. No contractions or abbreviations are expanded. An author’s additions above the line are printed interlinearly. Marginal material is set in the margins. Passages crossed out by a line are rendered in canceled type (canceled type). The author’s format and spacing are followed exactly. Headings, {p.148} titles, datelines, greetings, and salutations are set line for line so that the line breaks in the printed version mirror those in the original.

Modern typographic facsimiles of handwritten material customarily do not present pages whose breaks correspond to changes in the original’s pagination, but those page breaks are recorded within the editorial text, usually by an editorial interpolation within some form of bracket. Such editorial interpolations as “[2]” or “[new page]” can appear in the body of the text (when the author begins a new page in the middle of a sentence) or in the margin (when the new page represents a new section of the document).

Examples of a pure typographic facsimile, without the use of any textual symbols or other editorial conventions, are rare in modern editions. An approximation of facsimile technique can be found in Julian Boyd’s “literal presentation” of Jefferson’s drafts of the Declaration of Independence (Jefferson Papers, 1:417–27), in the texts of letters in Shelley and His Circle, and in the Cornell Wordsworth edition.

Because typographic facsimiles can be more time-consuming and costly to produce than other formats, they are generally used for nonprinted source texts only when textual symbols will not adequately communicate the nuances and complexities of the original. The users of a series of colonial laws, for instance, would be more concerned with the final versions of statutes—the words or phrases that were promulgated by the provincial government—than with a clerk’s corrections of his copy of that law. Here a typographic facsimile would be a needless luxury. In contrast, the editor of documents relating to the legislative history of a colonial assembly would have good reason to consider a typographic facsimile of a draft of an important legislative report. In such a document, the position of interlinearations and marginal additions can indicate the evolution of the final version of laws, often essential to understanding their history.

Some elements of typographic facsimile techniques may be used for individual documents in editions that otherwise do not follow this method. One legal historian has remarked, for instance, that the reproduction of the format of Hamilton’s draft legal papers in the Hamilton Law Practice volumes “allows the reader to see Hamilton’s mind at work.” This edition uses conservative expansion of some archaic abbreviations and contractions and thus cannot be described as a true typographic facsimile, but its editors recognized that there were special {p.149} documents whose full meaning could be conveyed only by retaining Hamilton’s format.

Typographic facsimile has always been the method of choice for reproducing printed documentary source texts. Even editors who took substantial liberties with manuscript source texts did their best to reproduce as exactly as possible the appearance of printed sources such as pamphlets and newspapers (see, e.g., the treatment of “A Representation of Facts” in vol. 5 of the Laurens Papers). This rule for the treatment of printed source texts is consistent among editions of eighteenth-century materials, but editors of more modern documents have departed from it. In the Woodrow Wilson Papers, for example, the same rules for emending spelling and punctuation were applied to both manuscript source texts and Wilson’s published works. The editors of the Harold Frederic Correspondence allowed themselves far greater latitude in emending printed and even typewritten source texts of letters than in treating handwritten sources.

The economies offered by modern technology make this textual approach economically practical for many more editions. The first volume of the Documentary History of the Supreme Court, published in 1985, offered readers page after page of documents in which superscript letters, marginal additions, interlineations, and archaic symbols were reproduced literally in type. By then, the project’s editors could put encoded editorial copy for both documents and notes into computer files, which were translated into print by Columbia University Press. The press estimated that the cost of producing each page of the Supreme Court volumes in printed facsimiles in 1985 was a bit less in real terms than the per-page cost for the more standardized texts of a volume of the Hamilton Papers in 1978.
Although theory and technology now combine to make facsimile editions more popular, the most important theoretical consideration in publishing facsimiles remains the same: editors and readers alike must remember that the facsimiles themselves are edited forms or versions of their sources, with cameras or scanning equipment performing the editorial function. They are their own form of editorial text, and they demand the same exacting review readers would expect from printed emended texts. In his excellent historical survey “Reproductions and Scholarship,” G. Thomas Tanselle reminds us that facsimile editions, unlike conventional photoreproductions, involve “thorough proofreading and the writing of notes on potentially misleading spots” (75).

(p.152) Figures 1 and 2 display the results of imposing facsimile treatment on this chapter’s sample document. Figure 1 shows a photographic facsimile, which can be compared against the typographic facsimile in figure 2. In a scholarly edition, readers would be informed that the letter was written in April 1982, when the author was engaged in arranging the date for a meeting in Bloomington, Indiana. There she was to discuss the final details of the Guide with an executive subcommittee composed of David Chesnutt of the University of South Carolina, David Nordloh of Indiana University, and Paul Smith of the Library of Congress. It was expected that another person might attend these meetings—Don Cook, then president of the ADE and a member of the Indiana University Department of English.

Figure 1. Photofacsimile
Dear NICK,

I’ve begun annoying members of the Executive Sub-Committee to make up their minds about the new date of our planning meeting in Bloomington. Chairman Dave N. has his shuttle school teaching schedule to reconsider. Dave C. may be going to Italy. Paul S. doubtless has vacation plans.

Another week in the life of the author of the ADE guide: one Clizard (Tuesday) and one lunch with a member of the Committee on Manuscript (Wednesday).

I’ll start once again with editing documents: is it more fun than writing about editing them? There are only so many ways to say: “The responsible editor should...”

Often more frequently I must ask myself, from typing: “Seek fellow editors; this is the way to do it.”

Realize that we are the tongue we should string for.

I’ll let you know when the date for the Sub-Committee meeting is set — hope that D.C. can be one of our party, but that must be left to chance. I can hardly believe him for finding an excuse to miss two days’ worth of arguments about the history of the apple tree.

Most obviously, this could be disagreement among members of the Sub-Committee. Let me argue that the entire book should be written in FORSAND and foliage may insist that the book is written too simplistically — that it isn’t intended for these Diogenes to confuse, or even to play peacemaker.

I’ll keep you advised of all developments. Any advice in advance will be welcomed.

And, perhaps, needed.

Yours in a quandary —

Mary J.

9 April
Dear Dick

to nag

I've begun nagging members of ye Executive Sub-Comm. to make up their minds about the for date of our planning meeting in Bloomington. Chairman Dave N. has his summer school teaching schedule to consider. Dave C. may be going to Italy. Paul S. doubtless has vacation plans.

Another week in the life of the author of the ADE "Guide (\)": one blizzard (Tuesday) and one lunch with a member of the "Comm\'ee on ye Manual" (Wednesday).

I'll state once again that editing documents is a lot more fun than writing about editing them. There are only so many ways to say "the responsible editor should . . . . " More and more frequently I must restrain myself from typing, "Look fellow editors, this is the way to do it!"

I realize that isn't the tone we should strive for. I'll let you know when the date for our Bl'ton meeting is set— I hope that D.C. can be one of our party, but that must be left to chance. I can hardly blame him for finding an excuse to miss two days' worth of arguments about the history of the angle bracket as a symbol for authorial cancellations.

More seriously, there could be disagreement among members of the Sub-Comm. Chesnutt will argue that the entire book should be written in FORSAN. Nordlieh may insist that the book is written too simplistically—that it isn't intended for those so ignorant to confuse a stemma with a lemma. I wish that you could be there to play peacemaker.

I'll keep you advised of all developments. Any advice give in advance you will be welcomed and, perhaps, heeded.

Yours in a quandary—
Mary-Jo

9 April

III. Editorial Texts Requiring Symbols or Textual Annotation

Only editors who prepare image facsimiles are spared the use of editorial devices that communicate original inscriptive details that cannot readily be duplicated in a typeset version. As they have put general theories of textual method into practice, American editors have used a broad, sometimes bewildering variety of mechanical
devices and techniques. In a diplomatic transcription, symbols or abbreviations appear in the documentary text itself. In inclusive or expanded texts, the symbols may appear both in the text and in supplementary notes. In clear text, symbols and notes appear in the back-of-book records of editorial emendations and details of inscription.

A. Textual Symbols

The use of textual symbols goes back to the work of classical scholars in their editions of ancient works. The first systematic use of textual symbols for more recent materials appeared in the early 1920s in the Malone Society's Reprints series, edited by W. W. Greg. The society's editions of British literary works of the Renaissance and the early modern period employed two sets of characters to indicate details in manuscript sources: angle brackets (« ») to enclose passages lost through mutilation or other damage and restored by the editors, and square brackets ([ ] ) to enclose passages deleted in the original manuscript.

After World War II, these and other symbols were applied to American documents and literary works. The first American system of textual symbols was devised by Julian Boyd and Lyman Butterfield for their Jefferson and Rush editions. They confined themselves to the same two pairs of symbols used by the Malone Society, but they modified the meaning of each set of brackets. Because common American usage had already assigned to square brackets the function of setting off interpolated material, these symbols were given a similar meaning in the text of American documentary editions. Instead of denoting authorial deletions, square brackets in the Jefferson and Rush volumes indicated some form of editorial intrusion into the text—the insertion of characters or words not physically a part of the original, whether these were added to restore mutilated passages or interjected to explain some aspect of the text. Angle brackets, on the other hand, had no generally accepted function in American usage of the time. Since they were a “neutral” symbol, the editors of the Jefferson and Rush editions arbitrarily assigned them to enclose restored canceled material in their source texts.

During the thirty years following publication of the first volume of the Jefferson Papers, scholarly editors here and abroad devised symbols and abbreviations for almost every detail of inscription of which the human mind, hand, and pen could be guilty. Symbolic description of textual detail won popularity so quickly that many of its adopters were unaware of the novelty of their methods. In the statement of textual method for the first volume of the Emerson Journals, published in 1961, the use of angle brackets to enclose restored authorial deletions was described as traditional, although that tradition was only a decade old. The symbol has now become so widely used that some editors no longer bother to define its meaning in prefaces to their volumes of documentary texts.

Other conventions have won less universal acceptance and acclaim. The list that follows records only some of the devices that modern editors have devised to represent textual details—what Lewis Mumford dubbed the “barbed wire” of modern American scholarly texts:

1. Passages Deleted by the Author

« italic » Original device used in Jefferson Papers. Textual notes now record deletions
<br>
roman Emerson, Howells, Irving, and Frederic editions; most historical editions
<br>
<< italic or roman >> A deletion within a deletion in an edition that employs single angle (or “broken”) brackets for a primary deletion
<br>
{p.154} canceled type Grant and Hamilton Papers; most editions of literary works; almost every edition inaugurated since 1990
<br>
< A crossed-out deletion, Billy Budd
<br>
< An erasure, Billy Budd

2. Unrecoverable Gaps in the Source Text

[ . . . ] The number of suspension points within the square brackets usually offers a clue to the length of the lacuna, or unrecoverable material. In the Jefferson and Adams Papers, “[. . . ], [. . . ]” indicates one or two missing words; if a footnote number follows the brackets, the lacuna is longer, and a note estimates the number of missing words. The Cooper edition employs a similar technique. In the Grant Papers, the number of points represents the approximate number of missing letters, not words.
<br>
[ ] A missing portion of a number, Jefferson Papers
<br>
|| . . . || The number of suspension points approximates the number of missing words in the Emerson Journals, with three dots representing one to five words; four dots, six to ten words; and five dots, sixteen to thirty words. The abbreviation msm within the vertical lines stands for “manuscript accidentally mutilated.”
<br>
{} Missing words or pages in the Margaret Sanger edition, with text between the styled brackets indicating the amount of lost text (“{two words missing}” or “{one page missing}”)
<br>
xxx Missing letters in the Emerson Journals, with the number of x’s approximating the number of
lost characters

[---] Missing words in the Grant and Wilson Papers, with the spaced hyphens indicating the number of words lost. The editors of the Ratification series employ the same symbol, but three hyphens are used regardless of the length of the lacuna.

... Lacunae in the Hamilton Papers, with the suspension \{p.155\} points representing the approximate number of unrecoverable characters

//...// Illegible words in Mark Twain's Satires

[ *** ] Unrecoverable shorthand characters, Wilson Papers

Countless variations upon this theme are possible. The Irving edition, which employs angle brackets to enclose deleted passages, combines that symbol with italicized descriptive words or phrases to indicate unrecoverable canceled passages, as with "<illegible>" for a hopelessly obliterated section.

3. Additions to the Original

> An insertion with Melville’s caret, Billy Budd

< An interlinear insertion without a caret, Billy Budd

^roman^ All insertions (interlinear and marginal), Mark Twain's Satires; interliniations only, Frederic Correspondence; interliniations made with author’s symbol for an interlined addition, Emerson Journals

↓roman↑ “Substitutions for a deletion,” Mark Twain’s Satires; interliniations in Emerson Journals, Irving edition, and most other “literary” series and the Margaret Sanger Papers

roman/ Marcus Garvey Papers

I roman I Marginal additions, Emerson Journals

< > Jane Addams Papers; interlineation appearing within the brackets

w.o. Superimposed addition (“written over”), Billy Budd

4. Underlining in the Source Text

italic Single underlining. This method was once almost universal, but most projects now simply print a word with its underline.

SMALL CAPITALS Double underlining. This convention, too, was once almost universal. Modern editions print the text with two underlines.

\{p.156\} 5. Authorial Symbols

þ The handwritten thorn, which had been formalized to y by the mid-eighteenth century, is customarily printed as a y or as th for materials in American history and literature.

~ Many eighteenth- and early nineteenth-century authors employed the curved tilde or a simple straight line at the point in a word where characters were omitted to form a contraction. The tilde is reproduced in type by the Laurens, Ratification, and Burr editors. Many editions, following the lead of the Jefferson Papers, once ignored the tilde and silently expanded the resulting contraction where it was employed. Today, the Jefferson editions reproduce the tilde without expanding the contracted form.

p The “tailed p” is either rendered by the character for this symbol in print (Jefferson Papers) or expanded to the intended form of per, pre, or pro (Hamilton Papers, Letters of Delegates). If the meaning of the symbol is unclear, an edition that ordinarily expands it must indicate an ambiguous usage by “p[er?]” or some other method reserved by the edition for conjectural readings. Changes in modern-day typesetting are rapidly eliminating the need to translate this symbol, but it is still one of the characters unavailable for display in HTML.

* The asterisk is the most commonly employed rendition of an idiosyncratic symbol used by authors to indicate their own footnote numbers (Emerson Journals). However, when the author uses a standardized (even if rather archaic) form of citation, it is preferable to retain the form when it has an equivalent in modern type fonts, such as a dagger (†).

☞ A “fist” or “index” drawn by the author in the margin to call attention to a passage in his or her text can be translated to the printed “fist” that survives in many typefaces.

\{p.157\} Bracket used in the Emerson Journals to enclose page numbers supplied by Emerson himself

¶ Author’s marking for a new paragraph
6. Line Breaks in the Source Text

/ Most historical editions; Howells Letters
| Hawthorne, Whitman, and James editions

7. Editorial Supply

[roman] The most common device for both literary and historical series; if doubt exists concerning the
supplied material, a question mark precedes the closing bracket ([reading?]).

<roman> Hamilton Papers
|| roman || Emerson Journals
| roman | Howells Letters
{roman} Frederic Correspondence
-{roman}- Wilson Papers device for “word or words in the original text which Wilson omitted in
copying”

8. Editorial Expansion of Abbreviations or Contractions

[roman] Universally accepted symbol in editions that expand such forms within the text

9. Editorial Omissions

[ . . . ] Emerson Journals
. . . Booker T. Washington Papers

10. Alternative Readings

/ Introduces alternative readings in Mark Twain’s Satires
/roman/ The virgules enclose alternative readings in the Emerson Journals
\{p.158\} When variant copies provide alternative readings in documentary materials, it’s far more common to
describe the variations in notes than to represent them symbolically within the text.

11. Editorial Interpolations

[italic] This is the most commonly used device, although both the Emerson and Ratification volumes
employ “[roman].” If an edition contains a substantial amount of material that must be
represented within square brackets for other reasons (lacunae in the text, expanded
abbreviations, supplies of mutilated passages), the editorial apparatus must distinguish
clearly between bracketed material that can at least be inferred from the source text and
bracketed contributions that do not stem from the source (corrections of outrageously
misspelled words, catchwords such as illegible, and so on). The reader must not be left to
wonder to which category the bracketed material belongs. If brackets are used sparingly in the

text, and there is no possibility of confusion, then all bracketed letters, words, and phrases
can be in the same typeface.

italic Frederic Correspondence

B. Some Rules for Using Editorial Symbols

The variety of symbols used by different editions to represent the same textual detail is so great that it’s hard to
escape the conclusion that some symbols were adopted because editors were unaware of the conventions already in
place. Others may have tried to prove their inventiveness by adopting a new form of bracket or a new arrangement
of virgules (slashes) instead of imitating another edition’s practices. Nothing short of an editorial Council of Trent
could impose order on the symbolic chaos in existing documentary editions. We can, however, offer a few
observations and suggestions.

A careless choice of symbols can make the reader’s task more difficult. If two symbols represent related details in
the manuscript, the symbols themselves should have a visual relationship. The reader’s memory will be burdened
sufficiently without the addition of sets of characters that contradict each other in appearance and meaning.
Following this rule, \{p.159\} the editors of Melville’s Billy Budd used variations on the opening half of a pair of
angle brackets (<) to indicate different kinds of authorial cancellation, and employed variants on the closing angle
bracket (>) to represent two methods of interlineation in the source text. If the edition requires a lengthy series of
textual symbols, the editor should use as many devices as possible that mirror the physical appearance of the
original. If an author cancels material both by lining out phrases and by erasing them, the lined-out deletions
should appear in canceled type, whose meaning is easily grasped, with symbolic designation reserved for the
erasures.

Fortunately, computerized publishing technology makes it possible to eliminate a good many symbols
altogether. Many of those details can now be reproduced with no more trouble to the editor or expense to the
publisher than inserting codes for symbols. Arbitrary symbols can be reserved for problems that cannot be so easily
translated, and the list of symbols the reader is required to memorize will be reduced correspondingly. Online
editors can provide hyperlinks from transcribed editorial text to points in images of the source texts where
questions or confusion might arise.

C. Descriptive Textual Notes

Despite the thicket of editorial symbols cultivated by modern scholars, many details of inscription stubbornly defy
symbolic representation, and some editors simply prefer to avoid the use of textual symbols altogether. Instead of
symbols, their textual notes use verbal descriptions of textual problems whenever possible. Such textual
descriptions can be provided in three ways: (1) in the documentary text within square brackets (as in the Calhoun
Papers); (2) in footnotes whose numbers are keyed to the location of the cancellation, interlineation, marginal
addition, or other detail in the edited text of the document (Franklin Papers); and (3) in a back-of-book textual
record. The reasons for choosing each variation on the method are instructive.

The format of the Calhoun Papers allows for no footnotes of any kind. All editorially supplied historical notes as
well as textual explanations are presented within square brackets in the text. This design does not lend itself easily
to arrows and other symbols within the initial brackets. In the Franklin Papers, William Willcox explained that he
found the use of symbols within the texts of printed documents to be “disfiguring,” and his choice of descriptive
methods rested on this aesthetic preference.

The format of the Calhoun Papers allows for no footnotes of any kind. All editorially supplied historical notes as
well as textual explanations are presented within square brackets in the text. This design does not lend itself easily
to arrows and other symbols within the initial brackets. In the Franklin Papers, William Willcox explained that he
found the use of symbols within the texts of printed documents to be “disfiguring,” and his choice of descriptive
methods rested on this aesthetic preference.

The format of the Calhoun Papers allows for no footnotes of any kind. All editorially supplied historical notes as
well as textual explanations are presented within square brackets in the text. This design does not lend itself easily
to arrows and other symbols within the initial brackets. In the Franklin Papers, William Willcox explained that he
found the use of symbols within the texts of printed documents to be “disfiguring,” and his choice of descriptive
methods rested on this aesthetic preference.

(p.160) Traditionally, editions of the writings of American literary figures have carried a complete textual
record. Here the use of descriptive rather than symbolic methods rests on very different and more complicated
grounds. Some of the details and patterns customarily reported do not lend themselves to symbolic treatment. The
resulting number of textual notes would make it impossible to report these details within the text of a document or
even in numbered footnotes, whose profusion would make the text a field of numerals rather than words and
phrases. Therefore, the textual record usually has to be presented in unnumbered notes keyed to the line and page
of the printed edition, appearing either in a section of notes following the text or in a back-of-book section. In either
case, the textual note is at some distance physically from the section of edited text to which it referred. Many textual
problems requiring such notes are so complicated that readers need special cues for understanding the significance
of emendations or omissions, and brief verbal descriptions frequently serve this purpose better than symbols.

Fredson Bowers was the leading exponent of descriptive rather than symbolic textual annotation, and his edition
of Leaves of Grass is the best example of a diplomatic transcription employing the technique. The notes follow each
poem and can be consulted easily. Bowers later argued for the use of descriptive textual annotation for inclusive and
clear texts as well, and the practical effect of the method in such an edition was first seen in the Hawthorne
Notebooks. The system became more highly refined in Bowers’s textual notes for the William James edition, and his
exposition of the method appears in his “Transcription of Manuscripts: The Record of Variants.” The back-of-book
records of emendations and inscriptive details in these series follow the traditional format for literary works. Each
line-page reference is followed by the lemma (a word or phrase in the editorial text that indicates the site of editorial
activity), followed by a left-opening bracket ([), which divides the lemma from the reading in the source text. The
method can be seen at its simplest in the record of the first editorial emendation in Hawthorne’s French and Italian
Notebooks:

6.3 a little more or less] A little more less

This merely indicates that in line 3 on page 6, the editors have supplied the word or, which Hawthorne omitted from
this entry in his notebook.

(p.161) Usually, such textual tables need only three symbols in addition to the left-opening bracket that follows
the lemma: the slash (/), which indicates a line break in the source text; the mathematical symbol ~, which
represents a repetition of the word that appears after the bracket; and a caret on the line (^), which can indicate the
absence of punctuation in the source text. Thus, in the same Hawthorne volume, “of] of/ of” shows that the editors
have omitted the of that Hawthorne repeated when he began a new line. And “Liverpool,] l~ˆ” shows that the
editors have added the comma after Liverpool.

A scheme of descriptive textual annotation should be designed as carefully as one of symbolic representation.
The words and phrases that describe more complicated emendations and details in the source can be written out or
abbreviated. The Thoreau Journals contain descriptive textual notes without abbreviations for the best and most
practical of reasons: brevity. The notes seldom run to more than one line with unabridged descriptive forms, and
using contracted forms would have saved no space while forcing readers to master a table of abbreviations. Still,
many editions require abbreviations to prevent the descriptive textual notes from becoming unmanageable; among
the most common are del. for deleted, ab. for above the line, and interp. for interlined. A descriptive textual record is
offered for the clear text version of the sample document in figure 5, below. Descriptive annotation is still very much alive, and a new, alternative system appears in David L. Vander Meulen and G. Thomas Tanselle’s “A System of Manuscript Transcription.”

**IV. Diplomatic Transcriptions**

In modern American editions, a diplomatic transcription is one step removed from the typographic facsimile. The editor uses carefully chosen critical symbols or abbreviations to indicate details of inscription such as interlineations and cancellations instead of reproducing their physical appearance in the original. Editors of diplomatic transcriptions often standardize the placement of such routine elements of the source text as datelines, greetings, salutations, titles, and the indentation of paragraphs, and they may also supply missing punctuation, expand ambiguous or archaic abbreviations and contractions, or even supply words unintentionally omitted by the author or destroyed by mutilation of the original source text. None of these corrections or emendations, however, is made silently: each is given within a form of brackets indicating such editorial activities. If some emendation or detail of original inscription cannot be described conveniently with symbols or a bracketed interpolation, a footnote immediately adjacent to the text explains the problem at hand.

*Figure 3. Diplomatic transcription*
Examples of exhaustive diplomatic transcriptions are almost as rare as printed facsimiles among editions of modern American materials. Perhaps the best known are Fredson Bowers's *Leaves of Grass*, the Hayford-Sealts genetic text of Melville's *Billy Budd*, and Donald Reiman's edition of *Shelley and His Circle*. In all these, the reader has immediate access to the details of the original manuscript source text in the diplomatic transcription and to a critically edited reading text, which represents the author's apparent final intentions. In *Leaves of Grass* and *Shelley and His Circle* the reading texts and the diplomatic transcriptions are presented as parallel texts on facing pages. The Hayford-Sealts reading text of *Billy Budd* precedes the genetic text of the author's manuscript. For materials whose diplomatic transcriptions rival these in complexity, providing a reading text is not only a kindness to the reader but a necessity.

Figure 3 offers a diplomatic transcription using the following textual symbols:

`<>` Deleted passages
V. The Middle Ground: Inclusive Texts and Expanded Transcriptions

Most editors compromise to one degree or another between a detailed diplomatic text and a clear reading text. Among editions in the CEAA/CSE tradition, such methods are described as “inclusive.” For historian-editors the practice became known as expanded transcription. The difference between the inclusive and expanded methods is not so much the editor’s basic conservatism or liberalism in emending and standardizing the text as the degree to which such editorial tinkering is reported to the reader. In both inclusive texts and expanded transcriptions, certain classes of emendations (usually relating to physical format) are performed silently in individual cases although described to the reader as classes of correction and standardization in a general statement. Editors of inclusive texts that meet CSE guidelines report any emendations in other categories individually in accompanying textual notes. Expanded texts in historical editions offer no such supplementary record of emendations or suppressed details beyond what appears in the text and in footnotes immediately adjacent to the documentary text.

In both techniques, some details of inscription are reported overtly in the editorial text, that is, through the use of textual symbols or numbered footnotes adjacent to the text. Both breeds of editors may also standardize certain elements of the format of the source text silently. Datelines and place lines in letters are usually printed above the greetings and text, no matter where they appear in the original. Headings for diary and journal entries are standardized for easy reference. Paragraph indentations and dashes of varying length are made uniform.

Beyond such standardization of the format, inclusive and expansive editors often emend the text without giving any overt indication. In decades past, superscript characters were commonly lowered to form an abbreviation or contraction on one line, with a mark of punctuation placed after the resulting form. Archaic holographic symbols were generally rendered in their closest equivalent in print (as th or y for the thorn, or per, pre, or pro for the tailed p). Abbreviations or contractions that editors judged unfamiliar to their readers were not reproduced exactly—they were expanded silently or overtly, with brackets to indicate that an editorial hand was at work or a change in typeface to signal supplied words or characters. Erratic punctuation was standardized—silently or overtly—and missing marks of punctuation were supplied. Modern editors of long-lived projects that began with broad patterns of emendation have generally modified these policies in favor of more conservative treatment.

A. Inclusive Texts

The most cogent description of requirements for an inclusive text appeared in the CEAA’s revised Statement of Editorial Principles of 1972 (9). It stated that inclusive methods should be employed for a source text whose audience is “limited mostly to scholars and specialists,” and they were specifically recommended to editors whose source texts are “manuscript letters or journals or notebooks” for which no authoritative published version exists. The Statement also pointed out that this method is preferred whenever “reporting the author’s process of composition directly is important.” Once the editorial text is established, deletions and revisions might be reported on the same page as the editorial text, as for a diplomatic transcription; in accompanying textual footnotes at the bottom of the page; or in notes placed between such “separable” items as letters and entries in a journal. These textual notes can explain the use of symbols in the text and record details of inscription that the editor has been forced to omit from the text itself.

In addition to such notes, the CEAA/CSE inclusive editor was usually required to furnish “in some form” a record of editorial emendations of the source text that were not clearly indicated in that text or its adjacent notes. Thus, an inclusive text might be followed by a back-of-book record of emendations similar to the ones that appeared in CEAA/CSE editions of literary works, with their references keyed to the line and page of the edited volume where they occurred. An inclusive text of documentary materials, like any CEAA/CSE edition, was also to be accompanied by a report of “editorial decisions in the handling of possible compounds hyphenated at the ends of lines in the copy-text, along with an indication of which end-of-line hyphenations in the newly edited text should be retained in quotations from the text.”

In practice, few multivolume CEAA/CSE editions consistently followed these rules for documentary materials. The Washington Irving series was among this small number. Irving’s Journals and Letters used editorial symbols in their texts for most details of inscription. Numbered footnotes at the bottom of the page (Journals) or at the close of individual items (Letters) reported emendations and details suppressed in the reading text.

Editors experimented with a variety of methods for documentary source texts. The Emerson Journals provided an example of inclusive methods, with editorial symbols within the journal entries for major details of inscription and a further back-of-book record of editorial emendations not apparent in the reading text itself. The Emerson Lectures, however appeared in clear text. Still another series bearing the CSE seal, the Harold Frederic Correspondence, imposed different editorial methods on different groups of source texts in the same volume. The 203 letters whose source texts are holographs are offered as inclusive texts, while the 163 letters in the volume that
One edition adopted inclusive methods midstream in its progress. The first volumes of documentary material in the Mark Twain Papers—Twain’s Letters to His Publishers and his Letters to H. H. Rogers—were heavily emended by the editors, with neither a report of canceled passages nor a record of other details suppressed in the reading text. Ensuing volumes of documentary materials, however, adopted inclusive methods. The volumes of Twain’s Notebooks and Journals are an example of inclusive textual editing at its best. The reading texts employ conventional editorial symbols to indicate such items as legible canceled passages, revisions, and other significant details of the source texts. A back-of-book record conveniently divides textual notes between “Editorial Emendations and Doubtful Readings” and “Details of Inscription in the Manuscript.” The first alerts the reader to elements of the reading text that result from editorial judgment, while the second records elements of Clemens’s accidentals and substantives in the source texts that either did not warrant inclusion in the reading text or could not be reproduced symbolically on the printed page.

In the light of experience, the editors of the Twain edition later devised a new form of inclusive textual editing dubbed “plain text” for their author’s correspondence. This system is discussed at length in Robert Hirst’s “Guide to Editorial Practice” and “Guide to Textual Commentaries” in Mark Twain’s Letters (i:xxv–xlv, 447–63). The Twain editors do not pretend that their reading texts are print facsimiles or even diplomatic transcriptions, but these conservatively emended inclusive texts use typographic equivalents for certain details of the original rather than the editorial symbols customarily used in inclusive texts. Samuel Clemens’s use of the conventions of nineteenth-century typography in his personal letter writing and the economies of modern computerized typesetting made this effective, commonsense approach a practical possibility. The letters that appeared in the two earlier volumes, Letters to His Publishers and Letters to H. H. Rogers, are being reprinted in later volumes of the Letters series, appearing at their appropriate chronological places and reedited in plain text.

None of these inclusive or plain text editions pretends to report all the details of their sources. Two forms of emendation are made silently within the text, with no record in the textual apparatus: (1) standardization of the manuscript’s format, including the placement of such elements as a letter’s dateline, salutation, and complimentary close; uniform spacing between lines and uniform indentation of paragraphs; and (2) (p.167) standardization of irregularly formed letters of the alphabet and marks of punctuation. Inclusive editors seeking CSE approval may enlarge this list of silent emendations, listing additional categories in statements of textual method in their volumes.

Certain details of inscription are customarily reported overtly, by using symbols or facsimile printing in the text or by placing footnotes adjacent to the text. Those details include legible canceled passages, especially those that reflect a change in the substance of the author’s thought, and legible additions to the original passage, such as interlineations, on-line additions, and marginal insertions. An interlinear spelling correction, for instance, would not fall into this category.

Details of inscription in the following categories are usually omitted from the inclusive text and recorded only in back-of-book textual notes: (1) false starts so brief that they give no sense of the author’s preliminary intention; (2) slips of the pen such as dittography (words repeated unintentionally) or minor errors of spelling or punctuation that the author did not correct; (3) authorial corrections of spelling and punctuation (whether as write-overs, interlineations, added characters, or marginal insertions) that do not indicate a change in the desired sense of the passage; (4) illegible canceled passages; (5) catchwords at the bottom of a page that are repeated at the top of the following page; (6) a change in any of the media used in the original manuscript (i.e., variations in paper, ink, or pencil within the same document); (7) symbols that cannot be reproduced in set type in any readable form and must be described rather than represented by visual symbols; and (8) authorial revisions so complicated that not even diplomatic transcription or facsimile printing could represent them clearly.

Similarly, some types of editorial emendation are usually reported within the text rather than in the back-of-book record of an inclusive edition. The following editorial contributions are signaled by symbols in the text (usually a pair of square brackets) alerting the reader to possible ambiguity: (1) supply of a word, phrase, or a single mark of punctuation omitted by the author; (2) supply of mutilated or obliterated material; (3) any change in the identity of the source text, when two or more sources are conflated to produce the complete text of an item; (4) any editorial expansion of authorial shorthand necessary to make a passage read sensibly, such as completion of dates and expansion of ambiguous contractions and abbreviations or of a set of initials to a full name; (5) any editorial interpolations of factual material; and (6) any editorial omissions of (p.168) material that the author clearly intended as part of the final letter, diary entry, or essay (e.g., standardized headings).

Finally, some forms of emendation are made silently within the text, with some note of their existence in the back-of-book record: (1) expansion of unambiguous but obsolete contractions or abbreviations; (2) the supply of a missing punctuation mark that is part of a set, such as half of a pair of quotation marks, or one or more commas in a series; (3) the supply of terminal punctuation when the author began a new sentence with a capital letter but omitted the period, question mark, or exclamation point that should have preceded that character; (4) correction of an author’s lapse from his or her usual patterns of punctuation or spelling; and (5) the supply of breaks for paragraphs in a lengthy passage.

**B. Expanded Transcription**

categorized the methods used for American historical materials into three groups: “the Literal, the Expanded, and the Modernized.” As an aside, he pointed out that “in addition there is one that we might call the Garbled or Bowdlerized, which should be avoided” (95). Morison’s rules for all three of the recommended methods included providing clues to the provenance of the source text; standardizing the address, dateline, and greeting of letters; marking all editorial interpolations by square brackets; indicating editorial omissions by suspension points ( . . . ); lowering interlineations to the line; and rendering words underscored once in italics and those underscored twice in small capitals. A final instruction to all editors advised them “to prepare a fresh text from the manuscript or photostat” instead of relying on an earlier printed version (97).

Morison spent little time discussing “literal” techniques or “modernized” methods, which he approved for use only for English translations from other tongues and for “an early document, chronicle, or narrative [whose] average reader [might be] put off by obsolete spelling and erratic punctuation.” Most of his attention was directed to describing the expanded method of textual presentation that had been used in the first volumes of Boyd’s Jefferson edition, then a new and exciting addition to the literature. Unfortunately, Morison confined himself to describing his preferences for expanded transcription and noting his minor differences with Boyd, never explaining the goals or rationale of the method. His six rules for expanded transcription were hardly helpful, for they included such advice as “Retain the spelling, capitalization, and punctuation of the source text, but always capitalize the first word and put a period at the end of the sentence no matter what the writer does” (98).

Studying Morison’s precepts for expanded texts is less illuminating than examining Julian Boyd’s reasons for adopting the methods or later editors’ motives in modifying the technique in what became known as the historical tradition. The patterns of silent emendation cited by Morison were worked out by Boyd and Lyman Butterfield during their fruitful partnership at the Jefferson project in the late 1940s. Acutely aware of the loss suffered by the transfer of any eighteenth-century source text to a twentieth-century printed page, they cast about for some device that would preserve the flavor of the original materials: their solution was to print manuscript materials more or less as they would have been printed at the time of their inscription. While the device itself was not new, an attempt to apply it to American materials was. Consulting products of Benjamin Franklin’s press, the two scholars compiled what could be termed a style sheet for compositors of the late eighteenth century. The conventions used in the Jefferson, Rush, and Adams volumes edited by Boyd and Butterfield were largely those employed by the printers who were contemporaries of their statesmen subjects. In printing houses of that day, for instance, the ampersand was retained only when it was part of a name of a business firm or part of the abbreviation &c or &ca for etc.; otherwise the symbol was rendered as “and.” This usage was transferred to the printed texts published in the 1950s and 1960s. Similarly, because the thorn was no longer used in printing by the late eighteenth century, Boyd and Butterfield silently translated it to th when it appeared in a manuscript.

Unfortunately, Boyd and Butterfield assumed that their readers and fellow editors would recognize the patterns of silent emendation for what they were—the printing conventions of Revolutionary America. They believed that they were justified in imposing these conventions on the materials at hand because Jefferson, Rush, and Adams were literate men who would have expected to see such conventions imposed on any holograph materials that they submitted to a printer. In effect, Boyd and Butterfield sought to publish volumes of documents edited as Jefferson, Rush, and Adams themselves would have edited them. Butterfield recognized the method’s limitations for John Adams’s own family: letters written by eighteenth-century Adams women, often denied conventional education in standard English usage and punctuation, were emended far more conservatively lest nuances of the original be lost by imposing inappropriate conventions. Other scholars, however, missed this central point, assuming that Boyd and Butterfield’s patterns of emendation were designed to serve any documentary edition. Some attempted to transfer these methods to manuscript source texts of later periods, when the reasons for the Boyd-Butterfield printing conventions had no validity. Worse still, some used these methods with documents composed by semiliterate men and women, thus obscuring almost every bit of the original texture and flavor that Boyd and Butterfield had hoped to preserve.

Not only did editors following this tradition expand certain forms abbreviated in the original, but they also included within the text or its accompanying footnotes elements of the source text that were excluded from the final version of the document. Thus, Boyd and Butterfield and their followers, like inclusive editors in the CEAA mold, reported cancellations and insertions whose existence seemed likely to be of importance to their modern readers. Some editors used textual symbols, others used footnotes adjacent to the text. Unlike inclusive editors of the CEAA/CSE school, however, expansionist editors did not pretend to record every detail of inscription by providing a back-of-book record of such suppressed details or of their emendations of the source text.

Expanded texts can be constructed conservatively or liberally. The text of letters in the Papers of Ulysses S. Grant, for instance, is close to a diplomatic transcription of the sources. If the editors provided a back-of-book record of minor emendations, their volumes would easily qualify for the CSE seal. Volumes in the Adams and Jefferson series, however, continued the practices of traditional expansionist emendation until the first years of the twenty-first century.

While part of this diversity came from the traditions of the editions in question, another factor played an important role. Ulysses S. Grant’s letters simply required fewer silent emendations under the definitions of expanded text because they were not eighteenth-century materials. Grant and his correspondents used neither the thorn or the tailed p. Their style was already a century closer to modern conventions, and the editors of the Grant Papers did not need to bend the rules of expanded methods to create their near-diplomatic transcriptions, for those
rules simply did not apply to the textual problems they faced.

The fathers of expanded transcription based their method on another assumption that Morison and their followers ignored. Although Boyd and Butterfield were selective in including details of the source text in their volumes, they did not ignore the requirements of readers who needed access to the contents of the originals. They assumed that microfilm editions of their projects’ archives would soon make facsimiles of these source texts available to a wide audience, thus giving readers access to any details of the originals that the printed texts ignored.

The Library of Congress did issue an indexed microfilm of Jefferson’s surviving personal papers in that institution’s Presidential Papers series (now even more conveniently available online thanks to the American Memory Web site). A few other major repositories issued microfilm collections of Jefferson materials, but there has never been a supplementary publication of the items from other repositories whose materials exist as surrogate photocopies in the files of Jefferson Papers project. Similarly, Butterfield and his staff issued a microfilm edition of the Adams Family Papers at the Massachusetts Historical Society, but the project never prepared a similar microform edition of photocopies of Adams materials collected later from other repositories.

Figure 4 presents the text of this chapter’s sample document in what might be its inclusive or expanded form, followed by the textual record that might accompany the inclusive text in an edition bearing the CSE emblem. In this expanded form of the document, the abbreviation “D.C.” in line 19 has been expanded within brackets, unlike other abbreviations for personal names, because its meaning would otherwise be ambiguous (Don Cook or David Chesnutt?).

Figure 4. Inclusive text, or expanded transcription
Dear Dick

Another week in the life of the author of the ADE

"Guide": one blizzard (Tuesday) and one lunch with a
member of the "Comm[itt]ee on the Manual" (Wednesday). ¹

I've begun (nagging) to nag members of the Executive Sub-
Comm[itt]ee, to make up their minds about the date for our
planning meeting in Bloomington. Chairman Dave N. has his
summer school teaching schedule to consider. Dave C. may be
going to Italy. Paul S. doubtless has vacation plans.

I'll state once again that editing documents is a lot more fun
than writing about editing them. There are only so many ways
to say "the responsible editor should. . . ." More and more
frequently I must restrain myself from typing, "Look, fellow
editors, this is the way to do it, and don't give me any [. . .]
arguments!" ² I realize that this isn't the tone we should strive
for.

I'll let you know when the date for our Bl'ton meeting is
set—I hope that D[on]. C[ook]. can be one of our party, but
that must be left to chance. I can hardly blame him for finding
an excuse to miss two days' worth of arguments about the
history of the angle bracket as a symbol for authorial
cancellations.

More seriously, there could be disagreement among
members of the Sub-Comm[itt]ee. (Chesnut) One member
will argue that the entire book should be written in
FORSAN. (Nordloh) Another may insist that the book is
written too simplistically—that it isn't intended for those (too)
so ignorant that they confuse a stemma with a lemma.
I wish that you could be there to play peacemaker. I'll keep
you advised of all developments. Any advice-in-advance you
give will be welcomed—and, perhaps, heeded.

Yours in a quandary—
Mary-Jo

¹This paragraph was initially the second paragraph in the body of
the letter. The author has marked it for insertion at the text's
beginning.

²The phrase ", and . . . arguments" entered in the margin, with a
guideline for its insertion at this point.

C. Textual Record

The textual record for the inclusive text (see fig. 4) uses the traditional "lemma]" form for locating alterations in the
reading text. Line numbers precede the first lemma in each line.

1 9 April [1982] entered at the foot of the letter in the MS
3 "Guide";
5 Comm[itt]ee Commee; the] ye
6 the] ye
7 for] <of> for↑
VI. Clear Text

The term clear text has traditionally described the preferred method for presenting the critically edited texts of published works. The texts themselves contain neither critical symbols nor footnote numbers to indicate that an emendation has been made or that some detail has been omitted. All such emendations and alterations are reported in back-of-book tables whose citations are keyed to the page and line numbers of the new printed edition. With the publication of the Hawthorne Notebooks in the early 1970s, however, clear text was applied to CEAA-approved volumes of writings not originally intended for publication—the private writings that had hitherto received more conservative textual treatment. The decision was soon imitated by editors of the Howells Letters, the Thoreau Journals, and portions of the Harold Frederic Correspondence. Each series won CEAA/CSE endorsement, even though the CEAA and CSE guidelines of the day urged the adoption of more inclusive methods for source texts of this kind.

While there is no official standard for the adoption of clear text in documentary editing, the experience of editors who have used the technique furnishes some useful guidelines. Elizabeth Witherell, longtime editor of the Thoreau edition, provided words of caution when she recommended that clear text for manuscript source texts of private writings be used “only when a great deal of editorial emendation is required, or when almost none at all is necessary.” This apparent contradiction is easily explained. Thoreau revised many journal passages for use in lectures, essays, and books, so a page of his manuscript may contain two distinct versions of the same passage: Thoreau’s original jottings in pen, along with his later revisions of these entries in pen and pencil. Although an ingenious book designer might have improvised a method to present a single typographic facsimile, diplomatic transcription, or inclusive text giving simultaneous access to both stages of composition, the reader would be hard-pressed to make sense of the results. The Thoreau editors chose to present a clear text of the earliest version of the journal entries and to complement each volume with tables of emendations (applying only to that first version), as well as with tables of significant variants between the earlier and later stages of the journals’ physical contents. The editors did not pretend that their reading text conveyed all that could be learned from the manuscript journals; they simply provided a legible and reliable text of one of the two versions that exist in the same document.

At the other extreme, Witherell suggested that clear text is a practical option for documents that are themselves close to final versions of the documents that they represent. Neatly inscribed recipients’ copies of letters, fair copies of literary manuscripts or of political treatises—each of these is clean enough to serve for clear text. The extent of editorial intervention is so slight and of so little substantive importance that the editor can responsibly assign the record of such emendations to a back-of-book table.

The editor who weighs clear text as an option must carefully consider the requirements of the text and its prospective readers, for this treatment can suppress important inscriptional details or distort the documentary value of the resulting editorial text. If a writer customarily sent correspondents copies of letters containing canceled passages, a clear text of such letters would seriously distort the document by omitting such cancellations from the reading text. The edition’s readers would be denied immediate access to words, phrases, and paragraphs that were easily read by the letters’ recipients. Similarly, if a writer consistently dispatched carelessly proofread letters in which words were inadvertently omitted, a clear text that silently supplied the missing words would be a disservice. In both cases, the editor would have gone too far, and clear text would not be a practical and honest solution.

Before choosing clear text, an editor should analyze the emerging patterns of emendation and details of transcription that would have to be relegated to back-of-book records. If these fall into the categories that the inclusive editor would ordinarily report within the text (legible canceled passages with substantive implications, editorially supplied material for a mutilated document, and so on), then the source is not an appropriate candidate for clear text. If the evolving patterns fall into the categories that inclusive editors normally consign to the back-of-book record or emend silently, however, the source is clearly a perfect subject for clear text methods.

Clear text is justifiable in a scholarly edition only when that reading text will be accompanied by a full record of editorial emendations and suppressed inscriptional details or a companion facsimile. Editorial intervention of any
degree can be justified only when readers are provided with a complete report of what they have been denied by the editor's decision to be exclusive rather than inclusive.

Figure 5 provides the clear text version of the sample document shown in figure 1. It is followed by samples of the textual record that might accompany it, one using symbolic methods and the other the descriptive approach.

Figure 5. Clear text

<table>
<thead>
<tr>
<th>1</th>
<th>9 April 1982</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Dear Dick</td>
</tr>
<tr>
<td>3</td>
<td>Another week in the life of the author of the ADE “Guide”:</td>
</tr>
<tr>
<td>4</td>
<td>one blizzard (Tuesday) and one lunch with a member of the</td>
</tr>
<tr>
<td>5</td>
<td>“Committee on the Manual” (Wednesday).</td>
</tr>
<tr>
<td>6</td>
<td>I’ve begun to nag members of the Executive Sub-Committee to</td>
</tr>
<tr>
<td>7</td>
<td>make up their minds about the date for our planning meeting in</td>
</tr>
<tr>
<td>8</td>
<td>Bloomington. Chairman Dave Nordloh has his summer school</td>
</tr>
<tr>
<td>9</td>
<td>teaching schedule to consider. Dave Chesnutt may be going to</td>
</tr>
<tr>
<td>10</td>
<td>Italy. Paul Smith doubtless has vacation plans.</td>
</tr>
<tr>
<td>11</td>
<td>I’ll state once again that editing documents is a lot more fun</td>
</tr>
<tr>
<td>12</td>
<td>than writing about editing them. There are only so many ways to</td>
</tr>
<tr>
<td>13</td>
<td>say “the responsible editor should . . . .” More and more</td>
</tr>
<tr>
<td>14</td>
<td>frequently I must restrain myself from typing, “Look, fellow</td>
</tr>
<tr>
<td>15</td>
<td>editors, this is the way to do it, and don’t give me any</td>
</tr>
<tr>
<td>16</td>
<td>arguments!” I realize that this isn’t the tone we should strive for.</td>
</tr>
<tr>
<td>17</td>
<td>I’ll let you know when the date for our Bloomington meeting</td>
</tr>
<tr>
<td>18</td>
<td>is set. I hope that Don Cook can be one of our party, but that</td>
</tr>
<tr>
<td>19</td>
<td>must be left to chance. I can hardly blame him for finding an</td>
</tr>
<tr>
<td>20</td>
<td>excuse to miss two days’ worth of arguments about the history of</td>
</tr>
<tr>
<td>21</td>
<td>the angle bracket as a symbol for authorial cancellations.</td>
</tr>
<tr>
<td>22</td>
<td>More seriously, there could be disagreement among members</td>
</tr>
<tr>
<td>23</td>
<td>of the Sub-Committee. One member will argue that the entire</td>
</tr>
<tr>
<td>24</td>
<td>book should be written in FORSAN. Another may insist that the</td>
</tr>
<tr>
<td>25</td>
<td>book is written too simplistically: that it isn’t intended for those</td>
</tr>
<tr>
<td>26</td>
<td>so ignorant that they confuse a stemma with a lemma.</td>
</tr>
<tr>
<td>27</td>
<td>I wish that you could be there to play peacemaker. I’ll keep</td>
</tr>
<tr>
<td>28</td>
<td>you advised of all developments. Any advice-in-advance you give</td>
</tr>
<tr>
<td>29</td>
<td>will be welcomed—and, perhaps, heeded.</td>
</tr>
</tbody>
</table>

Yours in a quandary—
Mary-Jo

A. Textual Records

1. Symbolic Method

The textual record for the clear text in figure 5 is presented in the symbolic format used by the Howells Letters. Editorial emendations are reported in the “lemma] manuscript reading” style, with heavy reliance on traditional textual symbols. Suppressed details of inscription are indicated without the citation of the lemma unless confusion might result from abbreviated treatment.

1 9 April 1982] “9 April” entered at the foot of the letter in the MS; “1982” added by the editors.
3 Guide] ~ <, >
3–5 This was initially the second paragraph of the body of the letter. The author marked it for insertion at this point.
5 Commee
<nagging>↑to nag↓ they; Sub-Comm.

date <of>↑for↓

Nordloh] N.

Chesnutt] C.

Smith] S.

a <h>lot

Look] ~ ^

“and don’t give me any [. . . ] arguments” added in the margin, with a guideline directing its insertion after “it”

{p.177}

Bloomington] Bl’ton

Don Cook] D.C.

I <co> can

Sub-Comm. <Chesnutt>↑One member↓

<Nordloh>↑Another↓

<so>↑<too>↓ ignorant to ↑<that they>↓ confuse

“I wish . . . peacemaker. I’ll keep . . . of all developments.” Originally, “I wish . . . peacemaker” was the concluding sentence of the preceding paragraph. The author marked it to begin the letter’s last paragraph and marked the following sentence to run on as part of that paragraph.

↑give↓

↑—↓

2. Descriptive Method

This alternative textual record for the clear text in figure 5 employs the descriptive method for details of inscription, which employs the following abbreviations:

*ab.* = above  *del.* = deleted  *interl.* = interlined

*aft.* = after  *inserted.* = inserted

9 April 1982]↑“9 April” entered at the foot of the letter

“Guide”] comma *del. aft.* Guide

This was initially the second paragraph of the body of the letter. The author has marked it for insertion at this point.

Committee] Commee


for]↑interl. ab. del. “of”

Nordloh] N.

Chesnutt] C.

Smith] S.

{p.178}

lot] aft. del. “h”

Look] ~ ^

and . . . any arguments] inserted in margin with illegible cancellation aft. “any”

Bloomington] Bl’ton

Don Cook] D.C.

can] aft. del. “co”

Sub-Committee] Sub-Comm.; One member] interl. ab. del. “Chesnutt”

Another] interl. ab. del. “Nordloh”

so . . . confuse] del. “too” ab. del. “so”; “that they” interl. ab. “to”

I wish . . . of all developments.] Originally, “I wish . . . peacemaker” was the concluding sentence of the preceding paragraph. The author marked it to begin the letter’s last paragraph and marked the following sentence to “run on” as part of that paragraph.

give] interl. aft. “you”

—] interl. with caret
VII. Electronic Publication

Electronic publication of facsimiles or editorial transcriptions of source texts broadens the choices open to editors and users of these editions. Digitally scanned images of the sources can conveniently be paired with a variety of parallel texts. Hypertext links, for instance, give users access to the image of the source as well as a machine-readable clear text transcription that might serve the needs of a novice, with links to a textual record that notes details of inscription and a diplomatic transcription that might serve an advanced student. The Web component of this Guide offers such an example for the sample document used in this chapter.

VIII. Conclusion

Even with the newest technology, the oldest rules apply. The source text, likely audience, and publishing medium will dictate the choices new editors will make concerning their editions’ textual presentation, and there are cautionary lessons to be drawn from earlier enterprises. The sample texts presented here provide a few basic lessons.

The fact that the author of that 1982 letter used the thorn as shorthand in drafting correspondence is of some interest, but it could easily be reported in a general statement on her style without the need to reproduce every such symbol as “ye” or “[th]e” or to supply a textual note every time the thorn appears. The treatment of revisions and substitutions in the clear text, however, conceals not only changes in wording and tone but also factual information, such as the identities of the committee members to whom she referred. In each text, the conventions peculiar to the editorial method used show their own virtues and limitations.

The history of these editorial methods provides even more important lessons. Boyd and Butterfield, for instance, grounded their justification of expanded transcription on the assumption that some kind of comprehensive image edition of their source texts would be available to the public in timely fashion. However, no such comprehensive facsimile edition has been published or is now contemplated. Boyd and Butterfield erred, too, in assuming that their readers would understand the reasons for the editorial policies they adopted. No editor is justified in automatically adopting conventions or policies of emendation used by a colleague. The textual methods of each edition must be designed to suit the materials that provide source texts. “Generally accepted,” “traditional,” and “time-honored” are adjectives that should not be used to justify the adoption of any editorial practice. In choosing textual methods, each editor must start afresh, making decisions.

Suggested Readings

The most valuable discussions of editors’ choices of one textual method over another appear in the introductions to the editions themselves. Much of this material is collected in Editing Historical Documents: A Handbook of Practice, edited by Steven B. Burg and Michael E. Stevens. An earlier attempt to survey a wide variety of methods is G. Thomas Tanselle’s “The Editing of Historical Documents.” An interesting example of the modern application of different methods to the same source text can be seen in the Franklin Papers edition of The Autobiography of Benjamin Franklin (New Haven, 1964), and in J. A. Leo Lemay and P. M. Zal’s genetic text of the same work.

An interesting summary of the arguments concerning the responsibility of editors of private source texts appears in CEAA Newsletter 3 (June 1970): {p.180} 16–21. Hilary Jenkinson’s “The Representation of Manuscripts in Print,” published in 1934, is worth rereading, if only for its reminder to editors: “It is possible to make the path too smooth.”

Documentary editors have confined their remarks on the use of symbols or descriptive notes for textual problems in unpublished writings largely to the introductions to volumes in which such symbols or notes are employed; many of these are collected in Burg and Stevens’s Handbook of Practice. The Handbook provides a survey of practices comparable to G. Thomas Tanselle’s analysis of methods for published literary works in “Some Principles for Editorial Apparatus.”


In addition to the essays cited in the text, the student of descriptive textual annotation should see Joel Myerson’s review essay “The Autobiography of Benjamin Franklin: A Genetic Text” in the Newsletter of the Association for Documentary Editing; and Fredson Bowers’s comments on Myerson’s remarks in a letter to the editor of the newsletter in September 1982.

For a more detailed discussion of the options available in the present century, we refer you to chapters 6 and 7, below, and to links from the Guide Web site.
Chapter Six

From Transcriptions to Editorial Texts

I. Putting Transcriptions to Work
   A. Establishing the Editorial Texts
   B. Back-of-Book Textual Notes

II. Exceptions to Some General Rules
   A. Documentary Problems with Textual Solutions
      1. Genetic Elements in Source Texts
      2. Multiple-Text Documents
      3. Conflation

III. Translations
   A. Foreign-Language Materials
   B. Shorthand
   C. Codes and Ciphers
   D. Telegrams

IV. Current “Good Practice”

V. Conclusion

Suggested Readings

Inevitably, the time comes when all editors must convert statements of practice and procedure into editorial products. Transcriptions will evolve into established texts using whichever sets of textual conventions have been chosen and whatever scheme of proofreading and verification is appropriate. Translations of materials in foreign languages, shorthand, or ciphers will join those transcriptions as working copy for the edition. And, of course, earlier confident statements of what is “right” for the edition will have to be modified as the editors become even better acquainted with those documentary texts.

I. Putting Transcriptions to Work

Even for editors who are their own transcribers, transcribing a source text is only the first step in a very long process. An editor-transcriber needs to concentrate on translating the originals as conservatively as possible instead of assuming, in advance, that he or she knows what textual methods will finally be applied to the initial transcription. When the duties of transcriber and editor are divided, responsibility for such emendations belongs to the editor, not to the initial transcriber. Making those transcriptions useful working copies begins with choosing their appropriate format and location.

It may be necessary to provide hard-copy printouts rather than expect that proofreading and annotation can be conducted with an on-screen image. This is a function of the visual capacities of those who’ll have to read them. If transcriptions are printed out, they need to be filed carefully in notebooks or kept in file drawers. Some find it convenient to separate each transcript from its neighbors by a sheet of tinted paper, this leaf to be used by the editor in drafting annotation. Whatever the method chosen, each transcription should be accompanied by some sort of separate annotation sheet or other convenient place for special queries or cautions concerning the source texts to which they refer. On this sheet the transcriber can remind the editor that the source text was an enclosure in another item or indicate special reasons for any apparent novelties in treatment.

Computers have simplified the process enormously. Word processing makes the design of special headings for such annotation sheets a simple matter. Editors who can handle on-screen editing will provide equivalent electronic files for annotation, and transcriptions will be saved with appropriately named files for these additions. Projects using a content management system can relegate all this to the electronic file that accompanies the document, eliminating much of the paperwork that needs to travel with the document.

A. Establishing the Editorial Texts

No matter the form of the transcription files, the initial transcriptions must be reviewed and reviewed again to ensure that they meet the standards of a documentary series. We’ll begin with a brief glossary of the terms and techniques that you’ll need in the process:
1. **Proofreading** traditionally indicates oral proofreading involving two or more people, where one member of the editorial staff “holds” the newer version of the text while a senior member of the team reads aloud word for word (or letter for letter in an old-spelling edition) and every punctuation mark from the earlier textual version on which the later transcription, galley proof, or page proof is based. Thus, when proofreading a source text’s initial transcription, one editor reads aloud from the source text while the second follows the characters on the typescript.

2. **Visual collation** occurs when a single editor compares two versions of the text visually. To increase accuracy, it is customary to place a ruler or other straight-edged device under each line in both versions so that one’s eye will not accidentally skip an inscribed, typed, or printed line in either copy. Some projects refer to this as “verification.”

3. **Mechanical collation** originally required a device that detected variants between two printed texts presumed to be identical. The earliest machine of this kind, the Hinman Collator, made possible the convenient collation of sample copies from each printing of a published work. Today, mechanical collation is more likely to take the form of more accurate electronic collation, in which two machine-readable versions of an edition’s text or notes are compared by computer.

4. **Verification** is a term with two distinct meanings: either checking the accuracy of editorial transcriptions made from photocopied source texts against the originals for each text, or checking the accuracy of the contents of informational annotation. Verification of texts may have to consist of a visual collation, although team proofreading is preferable. Verification of quotations in notes follows the edition’s general policies for proofreading or visual collation of documents printed in their entirety. The verification of other elements in the notes should always be performed by someone other than the original annotator.

5. **A record of corrections.** Any project should have a well-established system for recording corrections to transcriptions and the textual forms that follow. That record not only documents the changes that have been made but who made them—and when. This record should be established as soon as documentary transcriptions begin their trip through the editorial process.

Some of the terms and practices described above refer to processes employed by editors in the CEAA/CSE tradition. Editors without such stipulations can ignore some of these stipulations, but not the reasons for their existence.

1. No one denies that team proofreading is a more effective insurance against error than visual collation. It’s especially useful in the transcription of nonprinted source texts, for oral proofreading ensures that the person who reads aloud (preferably the senior editor responsible for the edition’s consistency) will not be influenced by the interpretation of the transcriber, even when a solo editor was the transcriber. Whenever the oral reading produces a character, word, or phrase at variance with the transcription, the editor naturally pauses to reevaluate his or her interpretation; but that reassessment should come with a fresh eye, uninfluenced by what someone else has seen in the source text.

2. For printed source texts, the differences in accuracy between proofreading and visual collation may be less marked, for fewer instances of subjective interpretation arise. But any veteran of the process can attest that the visual collation of printed sources easily leads to skipping lines in either the source or its transcription. Those who have experimented with both methods report that oral proofreading is far less tiring to the eye (although not, obviously, to the voice) than visual collation, where editors must continuously switch their fields of vision as they move from primary to secondary textual version while remembering what they’ve just seen.

3. Considerations of time and staff size often make impossible the number of independent proofreadings the editor might wish for projects that have externally produced page proofs. For projects that must compromise on the matter of independent, team proofreadings, the use of cassette tape recorders offers a useful supplement. The member of the team reading from the source can record that part of the proofreading process at his or her convenience, while the second team member can check those recorded, spoken words against the transcription or page proof later. Two or more staff members may do such checking against the same tape, thus producing semi-independent proofreading sessions. Tape-recorded proofreading has an advantage beyond that of easing scheduling problems. Any member of the team can interrupt proofreading when voice or eyes begin to falter and return to the task later without inconveniencing a colleague.

4. Even when the editorial text will not receive a number of multiple formal proofreadings, the editor should perform at least one proofreading against the source text. This should be done as late as possible in the editorial process, preferably immediately before copy is formatted for printing. This rule merely recognizes the fact that editors become progressively more familiar with the peculiarities of their source texts. In many cases the preparation of informational annotation will also make the source more intelligible. Proofreading against the source text immediately after transcription will be markedly less accurate than proofreading performed weeks or months later.

5. Ideally, all transcriptions should be perfected against the originals of their source texts, not merely against photocopied versions, but this is often impractical. When it is not feasible, the edition’s introduction should make this omission clear. Whenever verification of a given document or group of documents is performed by someone other than one of the editors, this fact should also be noted. Verification of transcriptions against originals is required of editions aspiring to CSE approval.
Traditionally, some editors preferred to proofread or collate page proofs of documentary texts against the source texts instead of against transcriptions. This practice provided a last chance to catch errors of interpretation or mistranscription. Editors who adopted this method usually did so only if the same kind of proofreading had been applied to the printer’s copy before it was sent to the compositor. Publishers do not look kindly on the long lists of author’s alterations to page proofs that could otherwise arise from the technique. Of course, computer technology has now so revolutionized American publishing that many editors would be hard pressed to define what constitutes “page proofs” for their editions. Some are themselves the producers of the page proofs that go to the press’s production department, while others use publishers who prefer to prepare materials for publication themselves. Each editor will have to identify the appropriate point at which to make such a final check.

Proofreading should be supplemented by visual collation and simply reading both documentary texts and annotation for sense. Transcriptions, compositor’s copy, galleys, and page proof should be read by as many people as possible for accuracy, clarity, and consistency. Solo editors and projects with small staffs are wise to enlist professional colleagues as reviewers of the edition.

B. Back-of-Book Textual Notes

For many editors, textual responsibility does not end with proofreading and perfecting the editorial reading text. Whenever an edition relegates all or part of its record of emendations or details of inscription to the back of the book in a print edition, or to a separate screen in an electronic one, the editor must establish an accurate reporting system for that apparatus as early as possible. Even though these records will be checked again and again, the possibility of error is reduced substantially if their format is established in advance.

The process of such reporting may begin with the transcriber. If instructed to keyboard only final authorial intentions in letters, journals, or draft works, the transcriber must also initiate a record of what is omitted—authorial deletions or interlineations or nonauthorial contributions in the source text. Word-processing equipment now provides the luxury of keying the deleted or interlined material within a set of tags reserved for each detail of inscription, for instance “∫” and “∫” for the beginning and end of interlined material instead of creating physically separate files. When an edition has already established a method for reporting such details (that is, through symbols, narrative description, descriptive abbreviations, or a combination of these methods), the transcriber can enter remarks in their proper predetermined form.

When the editor reviews the transcriptions, the process of correcting minor authorial errors or slips of the pen begins. Once, changes such as suppressed inscriptive details had to be recorded on index cards with references to the transcription’s page and line number. Editions that provided separate records of inscriptive details and editorial emendations had to distinguish between the two groups from the moment record keeping began. Word-processing equipment eliminates such dual record keeping, for the editor can designate separate codes for each sort of textual record in advance, and the appropriate codes can be entered for each note.

No matter what their policies in reporting emendations and inscriptive details, all CSE editors have traditionally supplied one kind of textual record in the back of the book—the report of ambiguous end-of-line hyphenations. This report lists all such ambiguous authorial hyphenations (possible compound words whose end-of-line division coincides with the position of the hyphen) in the source as well as any new ambiguous hyphenations introduced in a modern printed edition. Such a record allows scholars to quote accurately from the new text. To identify such ambiguities in the source, the editor can refer both to dictionaries contemporary with the document’s inscription and to the author’s customary usage. Most projects create an in-house record of their author’s preferences in hyphenating specific compounds. Complete consistency is too much to expect, but useful patterns emerge. Obviously, keeping such a lexicon in a computer file expedites processing of the entries, and patterns can be recognized earlier.

Once the editor has established that line-end hyphenation occurs at a point where the author would ordinarily have hyphenated a compound, that hyphen is marked in the transcription for retention in the print edition should typesetting place the word in the middle of the line. The Thoreau edition streamlined the scope of its hyphenations by forgoing a justified, or consistently even, right-hand margin in printed volumes of Thoreau’s private writings, thus ensuring that no new hyphens were added when the documentary texts were typeset. The Thoreau editors needed to report only the retention or omission of ambiguous line-end hyphens that appeared in the source texts themselves. Editors whose editions employ justified right-hand margins in their editions still need to check page proofs (or their electronic equivalents) for new and potentially misleading hyphens.

Since all printed back-of-book textual records are keyed to typeset lines, not to footnote numbers within the texts, the final preparation of their contents must await page proofs. Work on these lists will have begun far earlier, and records of emendations and inscriptive details can be keyboarded and proofread well ahead of time. The editor should first devise a format for such reports, generally one in which the first column for each entry in a list bears a reference to the line and page of the transcription or printer’s copy. After page proofs are reviewed, new page-line references to the print edition can be substituted.

Editors should remember, too, that they are responsible for making their textual records easier for readers to
reflects the development of Any document—letter, state paper, literary work, scientific essay, laboratory model—can survive in a form that revisers along the way. legislative records to show the genetic levels of their evolution and the specific contributions of collaborators and versions of these same documents. Documentary editors in the “historical” tradition long ago re-created drafts of American history showed little reluctance to provide their readers with editorial reconstructions of texts showing the historical documents. Paradoxically, scholars who bristled at suggestions that critically achieved, idealized clear do parallel texts in bound volumes. Versioning has a long and honorable history among editors of American more than one version of a text rather than a single clear text from which the various prior versions would need to compare for themselves two or more widely circulated basic versions of major texts” (167).

II. Exceptions to Some General Rules

Although American documentary editors have worked hard to create guidelines for the creation of documentary editorial texts for almost every form of recorded evidence, situations arise in which editors must admit defeat. After designing an editorial technique appropriate to the bulk of the sources for an edition, editors follow that technique until they encounter a situation in which the standard documentary formula of “one source equals one editorial text” doesn’t apply. Whenever the equation proves invalid, they turn to other traditions for appropriate solutions. Even here they tend to apply such borrowed techniques conservatively, pointing out to their readers where editorial judgment or guesswork has been employed rather than concealing that fact in the name of elegance or readability.

A. Documentary Problems with Textual Solutions

Even the editor following the most conservative general policies on emendation will encounter situations in which one source text will not provide one editorial text. The summary of methods of inscription and forms of documentary records in chapter 4 hints at some of the occasions on which a documentary editor borrows the methods of specialists in related fields. Any editor of orally communicated texts, for instance, {p.188} may have to deal with the theoretical implications of dissonant witnesses to a lost archetype, the central problem of the classicist. Even editors spared this special form of documentary record face the possibility that the best method for editing a specific source text may be one from the tradition of critical and not documentary publication.

In recent years, the debate among literary scholars over the value of eclectic clear texts versus editorial methods that retain more “documental” elements has presented useful lessons for the editor of historical documents that have little claim to literary merit. These discussions help determine just when noncritical texts serve the purposes of an audience and when critical methods are more appropriate. Donald Reiman discusses the question in his essay “ ‘Versioning’: The Presentation of Multiple Texts.” Looking back over decades as an editor of the writings of Shelley and His Circle, Reiman remarks, “I have become less and less confident that an eclectic critical edition is the best way to present textual information to scholars.” Even for literary works, he argues, the public may be better served by “enough different primary textual documents and states of major texts . . . that readers, teachers, and critics can compare for themselves two or more widely circulated basic versions of major texts” (167).

Reiman advocates “versioning” rather than “editing”: giving the read-ing public equally convenient access to more than one version of a text rather than a single clear text from which the various prior versions would need to be laboriously reconstructed from textual notes. Electronic publication, of course, offers one form of versioning, as do parallel texts in bound volumes. Versioning has a long and honorable history among editors of American historical documents. Paradoxically, scholars who bristled at suggestions that critically achieved, idealized clear texts showing the final intentions of the authors of letters or state papers served any useful purpose for students of American history showed little reluctance to provide their readers with editorial reconstructions of preliminary versions of these same documents. Documentary editors in the “historical” tradition long ago re-created drafts of legislative records to show the genetic levels of their evolution and the specific contributions of collaborators and revisers along the way.

1. Genetic Elements in Source Texts

Any document—letter, state paper, literary work, scientific essay, laboratory model—can survive in a form that reflects the development of {p.189} the author’s intentions, preserving not only a final text but also the false starts, preliminary wording, and stylistic evolution of that text. Few editors escape confronting source texts that carry intrinsic clues to their genesis. These are encountered most commonly when an editor’s source text is a manuscript obviously revised during composition. An edition of a draft letter or an author’s holograph corrections and additions to galley proofs or the pages of an early printing used to prepare a revised edition demand identification of original, intermediate, and final versions of the same document.

Genetic editions of texts try to offer the reader access to more than one level of textual creation within a single inclusive page. While the term genetic text edition came into usage in German textual studies more than fifty years ago (see Hans Walter Gabler, “The Synchrony and the Diachrony of Texts”; and Louis Hay, “Genetic Editing, Past
and Future”), it did not become current in American studies until the 1960s. In editing the successive draft versions of Billy Budd, Melville’s editors adopted the term genetic text to describe their diplomatic transcription of the manuscripts of that work, which had been left unpublished at Melville’s death. The genetic elements of the transcription were the result of their painstaking efforts to devise a system of symbols and descriptive abbreviations that would allow the reader to understand the order in which the changes were made by the author. In a single set of pages, a densely packed trail of symbols led the reader through two, three, sometimes four versions of the same passage.

The genetic text of Billy Budd is one of the most complicated and sophisticated products of modern scholarly editing. Simpler genetic texts have been with us since the first editor presented an inclusive or conservatively expanded text of a handwritten draft. Any editorial method that includes the use of symbols for deletions, insertions, and interlineations can present a genetic text for individual documents. Editors who eschew the use of textual symbols can instead give their readers clear texts of the final version and supply notes that permit the readers to construct their own genetic version.

a. Synoptic Genetic Texts

More sophisticated problems of conflation arise when the genetic stages of a document are recorded in not one but several source texts. If the variants between these preliminary versions are wide, the editor may print each document separately in parallel texts or treat each one as a distinct version. It may be, however, that these separately inscribed evolutionary stages of the text are so directly related that they represent a direct intellectual line of revision. In this case, the editor has the option of creating a “synoptic” text, another term and technique from classical scholarship.

This form of editing is as old as the synoptic Gospels, but the term synoptic text was borrowed for modern works by the editors of the James Joyce edition when they described their methods in editing Joyce’s Ulysses. Joyce’s revisions of the novel survived, not in a single draft manuscript, but in manuscript fragments, corrected galleys, and other forms. The Joyce editors combined the information contained in these separate documents to create a synthetic genetic text, a synopsis of information from several source texts combined in one new editorial text. The process of textual synopsis is not confined to biblical scholarship and editions of great literary works. The first volume of the Documentary History of the Ratification of the Constitution includes two synoptic texts that trace the evolution of the articles of the American Constitution through the debates of the Philadelphia Convention of 1787. The editors did not have at hand separate copies of the Constitution reflecting its wording at every stage of its consideration. Instead, they worked from four source texts: the draft constitution submitted to the Convention on 6 August 1787; the text of the Constitution recorded in the Convention journals on 10 September; a printed copy of the report of the Committee on Style, which revised the articles between 10 and 12 September; and the text of the Constitution as adopted on 17 September. To supplement these sources, the editors analyzed James Madison’s notes of debates in the Convention, records that indicated the date and nature of each revision of the frame of government.

The successive surviving versions of the Constitution, like those of Joyce’s Ulysses, qualified as source texts for a synopsis because they were similar enough to allow the editors to draw valid conclusions about the sequence of revisions at each point. The editors reprinted the articles of the Constitution as adopted by the Convention, thus supplying a reading text of the final stage of the evolution to parallel their synthetic texts.

Synoptic treatment of the records of deliberative political bodies is fairly common among editions dealing with governmental history. The Documentary History of the First Federal Congress: Legislative Histories provides three volumes of legislative histories of the bills considered by the U.S. Congress, 1789–91. They contain a calendar recording every action on a given bill or resolution, as well as transcriptions of surviving manuscript or printed sources of the texts for these measures as introduced to the Congress. Footnotes record amendments to the originals. In a few cases, when the original version of the item has vanished but its final text and a complete record of amendments survive, the editors re-created the original’s text by taking the final version and adding or subtracting words or phrases recited in the amendments. In cases of such heavy editorial intervention, the re-created text is accompanied by notes clearly tracing the work involved. In other instances, when a committee report of amendments to a bill has survived, the Legislative Histories provide readers with both a literal transcription of the committee report and the amendments identified in that report, reprinted as footnotes to any previous versions of the bill.

b. Collaborative Source Texts

The attentive reader has noticed that the examples of synoptic genetic texts from the Ratification and First Congress projects not only draw on multiple source texts but also represent the intellectual contributions of more than one author. Editors of the records of American political and military history have long been concerned with the need to identify not merely the stages of a document’s evolution but the separate authors of each revision. In recent years,
editors of literary works have come to share this interest.

Source texts that represent collaboration between two or more persons are often more challenging than genetic documents by a single author, and study of such collaborations received close scrutiny and analysis in the decade following Jerome McGann’s discussion of the socialization of literary texts. Ideally, both the identity and the specific contributions of each reviser should be recorded. Modern authors often work closely with publishers’ editors. Scholarly editors employing traditional copy-text theory to construct a single text reflecting final authorial intentions face serious difficulties in such situations, for they must somehow determine which suggestions from an author’s contemporary editor were imposed upon the author and which were accepted freely, perhaps with the author’s thanks. If authorial intentions are the overriding criteria, the critical editor must then exclude from the new edition passages added without the author’s full approval while retaining those to which the author gave hearty consent. The pitfalls of this approach are obvious, and attempts to impose it at inappropriate times and places provide critics of the application of emended copy-text to modern writings with some of their most telling attacks and led to the notion of “socialized” texts.

Editors of historical documents, especially in legislative and professional affairs, are veterans of dealing with such collaborative documents, for such records frequently represent action by committee. A manuscript report or public paper may contain passages in the hands of two or more legislators assigned to prepare that document. The rough draft of a state paper may reflect the fact that an executive assigned its preparation to one aide, circulated the draft to other advisers, while still other men may have approved or vetoed their suggestions—leaving the record of all these actions on the same scribbled, dog-eared set of pages.

The collaborative aspects of a document’s composition can often be represented quite easily if one contributor had primary responsibility for its drafting. This fact can be stated in the document’s source note, and the editor can focus on additions, revisions, or deletions in other hands. Such records may be provided by using a special form of symbol enclosing such additions (e.g., “The document was originally inscribed by AB. All revisions by CD appear in the text within square brackets.”). But if the collaborators’ contributions are fairly equal, or if a third or fourth writer is involved, the editor must consider descriptive notes to supplement the text. Each addition, deletion, or revision might be keyed to a footnote explaining that the words or phrases in question were “added above the line by CD” or “entered by EF in space left in the MS” or “rewritten by GH.” This is the method chosen by the editors of the Eisenhower Papers in the document discussed in chapter 3. Electronic editions of such source texts, of course, could represent these elements in different colors of type or special typefaces, with hyperlinks to any necessary editorial explanations.

c. The Physical Presentation of Genetic Texts

Obviously, electronic editions of any form of genetic text—single source, synoptic, or collaborative—can solve many problems of presentation gracefully and effectively. But even these new solutions demand careful thought, planning, and testing of formats, while book editions of such sources remain one of the cruelest tests of an editor’s ingenuity. Clearly, the choice among a truly genetic text, an inclusive text supplemented by textual notes, and a clear text is both a theoretical and a practical one. The editor will need to find the device that enables the reader to reconstruct inscriptions in the source text with the finest distinctions possible. The use of numbered footnotes with a clear text can serve this purpose, although the multiplication of superscripts necessary to record numerous significant revisions runs counter to the purposes of the editor who wishes to avoid “disfiguring” the documentary text. Clear text with a back-of-book record cannot serve the reader when the collaborative aspects of the source are central.

Some documents defy efforts to reproduce all the details of their genesis in comprehensible form: the proliferation of symbols or footnote numbers would make the result unreadable. In such cases, it has been traditional to offer the reader parallel texts of the same material. In the classic form, the two texts were truly parallel, printed in two columns on the same page or on facing pages (as in Bowers’s edition of Leaves of Grass), but modern editors use the term and technique more broadly.

The need to communicate genetic textual elements inspired much of the early investigation of electronic versions of source texts. Hypertext or hyperlink publication was an obvious supplement to or substitute for a printed text in these instances. Any editor who used word-processing software with the ability to switch between documents in “windows” easily saw the potential of electronic methods for linking the editorial transcription to different levels of a document’s genesis. This form of publication is a simple, effective, and logical way to communicate the genetic process to a reader.

2. Multiple-Text Documents

Similar treatment can be accorded multiple-text documents, a term coined by David Nordloh to describe sources inscribed in such a way that the reader could reasonably extract the texts of two or more distinct documents from the characters that appear on the same page or set of pages. In multiple-text documents, the entries on the same
page are so widely separated in time, intention, and even authority that they must be regarded as separate examples of one author’s writing or as examples of the writing of two or more authors. Their textual problems are distinct from those in ordinary drafts, in which authors leave records of their evolving intentions for works before their completion.

Notebooks used as literary commonplace books are an obvious example. Thoreau first inscribed entries in his journals in ink. Months or even years later, he reworked many of these passages for publication, considerately making his revisions and emendations in pencil and ink \({p.194}\) that could be distinguished from the original inscriptions. The Thoreau edition provides a clear reading text of Thoreau’s original entries. Textual notes in the Thoreau Journals reproduce only those revised passages that never achieved print publication, and the same notes can refer the reader to journal entries that served as the basis for works published in Thoreau’s lifetime and appearing in other volumes of the edition.

a. “Second Thoughts”: Authors Who Try to Rewrite History

While most authors of letterbooks or diaries are content to leave those documents untouched once their pages have been inscribed, others can’t resist the temptation to go back to improve youthful lapses in style or delete compromising passages. Historical figures who use their years of leisure and retirement to rewrite diary entries and other portions of their personal papers create a special form of purgatory for their editors. When these revisions were made for a published memoir, the editor can at least refer readers to an authoritative text of the final version of the material in another section of the edition or in the printed source. But when the revised sources remained unpublished in their own day, the modern editor is left with documents whose texts have been deliberately corrupted by their own creators. Providing access to such revised variant passages was a consistent problem for the editors of the Madison, Lafayette, and Washington papers.

When Madison set out to compile autobiographical material, he rewrote his own retained copies of correspondence (or recovered addressees’ versions from his correspondents or their heirs), revising the pages to suit his matured notions of style and discretion and adding marginal comments to the documents. Fortunately, Madison’s later emendations are usually distinguishable from his original inscriptions. When one of these “corrected” manuscripts had to serve as a source text, the Madison editors could recover the original words and marks of punctuation, discussing later revisions in footnotes. This method of emendation provides readers immediate access to the texts of letters and state papers in a form that gives them validity as documents of American political history. At the same time, the notes enable them to determine areas where Madison felt correction or suppression was necessary to make these materials ready for publication to the world.

The Marquis de Lafayette was a more systematic memoirist. In the early nineteenth century he revised not only personal letters but also his 1779 manuscript “Memoir.” These revisions were incorporated into \(\{p.195\}\) transcribed copies that Lafayette then sent to Jared Sparks, and most of Lafayette’s emendations were reflected in the published version, \(\text{Mémoires}\). The Lafayette editors collated all printed versions of the \(\text{Mémoires}\) against the emended manuscripts to establish the pattern of the author’s revisions, disregarding later revisions of the letters and the manuscript “Memoir” that were “purely stylistic.” “Significant passages” deleted in the manuscripts or omitted from the printed \(\text{Mémoires}\) appear in the new edition within angle brackets. Any other changes deemed significant by the modern editors are treated in footnotes.

George Washington’s motives in revising his early letterbooks are less clear than Madison’s and Lafayette’s. The volumes contained autograph copies of outgoing correspondence that Washington laboriously inscribed during the Braddock campaign on the western frontier in 1755. Beginning some thirty years later, he began to emend the letterbooks, directing a clerk to copy the revised texts of the letters into a new set of letterbooks. All but one of the original letterbooks vanished, and few correspondents saved the letters they had received from the young Washington. For those months for which only the emended, later copies of letterbooks survived, the Washington edition had no choice but to use those as source texts. For the summer of 1755, however, they had recourse to the original letterbook, which they used as source text and supplemented with notes reproducing Washington’s later notions of “improvement” as extrapolated from the later letterbooks.

One of the most remarkable recent feats in documentary editing was a volume, presenting a previously unpublished manuscript, that combined many of the problems faced by these other projects with a few unique to itself: Rosemarie Zagarri’s edition of David Humphreys’ “Life of General Washington” with George Washington’s “Remarks.” Humphreys set out to write the earliest and only authorized biography of his former military chief. Washington provided his former aide with source materials, as well as manuscript “remarks” for his guidance. Humphreys never completed his book, although he did incorporate many passages from his draft into shorter pieces about the first president. Questions of synoptic genetic texts, multiple-text sources, and authorial second-guessing arose as Zagarri reconstructed the results from fragments now scattered physically among several repositories and intellectually among other Humphreys writings, incorporating Washington’s comments in the text in angle brackets. George Billias, one of the sharpest contemporary reviewers of documentary editions of the Revolutionary era, remarked in \(\{p.196\}\) wonder, “Incredible as it may seem, this book actually contains new material about
George Washington, one of the most thoroughly researched figures in all American history” (review of Humphreys’s “Life of General Washington”).

b. Other Multiple-Text Documents

Several forms of multiple-text documents may present problems in transcribing the source but do not pose deep theoretical questions. For example, a writer may have used an existing document as scratch paper for drafting another letter or report. A sheet of paper may carry a letter received by John Smith on one side and Smith’s draft of his reply on the other. Some frugal eighteenth-century figures carried this practice to the extreme of drafting replies to a letter over that letter’s own lines, inscribing the new draft at right angles to the old lines. Such practices result in two documents that are part of the same physical whole. Although the textual notes that describe the provenance of each item must indicate that it is physically a part of the other, no special textual problems will arise.

A more common form of multiple-text document is created when an author becomes a reader, making notes or comments in the margins of the reading matter at hand. While marginalia are more commonly found on printed works, some writers were equally eager to record their own comments on newly received letters or other unprinted documents. Frederick Burkhardt discusses the problem of publishing both the letters sent to Charles Darwin and the naturalist’s autograph marginal remarks on these communications in “Editing the Correspondence of Charles Darwin.” Here at least, the editors could easily justify printing the comparatively brief texts to which Darwin’s comments referred.

Editors faced with the more conventional form of marginalia, a subject’s comments inscribed on the pages of books or articles that he or she was reading, have a greater challenge. Samuel Taylor Coleridge left behind a body of marginalia in copies of his own works as well as in books written by others, and that body of material was so complex and independently significant that his editors treated the Marginalia separately in a special series of five volumes within The Collected Works.

c. Nonauthorial Emendations and Additions

In many cases, additions to a manuscript made by someone other than the author or the document’s recipient may be ignored. These can include dealers’ notations, symbols entered by archivists who have cataloged the materials, and notes by collectors through whose hands the manuscripts have passed. It may not be necessary to reproduce them verbatim in an authoritative edition, but many readers will need to be alerted to their existence by a summary or description in an editorial note. To some readers, these may seem no more an intellectual part of the document’s text than an owner’s signature on the flyleaf of a rare book or pamphlet. Editors must remember, however, that to bibliographers or historians of collecting, such owners’ signatures, dealers’ notations, and archivists’ symbols are important historical evidence. With unprinted documents and rare printed materials, such entries can be helpful in determining the item’s provenance, and they may need to be included in the description of the source text’s history.

Still other categories of nonauthorial inscription require more careful notice and may even warrant reproduction in the edition itself. Few members of an author’s family have resisted the temptation to edit literary memorabilia with pen, pencil, or scissors. Perhaps the worst offender in this category was Sophia Peabody Hawthorne, whose contributions to her husband’s posthumous literary image are immortalized in the defaced notebooks and other manuscripts that she prepared for publication in her years of widowhood. Mrs. Hawthorne’s activities as editor and censor had such a pervasive influence on her husband’s reputation that they could not be ignored by the editors of the Centenary Edition of his works. No physical description of the notebooks would be complete without reference to Sophia’s emendations and mutilation, and students of American cultural history would be ill served by an edition that did not report them. Thus, the textual notes to the clear texts of Hawthorne’s writings include detailed descriptions of Sophia’s handiwork, as well as notes recording similar revisions by the Hawthornes’ son, Julian.

It is the degree of historical significance of any nonauthorial additions—their independent documentary value—that determines how fully they need to be recorded in a scholarly edition or whether they should be mentioned at all. If a writer’s spouse marked the deceased’s letters and papers for a print edition that was never published, the nonauthorial emendations clearly have less importance than they would do if bowdlerized texts appeared and influenced a wide reading public.

Some examples of posthumous editing by friends, relatives, and publishers can be ignored because the resulting printed texts have not been as influential as Sophia Hawthorne’s. The editors of the George Washington Papers, for instance, ignore Jared Sparks’s “styling” of punctuation and spelling on the pages of the Washington manuscripts entrusted temporarily to his care while he prepared his selected edition of Washington’s writings. Had Sparks’s Washington volumes been the only ones available to scholars and laymen in the century and a half before the inauguration of the new George Washington project, an argument could have been made for recording his emendations in the new edition. Luckily, in those intervening decades, scholars had access to the original Washington manuscripts on which Sparks based his texts. The source texts were used for several other, and better,
editions of Washington’s letters and papers, and the public was not left at the mercy of Sparks’s version of the documentary record.

In some ways, the treatment of such nonauthorial revisions in manuscripts is comparable to a critical editor’s approach to the works of an author whose publisher demanded or imposed changes in a manuscript for the sake of literary style or public acceptance. In these cases, the modern editor must offer readers both the text that the author originally considered final and the revised version that the public actually read. When an author accepted such revisions, they bear directly on her or his sense of craftsmanship and aesthetic convictions. Even if they were imposed over the author’s objections, it was they, and not the original words, that were circulated to the world and became known as that author’s text. Just as the writer’s original intentions form one historically significant document, so does the revised and published version that became a part of literary history through its influence on those who read it.

3. Conflation

Just as a single manuscript can contain many versions of the same document or even the texts of distinct documents, two or more sources may be combined to produce a single editorial text. Few documentary editors will entirely escape the task of conflating, or combining, the elements of two or more sources into one reading text, although some of their methods of presenting the new texts may differ from those traditional to critical textual editing.

a. Fragmentary Source Texts

Conflation occurs most frequently when the best source text survives in fragmentary form, while less authoritative versions exist with a more complete text. It is no novelty to catalog a manuscript letter whose last page has been lost but for which a contemporary copy, later transcription, or even printed text will furnish the missing material. David Nordloh and Wayne Cutler argued the problem of conflating fragmentary sources in the Newsletter of the Association for Documentary Editing in 1980 and 1982. Nordloh questioned Cutler’s treatment of a letter from Andrew Jackson to James Polk for which two manuscript sources survived. The first was a draft in the hand of Jackson’s secretary, revised and signed by Jackson. The second was a copy of the letter made by Polk from the version he had actually received. Polk’s copy contained a postscript added when the fair copy was made from the draft. Three editorial choices were available. Nordloh argued for clear text, in which the postscript would have been printed as part of the letter, with the change in authority indicated in a back-of-book note. An inclusive- or expanded-text editor might have printed the postscript as part of the editorial text, noting the change in authority in a footnote. Cutler chose the most conservative solution to the problem, printing only the contents of the draft as the letter’s reading text, with the postscript transcribed verbatim in a note adjacent to the text. Nordloh defended his position with a discussion of the primacy of authorial final intentions. Cutler explained his own decision by analyzing the special reverece of documentary editors for their source texts.

In any documentary edition, conservative methods of conflation best suit the reader. Even if the conflated passages appear in one reading text, notes adjacent to the letter alert the reader to editorial intervention and provide easier and more convenient access to the necessary information. In clear text, without a superscript number to indicate that annotation is present, readers would be ignorant of the crucial textual and evidentiary problem at hand and the need to consult the back-of-book textual record.

When overlapping fragments of the text of the same document survive, and when each version can be considered reliable in terms of substantives if not in terms of accidentals, overt conflation of the sources into one editorial text may be preferable. The fact and location of such conflation can be indicated using numbered notes or other devices. Even here, documentary editors resist the temptation to impose a single pattern of accidentals on the resulting conflated text, although this can produce a text in which one three-paragraph section represents the author’s usage in a surviving eighteenth-century manuscript, while another three-paragraph passage shows the style imposed on the text by a late nineteenth-century transcriber.

If all the pages of a manuscript source text survive in mutilated form—as with documents damaged by fire or water or defaced by descendants or collectors—the editor may have to supply missing words or phrases at regular intervals throughout the editorial text rather than conflate the texts at a single point where one source ends and another continues. If this is a consistent problem in the edition, editors often devise a system of symbols that indicate such routine conflation. It would be needlessly intrusive to accomplish this purpose with dozens of footnotes indicating the source of words or phrases from the supplementary source text. The editors of Mark Twain’s Letters give their readers a chance to evaluate supplied material in mutilated manuscripts by providing photo facsimiles of such pages in their back-of-book textual records.

Frequent conflation may also be required when the author’s drafts were routinely copied for transmittal as letters or other communications by a scribe who was less than conscientious. Some editors solve this problem by adopting a special bracket to enclose words in final versions of letters and state papers supplied from the more
A remarkable instance of what might be termed “facsimile conflation” can be found in Dickinson W. Adams’s twentieth-century facsimile edition of *Jefferson’s Extracts from the Gospels*. Jefferson arranged two compilations of New Testament texts, one known as “The Philosophy of Jesus of Nazareth,” and the other as “The Life and Morals of Jesus of Nazareth.” While “The Life and Morals,” a compilation of Greek, Latin, French, and English versions of Gospel verses, survived and was eventually preserved at the Smithsonian Institution, “The Philosophy of Jesus” collection was lost. All that remained were the two mutilated copies of the King James Version of the Gospels at the University of Virginia from which Jefferson had clipped selections, and what appears to be a copy of the list Jefferson followed in removing these passages from the New Testaments. Working with photostats of intact copies of the same editions of the Gospels used by Jefferson, Adams created a new body of clipped (p.201) photocopies. Annotated, indexed versions of the facsimile compilations were published as the first volume of the second series of *The Papers of Thomas Jefferson, Jefferson’s Extracts from the Gospels*.

b. Reconciling Accounts of Independent Witnesses

In classical scholarship, the surviving witnesses to a lost archetype are usually in the form of scribal copies. Each must be collated with the others to isolate patterns of error that indicate transcriptional descent, to determine whether one or more was a copy made from an earlier and thus more reliable copy. Once this process of textual filiation is complete, variants among the witnesses are used to reconstruct the missing archetype so that the editorial text can represent the best readings provided by the imperfect witnesses.

For editors of modern documentary materials, the problem of reconciling discordant witnesses is most likely to appear when only verbatim, even shorthand, accounts survive of words communicated orally in the form of a speech, conversation, or interview. Editors who confront this textual challenge may wish to consult the descriptions of widely differing treatments of such records in the Woodrow Wilson Papers (24:viii-xiii) and the Douglass Speeches (1:lxv ff.). Remember, though, that these editorial procedures were adopted in the 1960s and 1970s. Today, computer-assisted collation of machine-readable transcriptions of such accounts makes their comparison and analysis far easier and more accurate.

In both the Wilson and the Douglass series, the longest surviving verbatim account was chosen as the basic text when the editors needed to conflate variant accounts. (If nothing else, the longest report was usually made by the reporter who stayed alert after rival scribes had lost interest.) Collating each variant version against this basic text often showed that in some cases one variant was based on another. The editors could establish patterns of textual filiation and ignore those reports that were obviously the scribal descendants of another stenographic version. From that point in the editorial process, however, the two groups of editors followed different courses.

Once the Wilson editors identified those verbatim accounts with claims to authenticity, they isolated and analyzed every crux, or unaccountable variant between the texts. Wilson himself helped his later editors, for he often reviewed transcriptions of the shorthand reports of his speeches prepared by his personal secretary and corrected his aide’s inaccurate reporting. The pattern of variants and cruxes determined the final editorial treatment of Wilson’s oral texts. Variants were often comparatively minor in length, and many cruxes were easily explained in terms of the mishearing of similar spoken words or the misreading of a reporter’s shorthand notes. Here the Wilson editors silently combined the accurate words and phrases from two or more reports into one text. Only when variants were substantial and cruxes inexplicable did they intervene with brackets or numbered footnotes.

While the editors of Wilson’s oral communications turned to the techniques of classical and literary scholarship to solve this problem, other documentary editors performed conflations more overtly. In the Douglass Speeches, for instance, the editors could not afford the luxury of conflation and emendation, even with the use of a full textual record. Shorthand-based newspaper reports of Douglass’s speeches contained variants that were not only cruxes (in orally transmitted documents, anomalies in reporting the same words) but also reflections of inconsistent reporting of the same passages in the speeches. Such inconsistencies resulted in one newspaper’s publishing long passages, which must have taken twenty or thirty minutes to deliver, while another paper ignored this section of an oration completely.

The Douglass editors could not conflate such variant texts as gracefully as editors whose cruxes were largely confined to minor, easily explained anomalies. Whenever it was necessary to add passages to the basic text from a version that had reported a section more fully, the conflated material was added to the basic text in angle brackets, and its source indicated immediately in a note. If the basic text contained a summary of the sentences or paragraphs reported more completely in the second text, a dagger in the editorial text leads the reader to a note where the basic text’s summary version is reported verbatim.

Most editions of nineteenth-century documents follow the patterns used by the Douglass edition. It is rare to find even one complete stenographic report for a speech in this era, and the modern editor is spared the process of collating variants in two independent witnesses of comparable length and detail. Most, like the editors of the papers
of Thaddeus Stevens, use the most complete and reliable report of a speech as their basic source text. Footnotes supply significant variants from other reports of the same speech, that is, different versions of text in the source text or passages omitted from the source text but included in other reports.

Once they have achieved fame, twentieth- and twenty-first-century figures usually provide their editors with fairly reliable source texts for prepared speeches. Press releases containing advance copies of speeches were routinely supplied to newspapers by World War II. The files of wire services and major newspapers usually hold these documents even if they are missing from the records of the woman or man who delivered the speech. Radio and television recordings of speeches as actually delivered often survive to complicate the matter, providing future editors with two versions of an author’s “final intentions”—one on paper, the other on tape or a disc.

The experience of Margaret Sanger’s editors is instructive. Their approach to the various records of her speeches is a lesson plan in practical ingenuity. Although some of Sanger’s orally presented remarks are recorded in their entirety with both images and words (see above, chap. 3), others fared less well. Often, the editors could rely on the typescript of a speech that Sanger had delivered, but in one notable instance, there was neither a recording nor a typed reading copy for Sanger’s radio address at Vassar College, 5 August 1926. What survived were an abbreviated version published in the Birth Control Review and excerpts published in several Poughkeepsie newspapers. No two versions were close enough to permit conflation, so the editors sensibly printed the longest version available (that in the Poughkeepsie Evening Star), supplementing this text with material from other newspaper accounts absent from the source text (Sanger Papers, 1:447–49).

III. Translations

Initial decisions on handling foreign-language, shorthand, and ciphered materials are discussed, above, in chapter 3. A few lucky editors enjoy access to the foreign-language translations actually used by their subjects to read letters or other documents written in an unfamiliar language or code. Here the appropriate source text is that contemporary translation, whatever its shortcomings as an accurate rendering of the original.

The George Washington Papers editors had this luxury for documents in French sent to Washington during the Confederation and his presidency. In early volumes of the edition’s Presidential Series, the editorial reading texts for such foreign-language materials were transcriptions of the translations read by Washington, with the French texts transcribed in the last footnote to the document. In later volumes of the Presidential Series, the French texts can be omitted, as they’ll be readily available in the electronic edition of the Washington Papers.

Most editors, however, will have to find a way to translate such sources.

A. Foreign-Language Materials

Realistically, few editors will be completely fluent in every language used to create the materials relevant to their editions. (Editors working on the Papers of Jacob Leisler for instance, dealt with documents in German, French, Dutch, and pidgin English-Dutch.) The preparation of these documents for publication will be a cooperative and iterative process among editors and translators.

The translator must be expert not simply in the language in question but in the language as spoken and written at the time the document was created. Spelling, punctuation, and other usages do not stand still in any tongue. Facility in twenty-first-century Spanish is inadequate preparation for dealing with seventeenth- or eighteenth-century Castilian.

Because of the collaborative nature of preparing translations for a documentary series, the translator who is not a native English speaker must also have exceptionally high English-language skills. The translator must be able to understand not only the editors’ questions voiced in modern English but also the appropriate patterns of English usage for the period in which the documents were created.

Translating into English for a documentary edition raises many issues. While some editors might be tempted to retain the punctuation, capitalization, or sentence structure of the foreign-language source in the translation, this may create a less-than-readable document in English. A more reasonable course is creating in the translation a document that parallels the original in its style and tone. A formal document can be translated into formal English; an informal one, into more colloquial phrases. If translations will be published side by side with documents of the same period inscribed in English, translators can use the English-language documents as a guide in translating nouns, forms of address, or other diction that appears in both languages. The translation cannot replace the original foreign-language document, but it can give the English-speaking reader the original’s substance and a sense of the language in which that meaning was conveyed.

Editors intending to publish translations must decide on how to treat irregular spellings of proper nouns—whether to give them spellings that are correct and consistent in English usage or to retain unusual spellings from the foreign-language document.

The choice may be easy for geographical place-names: it would, for example, be disconcerting to retain the Spanish “Londres” for “London” in an English translation of a document. Other proper nouns will raise more
difficult questions. Editors have made different choices in this matter, basing their decisions on the nature of their documents and their audience. Some use the correct English spellings of proper nouns in translation (when these can be determined), providing in a footnote variant spellings used in the original; others retain the author’s usage from the original, untranslated document. These and any other translation policies should be made clear in an edition’s introduction or in source notes to the specific translations. Readers can then evaluate the translation’s relationship to the original.

Some of the original’s texture can be conveyed in a translation with little effort. Formal elements such as underlining and block letters can easily be retained in a translation. But if the editor’s goal is a readable text in English, other features may have to be ignored or relegated to notes.

Authorial methods of rewriting or correcting the text create especially difficult problems. Interlineations or insertions may have to be recorded in footnotes when significant. An attempt to reproduce them in the translated text itself may defeat the purpose of providing a convenient, readable English version, for words or groups of words in one language seldom translate into exactly the same number of words or phrases in another tongue. One word in English may be the perfect equivalent of a several-word phrase in a foreign language, or it may take half a line of English words to convey the meaning of a single word in Greek.

It may be impossible to find an appropriate place to insert translations of interlineations or substitutions when they differ in length from the original. For example, the common French term *il y a* (literally, “it there has”) is rendered in English as “there is” or “there are.” If, however, a French author carelessly omitted the “y” in drafting the phrase and then inserted it into the text with a caret, there would be no way to show this insertion in the translated phrase. An editorial note could dispose of the matter clearly and concisely.

A final issue is documents that intermingle languages. Again, the goal should be a readable text for the intended audience. In *Family Letters of Victor and Meta Berger*, letters that Victor Berger wrote entirely in German are printed entirely in English translation, but when Berger (p.206) wrote in English and dropped in an occasional German phrase, the German words remain in the text, with a translation in a footnote. The same system is used by most editors whose subjects sprinkle their correspondence with mottoes and sayings in modern and classical languages. For more complex language problems in a collection of documents, it may be necessary to experiment with several of the most difficult documents. Only after trying out a number of different solutions can an editor settle on the conventions for translation that will produce the most usable texts for the audience likely to consult the edition. The same conventions should be kept in mind if the editor later finds it necessary to provide a translation in annotation for an unfamiliar foreign-language word or phrase appearing in a text.

As a convenience for the readers of most editions, translation should immediately follow the foreign-language original. The edition will “privileged” that source text while giving readers the text most will need to understand it.

**B. Shorthand**

For source texts in some standardized form of shorthand or code, the translator will probably find it more convenient to work from an image of the source text, not a transcription. Experts in the Gregg system are still fairly easy to find, but the Woodrow Wilson editors’ search for someone familiar with Wilson’s favorite Graham shorthand remains a scholarly legend. Once translations from Graham to English were completed, more challenges arose.

Most of the source texts that served the Woodrow Wilson Papers were printed in a very conservatively emended form. There was no question of publishing hand-set printed facsimiles of the notes and drafts that Wilson made in Graham shorthand. The translations, however, still demanded editorial attention. Since Wilson usually indicated his preferred marks of punctuation and paragraph breaks in his shorthand materials, these could be honored in the editorial texts. Still, the editors never pretended that they could guess all of the author’s intentions for translation from Graham, and their standard for the treatment of shorthand materials differs from that for materials drafted by Wilson in clear English. A statement of these special methods appears in editorial notes (1:19, 128–31). Notes to individual documents alert the reader to the use of translated shorthand source texts.

Other writers use personal “shorthand” that may now be (p.207) comprehensible only to their editors. It’s only these experts who have the opportunity to compare and analyze a wide selection of such idiosyncratic forms, while scholars without access to the project’s editorial archives are denied this luxury. This imposes a special duty on the editor to determine the meaning of such forms and to represent them verbally in the editorial text for the reader’s convenience and enlightenment. If the nature of the shorthand allows the expansion of alphabetic contractions within square brackets, the editor’s responsibility is fulfilled. If the symbols make such expansion impractical, the editor may need to use a special typeface to represent translation of the author’s do-it-yourself shorthand, with footnotes (or hyperlinks in an electronic edition) explaining the symbols in the source.

Whatever policy is adopted, the initial transcription and deciphered version should reflect the peculiarities of the original. The editor can decide on any necessary emendations at a later stage. The reader must be warned that all such emendations have been supplied by the editor, and in the substance of texts translated from shorthand, the reader deserves to know when editorial guesswork or imagination has been employed.
C. Codes and Ciphers

With ciphered materials, as with shorthand, photocopies of the source text are a more reliable basis for translation than a transcription of numbers and symbols. The modern editors of personal or diplomatic codes and ciphers from earlier centuries will find no ready-made experts to assist them. The editorial staff itself will have to learn to work from surviving keys to these ciphered materials to create an understandable reading text.

Editors with access to the key to a cipher or code have a concern beyond the choice of typeface for encoded passages: the author’s accuracy in enciphering the passages and the recipient’s skill in decoding them. The translated clear text of coded documents cannot stop with enabling the reader to see just which sections were entered in code and cipher and what those codes meant. The text or the accompanying notes must also record what the editor has been able to determine about the recipient’s success in mastering the ciphered passages. Indicating which words, phrases, or sentences were significant enough to deserve encoding allows the reader to see exactly which information in the letter was judged confidential and which facts the writer felt free to leave open to prying eyes. Noting both the author’s skill in encoding his or her own words (p.208) and the correspondent’s accuracy in using the key to the same code is critical in showing the effectiveness of the transmission of the ciphered information.

Madison’s editors found that a recipient’s errors in deciphering the coded text had led earlier, less conscientious editors to publish inaccurate versions of significant political correspondence (see Madison Papers, 6:177–79). John Jay’s editors discovered that the inventor of one code misused his own system so badly that his correspondent was unable to decipher his letters (Papers of John Jay, 2:117–18). The editor should indicate instances in which a significant difference exists between what the author intended and what the second party was, in fact, able to comprehend. The simplest solution is an editorial text that approximates the author’s intentions, no matter how badly fulfilled or how poorly the recipient or other readers managed to grasp the writer’s meaning. Numbered footnotes can describe discrepancies between intentions and perceptions. (A detailed analysis of a cipher document representing the full array of such problems of communication can be seen in the Madison Papers, 6:177–79.)

With ciphers, as with shorthand, it may be impossible to guess the author’s intentions as to capitalization and punctuation. Writers in cipher and code often deliberately omit marks of punctuation or paragraph breaks to avoid assisting the efforts of enemy cryptologists. Many editors have decided against supplying any of these omissions. Writers argue that since the coded message’s recipient had to guess at punctuation, it may be more accurate to print the newly decoded text in the same ambivalent fashion. Other editors punctuate the deciphered text based on known patterns of the author.

Whatever policy is adopted, the initial transcription and deciphered version should reflect the peculiarities of the original, and the editor can decide on any necessary emendations at a later stage. As with shorthand translations, the reader must be warned when and why such emendations have been supplied by the editor. For ciphers, the usual device is to print ciphered passages in small capital letters or some other typeface that does not appear elsewhere in the editorial texts of handwritten sources. This method eliminates the need for additional footnotes and ensures that the reader cannot confuse it with any other textual device.

D. Telegrams

The textual problems presented by telegraphic communications are intimately related to those of other forms of coded transmission and (p.209) translation. Editors of nineteenth-century archives that contain a substantial number of telegrams follow a conservative policy in emending received telegraph texts. In this era, the decoded message was written out by hand in the receiving telegraph office, and the words and phrases were usually copied in conventional form, using upper- and lowercase letters and marks of punctuation.

Editors of cables received in the modern era face new decisions. The twentieth century brought automated printing of decoded telegraph messages in uppercase characters only, with the additional convention of writing out the names of marks of punctuation (e.g., “stop”) instead of translating the words back into the symbols themselves. Editors of World War II leaders like Marshall and Eisenhower chose readability over documentary fidelity, translating “stop” to a period and supplying appropriate patterns of uppercase and lowercase letters. This decision was further justified by the fact that these generals customarily saw the incoming messages as summaries neatly retyped by their aides, not in their original form as telegrams.

IV. Current “Good Practice”

Again and again in our survey of editorial treatment of the problems discussed in this chapter, we’ve used the terms “cautious” and “conservative” to describe the appropriate approach. These are now the accepted watchwords among documentary editors for considering all textual matters. The reasons are ruthlessly practical.

As a practical matter, the chosen textual method should be the one that best serves the majority of the sources being edited. This spares readers unnecessary announcements of exceptions to some editorial rule or other. Beyond
correspondence, diary entries, and the like, were heavily emended in one of the most liberal applications of
transcribed literally. When the Woodrow Wilson editors reached volume 31, they provided a revised introduction
in favor of a single general policy described in volume 13 of the series: “With only a few exceptions all documents are
expanded transcription on record. As the project continued, this dual standard of textual treatment was abandoned
to explain to readers that they had come to follow an increasingly rigorous adherence to the rule of
transcription, became more deliberately conservative in
text in near diplomatic form. The Franklin Papers editors’ textual policies, while still far from literal or diplomatic
the Mark Twain project, but the experience of other editors is equally helpful. Among literary series, the first volume
of the Howells Letters was emended far more heavily than any other number in the series. In part, this is because
Howells himself standardized his letter-writing style as he approached middle age; the later source texts simply
needed less emendation. But the frequency of emendation also decreased as the editors themselves grew more
accustomed to Howells’s usage. Patterns of punctuation or spelling that appeared odd or ambiguous in the early
years of the project no longer seemed to need correction or explanation because they had become familiar to the
project’s staff.

Historical editions, too, are not without their noticeably modified editorial methods. From its inception, the
Ratification of the Constitution series gave conservative, almost diplomatic treatment to certain documents (such as
government records), labeled “LT” to indicate a literal transcription of the source. Other materials, such as private
correspondence, diary entries, and the like, were heavily emended in one of the most liberal applications of
expanded transcription on record. As the project continued, this dual standard of textual treatment was abandoned
in favor of a single general policy described in volume 13 of the series: “With only a few exceptions all documents are
transcribed literally.” When the Woodrow Wilson editors reached volume 31, they provided a revised introduction
to explain to readers that they had come to follow an increasingly rigorous adherence to the rule of verbatim et
literatim. The Henry Laurens Papers, once an example of expanded transcription, adopted policies that presented
text in near diplomatic form. The Franklin Papers editors’ textual policies, while still far from literal or diplomatic
transcription, became more deliberately conservative in the late 1980s, and the same pattern has been seen in the editions of the papers of Jefferson and the Adams family in more recent years.

From the beginning, one group of editors ignored the temptation to emend or “improve” their source texts—
those who dealt with the records of less-educated people. Historians and historian-editors have become sensitive to
the significance of the evidence available in the writings and records of the poorly educated. Political, social, and
economic historians now focus on the less educated as well as on the literate elite in our society, and the
documentary records of such groups and individuals are the subject of scholarly editing as well as general scholarly
interest.

It should be no surprise that the methods employed in editions such as the papers of black Abolitionists and
Southern freedmen differ from those designed to serve the correspondence of Adams and Jefferson. Editors of the
traditional statesmen’s papers are concerned with documents that fall within the realm of conventional scholarly
research. The methods of emendation and normalization used in the earliest historical papers projects were chosen
to illuminate the writings of individuals who were not merely literate but exceptionally well educated. As newer
editorial projects confronted the texts of documents that recorded the words and thoughts of the ill educated, even
the wholly illiterate, they discovered that these traditional conventions were unsatisfactory. The documents
demanded different methods and even different skills. Randall M. Miller pointed out in his instructive 1985 essay
“Documentary Editing and Black History” that the evaluation of historical documents generated by the African
American experience might require expertise in fields as diverse as cultural anthropology, folklore, linguistics, and
musicology.

The imposition of normalized punctuation, for instance, is based on the assumption that a source text's author
understands the functions of such marks and that he or she would approve such repunctuation if given the chance
to be his or her own editor. But correction of spelling errors in the writings of an ill-educated writer imposes a false
sense of authorial intentions. Worse still, it can destroy much of the special value inherent in documents inscribed
by the semiliterate—the phonetic rendering of colloquial language and dialect that make such documents useful to
philologists and cultural historians.

In many instances, the records of the less educated suffered nonauthorial normalization long before they became
the subjects of scholarly editing. Illiterate individuals had no choice but to dictate letters or memoirs to
second parties, who imposed their own notions of correct spelling and punctuation and even syntax upon them. The
only way to emend such a dictated source text to bring it closer to authorial intent would be to make it less correct

— G. Thomas Tanselle’s 1978 evaluation of the methodology of documentary editions prompted some to
reexamine their methods, but most editors who revised their methods did so on the basis of their day-to-day
experience. This underscores the point that no statement of principles is likely to cover all the issues and problems
to be encountered in any documentary edition and that the fledgling editor is well advised to examine in detail the
practices of a wide range of editions before launching a new documentary project. In particular, the lessons of
editors who modified their method in midstream are instructive.

The best-known example of evolving editorial methods among editions of famous American authors is probably
this, American editors now generally agree that conservative emendation is more effective than liberal emendation.
A survey of veteran editors in the early 1990s revealed that nearly all had adopted less intrusive editorial policies
than the ones announced in their original statements of editorial methods. While several scholars who had become
directors of editions initiated by earlier editors had revised textual policies in the direction of more conservative
emendation, no successor editor had instituted a new policy of more generous editorial intervention. It is no longer
possible to find editors who endorse without question the practices sanctioning liberal editorial intervention
commonly accepted a decade or two ago.

(p.210)
syntactically and to introduce phonetic misspellings to match the author’s dialect. Luckily, no editor has been tempted to follow such a course. Documents dictated by the illiterate, like documents inscribed laboriously if incorrectly by the semiliterate, must be allowed to stand, even though they may reflect a degree of elegance superimposed by the amanuensis and completely foreign to the authors themselves. This phenomenon is addressed squarely by editors of the WPA “Slave Narratives,” in which patronizing white interviewers affected both the form and content of the reminiscences they elicited from former slaves (Documentary Editing 18 [1996]: 92).

In earlier decades, there was a flurry of debate over whether the writings of certain categories of professional authors or historical figures demanded special treatment in terms of textual methods. For a time, this centered on the desirability of special requirements for editing the writings of women. It has now been agreed that the author’s sex in and of itself dictates no special methods in establishing a text. It is the subject’s level of education or habits and patterns of writing and the needs and expectations of the edition’s audience that should mold such methods.

V. Conclusion

Reconstructions of vanished early versions of historic documents or sources that demand translation remain exceptions to the general rule of conservative treatment for documentary sources. Still, the important elements of such sources are more likely to survive the flexible application of a conservative editorial approach than more liberal editorial intervention. The reader with a taste for watchwords may be reminded of A. E. Housman’s comments on the “science” and the “art” of textual criticism:

A textual critic engaged upon his business is not at all like Newton investigating the motions of the planets: he is much more like a dog hunting for fleas. If a dog hunted for fleas on mathematical principles, basing his researches on statistics of area and population, he would never catch a flea except by accident. They require to be treated as individuals; and every problem which presents itself to the textual critic must be regarded as possibly unique. . . . If a dog is to hunt for fleas successfully he must be quick and he must be sensitive. It is no good for a rhinoceros to hunt for fleas: he does not know where they are, and could not catch them if he did. (“The Application of Thought to Textual Criticism,” 132–33)

Documentary editors, too, must be knowing and sensitive flea hounds. The fact that their editorial products will be used as documentary evidence imposes a special responsibility. Their imaginations should be directed toward reconstructing inscribed truth, not distracting their readers with uninformed guesses. Liberal policies of editorial emendation and intervention represent the “rhinoceros” approach to documentary editing, for they miss the fleas and crush the source texts under their own weight.

Suggested Readings

For a discussion of the nature of an established text see Fredson Bowers, “Established Texts and Definitive Editions.” Anecdotal but useful accounts of the lessons to be learned in establishing texts are found in Ronald Gottesman and David Nordloh, “The Quest for Perfection: Or Surprises in the Consummation of Their Wedding Journey.” As for the textual record required in an edition of private writings, see Nordloh’s “Substantives and Accidentals vs. New Evidence: Another Strike in the Game of Distinctions.” G. Thomas Tanselle addresses the problems of editorial records for literary works in “Some Principles for Editorial Apparatus,” but there are no comparable studies of the special problems arising in creating such an apparatus for documentary sources.

For cautionary words on an edition’s failure to recognize the limitations of a microfilm as a source text without verification of a transcription against the original, see Jo Zuppan’s review of The Correspondence of John Bartram.


For further discussion of Jerome McGann and the socialization of texts, see Jack Stillinger, Multiple Authorship and the Myth of Solitary Genius.

Sargent Bush discusses the array of “multiple-text” problems in Jonathan Edwards’s “Miscellanies,” the thirty-five-year record of Edward’s reflections on theology and philosophy, in “Watching Jonathan Edwards Think.” The results of editorial work on these texts can be seen in vols. 13, 18, 20, and 23 of The Works of Jonathan Edwards.

Edward A. Levenston’s discussions of foreign-language translation for scholarly editions in The Stuff of Literature: Physical Aspects of Texts and Their Relation to Literary Meaning provide much food for thought on both a theoretical and practical level. A detailed analysis of a cipher document representing the full array of such problems of communication can be seen in the Madison Papers, 6:177–79.
For still more arguments in favor of conservative methods in emending source texts with documentary elements, see Hershel Parker’s review of two Hawthorne volumes in *Nineteenth-Century Fiction*; Ernest W. Sullivan, “The Problem of Text in Familiar Letters”; and G. Thomas Tanselle’s “The Editing of Historical Documents,” as well as his remarks on the overrated virtues of readability in his essay “Literary Editing.” Examples of clear reading texts without the use of textual symbols can be found in such editions as the Howells Letters and the Franklin Papers.

Excellent discussions of the many facets of the documentary records of the ill educated can be found in John W. Blassingame’s introduction to *Slave Testimony: Two Centuries of Letters, Interviews, and Autobiographies*; and in C. Vann Woodward’s review essay “History from Slave Sources,” in the *American Historical Review*.

Finally, the examples of textual method and practice offered in Philip Gaskell, *From Writer to Reader: Studies in Editorial Method*, still remain among the most helpful available.
Chapter Three

General Principles of Transcription and Proofreading

Introduction to Transcription

3.1 Transcription is the process of converting textual and nontextual elements of original documents into readable, publishable, typescript form. In so doing, editors strive to represent original documents faithfully. All transcription, however, is a form of translation and requires editors to make innumerable decisions about how to present documents. Editors make choices about standardization of the form of the documents (placement of datelines, uniform indentation of paragraphs, etc.) as well as about how to emend the text (capitalization, punctuation, etc.). In the past, the typesetting of textual features such as superscripts, subscripts, canceled passages, interlineations, marginalia, drawings, and other marks was costly, and these features were reproduced for only the most important texts or were represented through editorial symbols. Modern typesetting and printing techniques have reduced the difficulty and expense of reproducing unusual textual characteristics, and, as a result, cost is no longer the primary consideration. Instead, editors evaluate the types of documents they will be transcribing, consider the needs of the audience, and then select the form or forms of presentation that best convey the information contained in the document. Transcription methods should be presented in an introductory statement and then consistently implemented.
Principles of Emendation

3.2 Emendations are textual changes made by editors when transcribing original documents. The most common forms of emendation include capitalizing proper nouns and first letters of sentences and supplying terminal punctuation, but editors might also supply or delete other types of punctuation, convert archaic forms into modern usage, spell out abbreviations, standardize accidentals, supply letters or words missing from damaged manuscripts, and correct spelling. Editors should carefully consider whether to alter textual characteristics and, if so, which ones. Most important, they must consider how those changes may affect the information contained in the documents.

3.3 Silent emendation modifies the text of a document during transcription without alerting the reader at the point that the text has been altered. Editors may inform their readers in the edition's introduction that certain standard emendations will occur throughout the volume, or they may list textual changes at the back of the edition. If editors employ silent emendation, it is incumbent on them to explain how the texts presented in their editions deviate from the original documents. Exceptions to these general emendation rules are usually indicated in editorial notes.

3.4 Overt emendation is the process of modifying the text of a document during transcription in a way that alerts the reader that a change has been made. Editors may use symbols such as arrows, lines, brackets, italics, small capital letters, or footnotes to indicate where and how a text has been altered.

Methods of Transcription

3.5 There is no single agreed-on method of transcription. Rather, editors use different types of editorial apparatus and different degrees of emendation depending on the nature of the documents and the intended audience. Editors choose among five major forms of document presentation (photographic facsimile, typographical facsimile, diplomatic transcription, expanded transcription, or clear text transcription) to find a style that best suits the needs of their audiences, the purposes of their editions, and their personal preferences. Expanded transcription is used most frequently by editors of historical editions, but all editors must evaluate the advantages and disadvantages each form of presentation might bring to their editions.

3.6 Photographic facsimile editions reproduce original documents and present them on the page, screen, or microform frame free from alterations caused by converting manuscripts or original print into modern type. For typed or clearly written documents, photographic facsimile offers an effective method of presentation that is quicker to execute and less expensive than transcription. Hard-to-read
Editors when transcribing documents are concerned with capitalization, punctuation, abbreviations, standardization of manuscript forms, whether to alter textual elements during transcription, and how the text of a document has been altered. Editors in their editions develop rules for textual emendation, such as the use of brackets, italics, small text, or other methods to indicate changes. Rather, editors may consider the needs of their audience. Expanded transcription suits the needs of their personal preferences, or may present alterations caused by typewritten or handwritten methods of transcription. Hard-to-read documents will remain hard to read as photographic facsimiles, although documents photographed for a book or microform edition may have better contrast than the originals, and electronic editions may allow readers to adjust the size, contrast, brightness, and resolution of the images. The editors of the book edition of the Johns Hopkins University Seminary of History and Politics: The Records of an American Educational Institution published their documents as photographic facsimiles:

The problem of transcribing and transcribing in the state of Pennsylvania by the historical method. It is one of several states which are being prepared on the same subject. The period under review from 1790 to 1844.

Fig. 3.6a Gettleman, Hopkins, 1:105

The editors of the microfilm edition of the Benjamin Harrison Papers published their documents as photographic facsimiles.

Fig. 3.6b LOC, Harrison, ser.1, reel 12
The editors of the *Emma Goldman Papers* published several documents on the project's World Wide Web home page as photographic facsimiles:

**The Emma Goldman Papers**

*Edited by* S. Tropel, in *Alas* (Feb. 1909) and *Kommune* (Oct. 1909)

_Copyright_ 1909 by The Emma Goldman Papers Project, Inc.

Editors who use typographical facsimile as a method of transcription transfer original documents into type so that the appearance and physical arrangement of words on the typescript page mimic that of the original. The *Documentary History of the Supreme Court of the United States* used typographical facsimile to put a page of original minutes _top_ into type _bottom_ but did not render the linebreaks of the original:

---

Fig. 3.6c Goldman, WWW

---

3.7 Proclamation opened.

The Art Genl moved for issue of the proclamation for the reasons of the state, and the Court agreed.

The Court proceeded to hear the evidence and arguments of the United States on the subject of the motion made by their counsel.
Present

William Cushing

James Wilson

John Blair

Esq

James Iredell

Associate

William Patterson

Justices

Proclamation is made and the court opened.

A rule to show cause why

The Attorney General move for a mandamus to be
directed to the Judge of the District of New York
to apprehend Henry Barré a deserter from the
corvette belonging to the Republic of France should not
issue.

The Court proceeded to hear Messrs. Lewis

and Murray on behalf of John Corbey

& others and Mr. Rawle Attor for the Pennsylvania District on behalf of

the United States on the subject of the petition of said Corbey & others & on

a motion made by their council.

Fig. 3.7a Marcus, Supreme Court, vol.1,1:387

Present

William Cushing

James Wilson

John Blair

James Iredell

William Patterson

Esq

Associate

Justices

Proclamation is made and the court opened.

A rule to show cause why

The Attorney General move for a mandamus to be directed to the Judge of the
District of New York to apprehend Henry Barré a deserter from the
corvette belonging to the Republic of France should not issue.

The Court proceeded to hear Messrs. Lewis & Murray on behalf of John Corbey
& others and Mr. Rawle Attor for the Pennsylvania District on behalf of
the United States on the subject of the petition of said Corbey & others & on
a motion made by their council.

Fig. 3.7b Marcus, Supreme Court, vol.1,1:386
3.8 Editors using diplomatic transcription transfer the text of a document into modern type but use symbols such as arrows, carets, and brackets to indicate the document's stylistic details and appearance, a system dubbed "barbed wire" by the author and critic Lewis Mumford. This method is often used for documents for which the reader may be interested in the composition process. The editors of the Papers of Henry Laurens began using diplomatic transcription in volume 10, marking insertions with up and down arrows (↑ inserted ↓) and canceled passages with strikethrough type (canceled):

[1st] Your Letter of the 25th November which inclosed [2nd]

Fig. 3.8a Hamer, Laurens, 10:524

The editor of the Complete Works of Washington Irving: Journals and Notebooks also employed diplomatic transcription, using arrows to mark insertions (↑ inserted ↓) and angled brackets (<canceled>) to mark canceled passages:

lasting impressions. A Frenchman is all life, gaiety & spirit - his feelings are acute and as easily wounded as delighted yet they do not retain long any sentiment either of anger or pleasure. Though his courage is no one is more sensible of an insult or more prompt to resent it - yet he is easily pacified the least concession restores him to good humour.

Fig. 3.8b Wright, Irving, 1:329

3.9 Expanded transcription encompasses a wide spectrum of editing styles, all of which standardize accidantals, datelines, and signatures; mark paragraphs with indentations; and do not attempt to reproduce the excessive spacing and physical layout of the text of documents. Editors who employ expanded transcription announce the ways they will emend the text and use those methods unless otherwise indicated. Editors have the greatest leeway for making decisions about the appearance of their text with the expanded

3.10 Editors who use a method of transcription. Consideration is given to the text as noted in Section 3.6 of the document into type as in the example History of Emancipation. transcription, maintaining the original wording, capitalization, and abbreviations. The editors provide extra space in the document to obscure ambiguous characters and indenting and printing in suitable

Hdqrs U.S.

165
Resp'y returned. The reports are false. My employees are discharged, and are entirely under the authority of the Statesman. The force is also false. They are not from my office requiring their livelihood. This order has been signed by Carney Sup't of Negroes in the Coun'ty of East Va., and approved by Major Gen'l & N.C.

For a more complete Communication of the matter, I am Coln. this Dept: forwarded White 1x Coln Corn

The editors of the Papers of Henry Laurens use expanded transcription to highlight significant emendation. Editors will make decisions about how to alter the text as noted in Section 3.6 of the document into type as in the example History of Emancipation. The editors also maintain the original wording, capitalization, and abbreviations. The editors provide extra space in the document to obscure ambiguous characters and indenting and printing in suitable
method of transcription. Choices range from a near-literal rendering to significant emendation. Editors who opt to use expanded transcription will face difficult decisions about how to best convey the substance of historical texts without distorting them or cluttering the page with editorial devices. (For more information on the execution of expanded transcription, see Chapter 5.)

3.10 Editors who use a more conservative style of expanded transcription alter the text as noted in Section 3.9 but otherwise try to convert the text of the document into type as literally as possible. The editors of Freedom: A Documentary History of Emancipation employed a conservative style of expanded transcription, maintaining the original manuscripts' syntax, spelling, punctuation, wording, capitalization, paragraph breaks, superscripts, contractions, and abbreviations. The editors limited intervention to indenting the beginning of new paragraphs, incorporating interlineations into the body of the text, providing extra space in the place of missing terminal punctuation, making ambiguous characters and punctuation conform with modern practice, and indenting and printing in small type extended quotations within documents:

Hdq" U.S. Forces. East" Shore Va. Eastville Feb7 15

/65/ Resp" returned. The Statements of B. L. Parish are entirely false. My employees are more loyal than the party making the charges, and are entirely worthy of the trust placed in them.

The Statements made concerning the hiring of colored persons by force is also false. These Statements are based upon an order issued from my office requiring all colored persons to labor for a livelihood.¹ This order in all its details has been approved by Major Carney Supt. of Negro affairs Dept: Va. by Brig: Gen Shepley Com" the Dist: of East" Va, by Major Gen¹ Ord Com² this Dept: and was approved by Major Gen¹ Butler when in Command of the Dept of Va & N.C.

For a more complete answer to within letter I would resp" refer to my Communication of the 6th inst: addressed to the Major Gen¹; Com² this Dept: forwarded through Hdq" Dist: East" Va.² Frank J. White Lt Col Com² U.S. Forces East" Shore Va.

Fig. 3.10a Berlin, Freedom, ser.1,2:220

The editors of the Papers of Ulysses S. Grant used a conservative form of expanded transcription that entailed lowering superscripts to the line, marking
canceled passages with strikethrough type, and reconstructing or noting torn, mutilated, illegible, and lost materials:

**Gen.**

I have just recieved yours of the 23d inst. informing me that I have given Lt. P. Andrews 2d Art’y credit for 100$ on my accounts for 2d quarter 1848. The name should read 1st Lieut. Geo. P. Andrews 2d Artillery.

I am Gen
Very Respectfully
Yr. Obt. Svt.
U. S. Grant
1st Lt. 4th Inf’y
A. A. C. S.

To Gen Geo. Gibson
Com.y Gen. Sub.

Fig. 3.10b Simon, Grant, 1:182

3.11 Editors who employ a more liberal style of expanded transcription standardize the text as noted in Section 3.9, but then make significant additional interventions. In the first volume of the Documentary History of the Ratification of the Constitution, the editors employed a liberal style of expanded transcription that heavily emended the text, as described in the volume’s editorial methods statement:

**Capitalization, Punctuation, and Italics in Manuscript Materials**

Capital letters are used to begin each sentence. Random capitals and italics are removed except when they are evidently used by the author for emphasis. Periods are placed at the ends of sentences instead of dashes, colons, or no punctuation at all. Punctuation is altered within sentences if needed to clarify meaning.

**Spelling**

With one exception, spelling is made to conform to present-day practice. For example, “labour” and “foederal” are spelled “labor” and “federal.” The exception to this rule is the spelling of names of individuals. While it is easy enough to correct the spelling of the names of a “Madison” or a “Washington,” there are hundreds of legislators and other men whose names are spelled in various ways in document after document, and sometimes in the same document.

3.12 Clear text editors attempted to produce clear transcriptions. Emendations sometimes a page of the back of those editors.
The editors therefore follow the practice of the editors of such modern publications as the papers of Thomas Jefferson, John Adams, and Benjamin Franklin, who print the names as they are spelled in each document.

**Abbreviations, Contractions, Superscripts, Numbers, Crossed-out Words, and Blank Spaces**

Abbreviations such as those for place names ("Phila." for Philadelphia, for example) and military titles are spelled out. Contractions such as "can't," "'tis," and "'altho" are retained. Superscripts are lowered to the line. Archaic forms such as "'yt" and "'ye" are spelled out, "'sc." is printed "etc.," and "'K" is printed "and." Numbers are printed as they appear in the documents. Crossed-out words in documents, if they are significant, are placed in editorial notes. Otherwise they are not reproduced. Spaces intentionally left blank in documents are indicated by an underline.

The style of expanded transcription employed by the editors of volume 1 of the *Documentary History of the Ratification of the Constitution* produced a text that replaced many of the abbreviations and unusual textual features of the manuscript with fully spelled-out abbreviations and modernized punctuation, capitalization, and spelling—as can be seen by comparing this volume's transcription of a letter (top) to the version provided by the Papers of James Madison (bottom), which used a more literal form of expanded transcription:

When the plan came before Congress for their sanction, a very serious effort was made by R[ichard] H[enry] Lee and Mr. [Nathan] Dane from Massachusetts to embarrass it.

When the plan came before Congs. for their sanction, a very serious effort was made by R. H. Lee & Mr. Dane from Massats. to embarrass it.

**Clear text editions use silent emendation and limited editorial apparatus to produce clean typescript transcriptions of documents.** Clear text editions offer a page of text that is free of editorial apparatus or commentary. Emendations and notes to clear text editions, if provided, are usually placed at the back of those editions, with page and line references to the text. Clear text
transcription is most often used for editions of literary works so that they might be easily read. The editors of Henry D. Thoreau: Journal presented Thoreau's journals in clear text, with all annotation and textual emendation noted in the back of the edition:

142 JUNE 24, 1840

...dust which day has raised— A column of smoke is rising from the woods yonder to uphold heaven's roof till the light comes again. The landscape, by its patient resting there, teaches me that all good remains with him that waiteth—and that I shall sooner overtake the dawn by remaining here, than by hurrying over the hills of the west.

"Morning and evening are as like as brother and sister. The sparrow and thrush sing, and the frogs peep—for both.

"The woods breathe louder and louder behind me—with what hurry-scurry night takes place! The wagon rattling over yonder bridge is the messenger which Fig. 3.12a Broderick, Thoreau, 1:142

662 TABLE OF EMENDATIONS

142.5 waiteth | possibly waited
142.8 and evening | an evening in MS
142.12 hurry-scurry | possibly hurry skurry

Fig. 3.12b Broderick, Thoreau, 1:662

3.13 An excerpt from Thomas Jefferson's draft of the Declaration of Independence is rendered below in each of the five forms of documentary presentation. This photographic facsimile is from the Works of Thomas Jefferson:

When in the course of human events it becomes necessary for a people to dissolve the political bonds which have connected them with another, and to assume among the powers of the earth, the station to which the laws of nature and of nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

Fig. 3.13a Ford, Jefferson, 2: facing 200
This typographical facsimile example was created for this handbook:

When in the course of human events it becomes necessary for a people to
advance from that subordination in which they have hitherto remained, & to as
separate and equal
same among the powers of the earth the equal and independant station to
which the laws of nature & of nature's god entitle them, a decent respect
to the opinions of mankind requires that they should declare the causes
the
which impel them to the change separation.

Fig. 3.13b

This diplomatic transcription example was also created for this handbook, whose authors marked insertions with arrows (↑ insertion ↓) and canceled words with strikethrough type (canceled):

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, & to assume among the powers of the earth the equal & independant station to which the laws of nature & of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

Fig. 3.13c

The editors of the Papers of Thomas Jefferson used expanded transcription to present the text as it read before it was edited and then noted the later deletions and insertions in the document's footnotes:

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, & to assume among the powers of the earth the equal & independant station to which the laws of nature & of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

Fig. 3.13d Boyd, Jefferson, 1:423
This clear text transcription is from Thomas Jefferson: Word for Word:

When, in the course of human events, it becomes necessary for a people to advance from that subordination in which they have hitherto remained and to assume among the powers of the earth the equal and independent station to which the laws of nature and of nature’s God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

Verifying the Text

3.14 Proofreading is the process of comparing a transcript of a document against the original text for accuracy of all textual details, such as correct wording, phrasing, spelling, punctuation, capitalization, paragraphing, and consistent emendation. Proofreading is an essential step in the editing process. Editors must ensure that all parts of the edition, including documents, notes, quotations, headings, and titles, have been accurately presented. The text must be rechecked for accuracy whenever errors might be introduced into the final text. For a list of proofreading symbols and an example of a marked-up proof, see the Chicago Manual of Style, 14th ed., figs. 3.1 and 3.2.

3.15 The most effective method of verifying a text employs a two-person team: a document holder who reads aloud from the original document or a clear photocopy, noting every capital letter, punctuation mark, or irregular spelling; and a reader who follows along, inspecting the typescript copy of the document. When a difference is discovered, the original document and the typescript version are compared to determine the correct form, and any changes are marked onto the typescript draft. After the entire document has been read, the typescript version is corrected. If time and resources allow, this process is repeated, with the two people exchanging roles or with two different people. This method may also be employed by a single editor who reads the original documents into a tape recorder and then listens to the tape while proofreading the typescript text. Proofreaders should be equipped with a style sheet explaining the rules of presentation and emendation that will be used in the volume.
3.16 Computers currently play a limited role in verifying historical texts because of their inability to accurately compare an original document and a typescript text. Spelling-check and grammar-check programs may detect flagrant misspellings or improper grammatical usage in modern documents, but they cannot verify that a typescript text accurately represents the spelling, wording, and punctuation of an original document, especially if the document was written by an author using nonstandard English. In addition, spelling checkers do not catch incorrectly spelled words that happen by chance to spell another word (e.g. “two” transcribed as “too”). Double-keying, a process whereby two different typists transcribe the same documents and a computer compares the typescript versions for dissimilarities, can be used to verify texts, but it is extremely labor-intensive and moreover does not guarantee an accurate text, which depends on the skill of the two transcribers and their familiarity with the original texts. As editors develop ways to use computers to check the accuracy of their texts, these new techniques must be judged by the standards used to evaluate other proofreading methods: their capacity to accurately detect inconsistencies between original documents and typescript texts.
Chapter Four

Transcription: Types of Sources

Sources

4.1 Different types of sources present editors with different transcription problems. Before establishing a transcription policy, editors should evaluate the texts to be published and consider which methods will accommodate differences in form, function, writing style, and content. Editors should strive to follow a consistent method of transcription from the initiation to the conclusion of a project, so any policy decisions should reflect the full range of sources that will be transcribed over the life of the project.

4.2 Editors seldom publish multiple versions of a document. Instead, they select as their source text the fullest or most reliable version or the one most consistent with the aims and principles of the edition. Editors choose between early and final drafts, published and unpublished versions, and retained copies and recipients’ copies of letters. Editors establish criteria for selecting the source text and then meticulously follow these guidelines. The editors of the Papers of Andrew Jackson provided readers with a ranking of their preferred genre for source texts:

In cases where several copies of a document have been available, the recipient’s copy has been preferred. Second choice has been a letterbook copy; third, a retained draft. Generally, printed copies have been used only in the absence of manuscript versions, but in cases where the document was intended for publication the contemporary newspaper copy has been used.

Fig. 4.2a Smith, Jackson, 1.xxxii
The editors of the *Papers of William Livingston* offered their system of prioritizing variant drafts from different sources:

**SELECTION AMONG MULTIPLE VERSIONS**

When multiple copies of a manuscript exist, the following priority system determines which version is to be published: (1) autograph letters or other documents, (2) broadsides, (3) contemporary newspapers, (4) drafts, (5) letterbook copies, (6) later printed copies. Significant deviations among the versions are noted, but minor variations are not.

Fig. 4.2b Prince, Livingston, 1:xix

If significant variations between drafts of a text exist, they may be noted in footnotes, as was done by the editors of the *Papers of Chief John Ross*:

1 In the draft copy of this letter are extra lines not in the autograph copy. These lines read:

   for example—when the white People first landed at these shores and placed their feet upon this Continent—they were but few in numbers and were consequently weak in physical strength but in knowledge they were strong for their minds were cultivated—and the red people were then numerous and strong, but in knowledge they were comparatively weak for their minds were uncultivated. The whites readily perceived this, and by artifice soon courted the friendship of the Indians and secured their hospitality & were kindly treated. For their own preservation the whites contrived & succeeded in creating quarrels and wars between neighbouring tribes, while they themselves sought refuge under the protection of their neutrality, until their numbers in process of time by emigrations from Europe became numerous & strong. They then were bold enough in conjunction with

Fig. 4.2c Moulton, Ross, 1:287

**Handwritten Sources**

4.3 Documents written by hand pose a number of challenges for transcribers. Individual authors introduce peculiarities into their manuscript documents, influenced by circumstances such as background, education, health, emotional state, and the physical conditions under which the document was written. Authors produce documents that contain indecipherable handwriting, nonstandard spelling, faulty grammar, obscure abbreviations, and inconsistent punctuation. Other peculiarities can be attributed to the writing conventions of particular eras. Documents written before the standardization of American English in the mid-nineteenth century contain irregular spelling, archaic usages, and nonstandard punctuation, while handwritten documents from the sixteenth century through the early twentieth century frequently used superscripts, subscripts, contractions, and abbreviations. Editors must clearly state the editorial conventions they use.

4.4 Editors transcribe the text of the manuscript as accurately as possible. They are unfamiliar to most readers. The editors retained many of the features that would cause confusion:

5. The text of the manuscript is transcribed:

a. Spelling mistakes are not marked with an overline, but are corrected through misspelling changes.

b. Capitalization of names of people and "w" is often altered. When readings are different in a sentence, the editor notes them.

c. Punctuation is added, when a sentence is not complete.

d. Words in parentheses and brackets are transcribed with the symbols.

e. Words in italics are transcribed with italicized text.

f. Slips of the tongue or persand are retained.

g. Contractions are retained, so "th" is expanded as "the," and "y" is preserved.

h. The use of "u" is expanded as "u" is also used to represent "y." Likewise, "w" is considered as "w." In the text of the manuscript, indicated by the editor in brackets, the letter is presented as a capital letter.
Editors transcribing handwritten documents from the seventeenth century encounter forms of capitalization, spelling, abbreviation, and usage that are unfamiliar to most modern readers. The editors of the *Papers of William Penn* retained many of these now-unusual features, changing only those that would cause confusion:

5. The text of each document is rendered as follows:
   a. Spelling is retained as written. Misspelled words are not marked with an editorial [sic]. If the sense of a word is obscured through misspelling, its meaning is clarified in a footnote.
   b. Capitalization is retained as written. In seventeenth-century manuscripts, the capitalization of such letters as “c,” “k,” “p,” “s,” and “w” is often a matter of judgment, and we cannot claim that our readings are definitive. Whenever it is clear to us that the initial letter in a sentence has not been capitalized, it is left in lower case.
   c. Punctuation and paragraphing are retained as written. When a sentence is not closed with a period, we have inserted an extra space.
   d. Words or phrases inserted into the text are placed within braces.
   e. Words or phrases deleted from the text are crossed through.
   f. Slips of the pen are retained as written, and are not marked by [sic].
   g. Contractions, abbreviations, superscript letters, and ampersands are retained as written. When a contraction is marked by a tilde, it is expanded.
   h. The thorn is rendered as “th,” and superscript contractions attached to the thorn are brought down to the line and expanded: as “the,” “them,” or “that.” Our justification for this procedure is that we no longer have a thorn, and modern readers mistake it for “y.” Likewise, since modern readers do not recognize that “u” and “v” were used interchangeably in the seventeenth century, we have rendered “u” as “v,” or “y” as “u,” whenever appropriate.
   i. The £ sign in superscript is rendered as “l.”
   j. The tailed “p” is expanded into “per,” “pro,” or “pre,” as indicated by the rest of the word.
   k. The long “s” is presented as a short “s.” The double “ff” is presented as a capital “F.”
In contrast to the fairly literal style employed by the editors of the *Papers of William Penn*, the editor of the *Correspondence of Roger Williams* modernized and standardized many of the peculiar features of seventeenth-century prose:

**Spelling**

Spelling is preserved as found in the manuscripts or printed sources, except for regularization, in accordance with modern practice, of the interchangeable letters u and v; i and j; w and w; and, less frequently, for initial letters i and y. Obvious slips of the pen or false starts are silently corrected.

**Capitalization**

Capitalization is retained as found in the original manuscripts, except in the following instances. Personal names and geographical names have been capitalized. When it is not certain whether Williams or other writers intended to use a capital or lowercase letter, modern usage is followed.

**Punctuation**

Punctuation is retained as found in the manuscripts and printed sources, except for the following conventionalizations. Williams's ubiquitous colons, when used as terminal punctuation, have been replaced by periods; colons used for medial punctuation, however, have been retained. In cases where it has been difficult to determine whether Williams intended a colon to serve as medial or terminal punctuation, the colon has been retained. Virgules have been replaced by commas or periods. Williams frequently used parentheses but commonly forgot to close parenthetical statements with a final mark. Because his style precludes most attempts to determine where a closing parenthesis should be placed, the opening parenthesis has been treated in such cases as a slip of the pen, and thus has been silently omitted. Sometimes Williams used parentheses within parentheses to such a degree as to cause complete confusion in his syntax; in those instances, the confusion has been silently remedied by replacing his interior parentheses with commas or dashes. Square brackets, used by Williams to set off quoted matter, have been replaced by modern quotation marks. Every sentence ends with a period unless it is not clear where a sentence ends, in which case the original punctuation—or lack of it—is retained. Bracketed punctuation is occasionally emended for clarification.

**Abbreviations, Constructions, and Word-Origins**

Abbreviations and constructions of manuscripts, but those that are not expanded in square brackets, are translated as follows: “yo” expanded silently to “our”; “et cetera” (etc) is converted to or not found in the original. Phrases, such as "vils & mo" printed as th. The tild is represented. The tailed Τ is expanded as it represents (i.e., per, pre, p) as monetary signs, are printed as an abbreviation, given the letter Θ. Williams's symbol for partially served as an abbreviation.

**Other Textual Procedures**

Superscript letters and integers are omitted without editorial comment. A canceled matter is deemed stored in the text and explained. Obsolete words and usage of a definition (usually a longer definition or explanation obsolete words or archaic. *Dictionary* have not been given.

4.5 Eighteenth-century documents issues as those dating from the eighteenth, obscure abbreviations, such as vanished symbols. The editor
ne editors of the Papers of Roger Williams modernized seventeenth-century prose:

scripts or printed xe with modern prac-
and j; v and w; and, us slips of the pen or

nal manuscripts, ex-
es and geographical cer-
unt whether Will-
or lowercase letter,

scripts and printed izations. Williams's s-
tuation, however, dif-
cult to determine me-
or terminal rules have been re-
y used parentheses ements with a final npts to determine the opening paren-
the pen, and thus used parentheses complete confusion has been silently s with commas or off quoted matter, s. Every sentence sentence ends, in of it—is retained. or clarification.

Abbreviations, Contractions, and Symbols
Abbreviations and contractions are preserved as found in the manuscripts, but those that are not recognizable to modern readers are expanded in square brackets after the first appearance in the text of each letter (e.g., Matie [Majesty]). Some abbreviations, however, are treated as follows: "yo" is expanded silently to "you"; "o" is expanded silently to "our"; "S" is expanded silently to "Sir." Ampersands are converted to read "and," though the abbreviation for et cetera (&c) is converted to read "etc." with a period added even if not found in the original. However, ampersands used in foreign phrases, such as "viis & modis," are retained. The thorn is always printed as th. The tilde is replaced by the letter or letters it represents. The tailed ñ is expanded silently and replaced by the letters it represents (i.e., per, pre, pro, and less often, pur). Symbols, such as monetary signs, are printed as closely as possible to the original representation, given the limitations of modern typography. Thus, Williams's symbol for pounds (money and weight), which essentially served as an abbreviation for livres, is printed 80 li.

* * * *

Other Textual Procedures
Superscript letters and interlineations have been brought down to the line. Words and phrases struck out by the writer are generally omitted without editorial comment. In some cases, however, when canceled matter is deemed important, deleted words have been restored in the text and explained in a note.

Obsolete words and usages are clarified for the reader by means of a definition (usually a synonym) placed in square brackets immediately following the obscure word or words, or when necessary a longer definition or explanation is supplied in a note. Definitions of obsolete words or archaic terms taken from the Oxford English Dictionary have not been given a citation.

Fig. 4.4b LaFantasie, Williams, 1:xxi–xxxii

4.5 Eighteenth-century documents raise many of the same transcription issues as those dating from the seventeenth century, including irregular spelling, obscure abbreviations, superscripts, the tilde, the thorn, and other now-vanished symbols. The editors of the Papers of General Nathanael Greene
outlined the basic issues presented by eighteenth-century handwriting and their methods of addressing them:

**Spelling**

Spelling is retained as written. If a misspelled word or name is not readily recognizable, the correct spelling follows in brackets. Names are correctly spelled in notes and index.

Inadvertent repetitions of words are corrected silently.

**Capitalization**

The author’s capitalization is followed, including the eighteenth-century practice of capitalizing words within sentences, except where necessary to conform to the following rules:

1. All sentences begin with initial capitals.
2. Personal names and titles used with them, honorifics (such as “His Excellency”), geographical names, and days of the week and months are capitalized.

**Abbreviations and Contractions**

1. Shortened word forms still in use or those that can easily be understood (as “t’was” or “twixt”) are rendered as written.
2. Those no longer readily understood are treated thus: “cmsgy [commission]” or “warr[an]t.”
3. Abbreviations of names or places—forms known only to the correspondents or their contemporaries—are also expanded in brackets, as in S[amuel] A[dams] or Chsn [Charleston].

**Symbols Representing Letters and Words**

When any of the following symbols are expanded they are done so silently:

1. The thorn, which by 1750 had been debased to “y” as in “ye,” is expanded to “th.” Such abbreviations as “ye,” “y’t,” or “y’m” are rendered as “the,” “that,” or “them.”
2. The macron is replaced by the letter(s) it represents—as in commission, which becomes commission, or happen, which becomes happen.
3. The ſ sign is expanded to the appropriate letters it represents (e.g., per, pre, or pro).

Where necessary, proofreading rules:

1. A period or question mark is placed after every sentence.
2. Dashes used in place of punctuation marks are replaced with periods, when used to mark a statement or thought.
3. No punctuation in the text.

The editors of the Papers have rendered these slips of the pen, usual in manuscript, in scrupulous fidelity, but the editors believe the letters have been brought to print with words with a tilde (~) to avoid confusion. Inconsequential errors in spelling, punctuation, and capital letters are corrected when necessary.

The editors have sought to render, beyond these exceptions, the manuscript with greater freedom regarding misspellings, capitalizations, and other differences persisting in the autograph manuscripts.
h-century handwriting and

d word or name is not
ws in brackets. Names

d silently.

uding the eighteenth-
enties, except where

m, honorifics (such as
days of the week and

ns
ose that can easily be
as written.
ated thus: “cmys [com-

s known only to the
so expanded in brack-

Words
nded they are done so
ed to “y” as in “ye,” is
“ym,” or “yn” are ren-
resents—as in com-
ich becomes happen.
te letters it represents

Punctuation

Where necessary, punctuation is changed to conform to the follow-
ing rules:
1. A period or question mark is placed at the end of every sentence.
2. Dashes used in place of commas, semicolons, periods, or question
marks are replaced with appropriate punctuation; dashes are retained
when used to mark a suspension of the sense or to set off a change of
thought.
3. No punctuation is used after a salutation.

The editors of the Papers of James Madison noted many of these same stylistic
forms and described their edition's method of transcription:

The editors have sought to follow the original texts of documents with
scrupulous fidelity, but several exceptions should be noted. Superior
letters have been brought down, thorns (“ye,” “y,” “yn,”) as well as
words with a tilde (“comissrs” and “comittee.”) have been expanded,
and confusing punctuation has been eliminated. The first letter of a sen-
tence is invariably capitalized, regardless of the writer's eccentricities.
Inconsequential decoding errors made by Madison, Jefferson, or Mon-
roe in their deciphered correspondence have been ignored, but substantial
mistakes have been corrected (with brackets) or annotated. Obvious
slips of the pen, usually repeated words, are silently corrected. Long
dashes or gaps have been interpreted to indicate a new paragraph. Be-
yond these exceptions, alterations are noted in textual footnotes.

4.6 Nineteenth-century writers discontinued many of the habits of
seventeenth- and eighteenth-century writing, particularly the use of the thorn
and the tilde, resulting in a style that is more familiar to modern readers. But
other differences persisted, such as greater use of superscripts and abbrevi-
ations, multiple ways of ending sentences and indicating paragraphs, and
greater freedom regarding capitalization. The editors of the Letters of Henry
Adams addressed the nineteenth-century stylistic conventions of their letters:

Textual editing has been relatively simple since there are but few in-
stances where we were not able to work from the surviving letters. Between
the autograph manuscript and any previously published version, the auto-
graph text has the authority of being the author’s intention. We have adhered to that text as closely as common sense allows. We have kept Adams’ misspellings and special usages. Readers may note his difficulty with *Tennessee* and *McLeish* and his omission of the first *a* from the name of his friend MacVeagh. They may also observe that he follows a common nineteenth-century practice and often puts apostrophes in *you’re* and *it’s* (as his father did before him).

Inscription errors (for example, *the* for *that*), which are by definition not what the author intended, are categorically different from misspellings. Such mechanical errors have been silently corrected unless they are possible psychological slips. In those cases we preserve the manuscript text or, when correction is called for in order to avoid confusion of meaning, we note the change. Corrected errors in dating, since they might make a letter hard for someone to find in the microfilm edition of Adams’ letters, have been noted. As to inscription changes which Adams himself made, we do not note them. He wrote fluently and made relatively few false starts or running corrections. A detailed record of changes, Adams’ own and the editors’, has been kept for the use of future scholars.

Simple procedures have sufficed for the making of copy text. Thanks to the clarity of Adams’ handwriting, the number of *cruxes* is remarkably small. The original letters are in a state of excellent preservation, and the rare cases of defective manuscript have been noted. Where Adams’ autograph contractions abridge words and add superscript letters, we preserve those which are immediately readable when superscript letters are brought down to the line (*Dr Mr* &c and leave-taking phrases like *Ye obt serv* or *Aflecty yrs*); where such transcription might not be immediately readable and a standard abbreviation exists, we substitute the latter (*Dec. for Dec’*); where Adams points his superscript contraction (*2* or *ob* &c) and modern usage does not, we again follow usage (*2d* or *obt*); where there is no equivalent, we expand autograph contractions (*which for wh*). We have not followed Adams’ inconsistent practice, but have normally put commas and periods within quotation marks.

Fig. 4.6a Levenson, Adams, 1.xlii-xliii

The editors of *Freedom: A Documentary History of Emancipation* outlined the basic stylistic issues and how they handled them:

THE RENDITION of nineteenth-century manuscripts into print proceeds at best along a tortuous path. Transcribing handwritten documents into a standardized, more accessible form inevitably sacrifices some of their evocative power. The scrawl penciled by a hard-pressed army commander, the letters painstakingly formed by an ex-slave new to the alphabet, and the practiced script of a professional clerk all reduce to the same uncompromising print. At the same time, simply reading, much less transcribing, idiosyncratic handwriting poses enormous difficulties. The records left by barely literate writers offer special problems, although they are created by better-educated military officers, or others.

The editors have a conviction that readers desire invention and, indeed, are frequently posed by conversion of a document. The principal intervention, one by near illiterates, is to survey the struggling intensely felt emotion following the document. For the modern reader, these documents may appear undiscerning until their understandable when such documents offer. Other writers, present the same haphazard adjunctions that their readership may be careless copyists or records. Both equate the writings to the same presentation on their unschooled. Indeed, their reflections about America, as about the talents.

Therefore, the text is reproduced—to the extent it appears in the most general principle we employed: All punctuation appears as paragraph breaks, quotation marks only above the line, contriving of a character in half of punctuation that comes rules is followed. Inference confidence from text is usually enclosed in brackets; question mark is a either inference or enclosed in brackets; is represented in the material.
problems, although these are often no more serious than the obstacles created by better-educated but careless clerks, slovenly and hurried military officers, or even the ravages of time upon fragile paper.

The editors have approached the question of transcription with the conviction that readability need not require extensive editorial intervention and, indeed, that modernization (beyond that already imposed by conversion into type) can compromise the historical value of a document. The practical dilemmas of setting precise limits to editorial intervention, once initiated, also suggest the wisdom of restraint. In short, the editors believe that even when documents were written by near illiterates, the desiderata of preserving immediacy and conveying the struggle of ordinary men and women to communicate intensely felt emotions outweigh any inconveniences inflicted by allowing the documents to stand as they were written. Fortunately for the modern reader, a mere passing acquaintance with the primer usually led uneducated writers to spell as they spoke; the resulting documents may appear impenetrable to the eye but are perfectly understandable when read phonetically. In fact, reproduced verbatim, such documents offer intriguing evidence about the spoken language. Other writers, presumably better educated, frequently demonstrated such haphazard adherence to rules of grammar, spelling, and punctuation that their productions rival those of the semiliterate. And careless copyists or telegraph operators further garbled many documents. Both equity and convenience demand, nonetheless, that all writings by the schooled—however incoherent—be transcribed according to the same principles as those applied to the documents of the unschooled. Indeed, a verbatim rendition permits interesting observations about American literacy in the mid-nineteenth century, as well as about the talents or personalities of particular individuals.

Therefore, the textual body of each document in this volume is reproduced—to the extent permitted by modern typography—exactly as it appears in the original manuscript. (The few exceptions to this general principle will be noted hereafter.) The editorial air is never employed: All peculiarities of syntax, spelling, capitalization, and punctuation appear in the original manuscript. The same is true of paragraph breaks, missing or incomplete words, words run together, quotation marks or parentheses that are not closed, characters raised above the line, contractions, and abbreviations. When the correct reading of a character is ambiguous (as, for example, a letter "C" written halfway between upper- and lower-case, or a nondescript blotch of punctuation that could be either a comma or a period), modern practice is followed. Illegible or obscured words that can be inferred with confidence from textual evidence are printed in ordinary roman type, enclosed in brackets. If the editors' reading is conjectural or doubtful, a question mark is added. When the editors cannot decipher a word by either inference or conjecture, it is represented by a three-dot ellipsis enclosed in brackets. An undecipherable passage of more than one word is represented in the same way, but a footnote reports the extent of the illegible material. (See p. xxxiii for a summary of editorial symbols.)
Handwritten letters display many characteristics that cannot be exactly reproduced on the printed page or can be printed only at considerable expense. Some adaptations are, therefore, conventional. Words underlined once in the manuscript appear in italics. Words underlined more than once are printed in small capitals. As for printed forms with blanks filled in by hand, the words originally in print are set in small capitals and the handwritten insertions appear in lower-case, with spaces before and after to suggest the blanks in the form. Internally quoted documents that are set off in the manuscript by such devices as extra space or quotation marks on every line are indented and printed in smaller type. Interlineations are simply incorporated into the text at the point marked by the author, without special notation by the editors unless the interlineation represents a substantial alteration. Finally, the beginning of a new paragraph is indicated by indentation, regardless of how the author set apart paragraphs.

Fig. 4.6b Berlin, Freedom, ser. 1, I.xxxv–xxvii

4.7 In the nineteenth and twentieth centuries, as grammar, spelling, and punctuation have become standardized through the development of style guides and dictionaries, transcribing handwritten documents has become less of a matter of addressing archaic stylistic conventions and more one of deciding how to handle violations of accepted standard practice. Common difficulties encountered in twentieth-century manuscripts include missing or inappropriate terminal punctuation, irregular paragraphing or sentence breaks, nonstandard spelling and grammar, and use of obscure abbreviations. Editors of modern documents evaluate whether to alter these irregular forms to make them more comprehensible or to make them conform to standard written English, and then consider how these changes might affect readers' appreciation of the historical documents. The editors of the Letters of Eugene V. Debs noted the issues presented by Debs's manuscript documents and their methods for addressing them:

As a general principle we have tried to reproduce the original text of each letter as accurately as possible. All the letters are transcribed in full, and idiosyncratic spelling, punctuation, and usage have been preserved.

* * * *

Errors in spelling, punctuation, etc. have been left in the text except in cases where it is judged that they are the result of a typographical error or a slip of a pen. In cases where a word has obviously been left out the word is supplied in brackets, with a question mark if the judgment is uncertain. For illegible words the transcriber has written "[illegible]" or "[two words illegible]," etc. Where a version of the text has been crossed out and rewritten but is still legible, the transcriber has included a line giving a version. Illegible words have been omitted. We have not made an exhaustive transcription by criteria of usage.

When a writer, as Debs often did, the transcriber has substituted a dash in place of an abbreviation or used a period. For example:

transcription: -initials
original: - initials

The editors of the Letters of Eugene V. Debs noted the issues presented by Debs's manuscript documents and their methods for addressing them:

1. Capitalization and deletions
2. The line breaks of the original are not always those of the printed version. The spacing and interword spacing of words are not always consistent in typescripts, or consistent (sometimes not even regularized) in printed versions.
3. Silent editorial insertions in the text, single or in position of two or more.
5. The author's names had been silently omitted but have been brought in as annotations.
6. Insertions in the text of the transcript are made by curly braces...
hat cannot be ex-
d only at consid-
rentional. Words
Words underlined
printed forms with
are set in small
lower-case, with
form. Internally
by such devices as
ind and printed
nto the text at
son by the editor:
on. Finally, the
ion, regardless of

ened, ser.1,xxv–xxvii

mmar, spelling, and
development of style
nts has become less
ore one of deciding
ommon difficulties
sing or inappropriate
breaks, nonstan-
ations. Editors of
ular forms to make
nard written Eng-
derers' appreciation of
Ve Debs noted the
their methods for

the original text
are transcribed
usage have been
in the text except
a typographical
obviously been
ition mark if the
ther has written
version of the
gible, the tran-
scriber has included the crossed-out version, followed by the corrected
version. Illegible crossouts, or errors crossed out and rewritten, have
been omitted. Words inserted above the line are indicated in the
transcription by curved brackets.

When a writer used a symbol for "and" (or a plus sign), as EVD
often did, the transcriber has used an ampersand. When a writer used
a dash in place of a period, as EVD often did, the transcriber has
used a period. Hyphens and dashes have been standardized in the
transcription: - indicates a hyphen, and — indicates a dash, regardless of
the usage of the letter writer.

The editors of the Papers of Martin Luther King, Jr. also specified how they
would address the stylistic issues of their manuscripts:

Our transcriptions are intended to reproduce the source document or recording
as accurately as possible, adhering to the exact wording and punctuation
of the original. Errors in spelling, punctuation, and grammar have been ne-
ither corrected nor indicated by sic. Such errors and stylistic irregularities may
offer important insights into the author's state of mind and conditions under
which a document was composed. Other features that could not be adequately
reproduced, such as signatures and handwritten marginal comments, are
noted and described in the text or footnotes.

* * *

1. Capitalization, boldface, symbols, subscripts, abbreviations, strikeouts,
and deletions are replicated regardless of inconsistency or usage.
2. The line breaks, pagination, and vertical and horizontal spacing of the
original are not replicated. A blank line signals a break in the text other
than a straightforward paragraph break. The transcription regularizes
spacing and indentation of paragraphs, outlines, and lists, as well as the
spacing of words, initials, and ellipses.
3. The underlining of book titles, court cases, or other words and phrases
in typescripts is reproduced. Since underlining practices were often in-
consistent (sometimes breaking between words, sometimes not), we regu-
larized the various types to continuous underscoring.
4. Silent editorial corrections have been made only in cases of malformed
letters, single-letter corrections of typescript words, and the superim-
position of two characters.
5. The author's use of hyphens is replicated, but end-of-line hyphens have
been silently deleted unless the usage is ambiguous. Dashes between
numbers are rendered as en-dashes. Em-dashes, which appeared in se-
veral styles in the original manuscripts, have been regularized.
6. Insertions in the text by the author (usually handwritten) are indicated by
curly braces ({{ }}) and placed as precisely as possible.

Fig. 4.7a Constantine, Debs, l:xxxiii–xxxiv

Fig. 4.7b Carson, King, 2:50–51,53
Technical or Professional Documents

4.8 In the course of performing their work, many professionals—including but not limited to architects, doctors, lawyers, business people, and scientists—produce documents that can be difficult to transcribe because of their technical terminology and unusual format. To transcribe professional documents, editors acquaint themselves with the language of the profession. They then develop methods for transcribing documents that present problems of size, layout, complexity, quantity, or organization. Different types of professional or technical documents pose their own unique challenges and require editors to consider how best to convert them into print.

4.9 Business documents, especially ledgers or account books, present editors with tables of numbers or monetary amounts and abbreviations that are difficult and often expensive to transfer into type. The editors of the Papers of John Marshall created a tabular format for Marshall’s account book that presented information without replicating the exact layout of the original page:

<table>
<thead>
<tr>
<th>Stockins for P.4 6 Dollars, mustard pot 9</th>
<th>£</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 yds. Oznabrugs at 1/3</td>
<td>5 17 9</td>
</tr>
<tr>
<td>Inkstand 8½ pomatum ½/6</td>
<td>8 9 6</td>
</tr>
<tr>
<td>Land warrants for my Father</td>
<td>1 15</td>
</tr>
<tr>
<td>Fee to the Attorney for Do.</td>
<td>1 2 6</td>
</tr>
<tr>
<td>Fee to the Register for a Patent for Do.</td>
<td>0 15 10</td>
</tr>
<tr>
<td>For Oats 36 Bushels</td>
<td>3 10 6</td>
</tr>
<tr>
<td>To one pair of Stockings</td>
<td>1</td>
</tr>
<tr>
<td>To trimmings for Mr. Keith</td>
<td>1 4 9</td>
</tr>
<tr>
<td>To advice fee given the Attorney for opinion on surveyors fees</td>
<td>1 2 6</td>
</tr>
</tbody>
</table>

Fig. 4.9 Johnson, Marshall, 1.294

4.10 Documents produced by scientists often use technical language and symbols that make them difficult to present in type. Editors need to exercise care not to distort the notation of scientific documents. The editors of the Papers of Joseph Henry used a combination of type and facsimile to represent Henry’s notes:

\[-\frac{1}{2}\int_{\Gamma} |E||S| r^2 d\Omega = \frac{1}{c^4} \int_{\tau} \int_{\Omega} (E||S|) r^2 d\Omega = -\alpha \]

4.11 Editors publishing documents will need to develop technical and engineering drawings, oversized book edition, and facsimile to reproduce Einstein's notes.
"RECORD OF EXPERIMENTS"

Henry Papers, Smithsonian Archives

Monday Oct 23rd 1843

Exper. on the holes pierced

in cards by the secondary current

Commenced this morning the Repetition of Matteucci's experiments to
determine the direction of the induced current by means of the pierced card.

The points were placed 1/4 of an inch apart, with one jar charge

60 <three> two small hol[e]s near the — minus point current

adverse.

Charge 80 three holes nearer the minus pole. The larger hole was

ragged—one jar distance same

Charge 90. Distance same three holes nearer the minus side. When

these holes are exami[na]ed with a magnifying glass the larger appears ragged and triangular.

Fig. 4.10a Reingold, Henry, 5:424

The editors of the Collected Papers of Albert Einstein also used both type and

facsimile to reproduce Einstein's notes:

\[
\int_0^\pi \int_0^{\pi/2} |E| |\mathbf{\hat{\tau}}| r^2 \ d\Omega = \frac{1}{c^2} \int_0^{\pi/2} dt \int_0^{\pi/2} \sin^2 \theta \ d\theta \left\{ \frac{\sin \theta \ d\theta}{2\pi \sin \theta \ d\theta} \right\} \\
= \left[ -\cos \theta + \frac{\cos^3 \theta}{3} \right]_0^\pi = 2 - \frac{2}{3} = \frac{4}{3}
\]

Fig. 4.10b Stachel, Einstein, 3:396

4.11 Editors publishing documents produced by engineers and architects

will need to develop techniques for presenting sketches and blueprints. The

editors of the Papers of Benjamin Henry Latrobe presented Latrobe's architectural and engineering drawings as annotated photographic facsimiles in an

oversized book edition, and they published facsimiles in the series' microfiche
edition (not shown).
4.12 Editors transcribing governmental and legal documents should be absolutely literal when these documents may be consulted by lawmakers, lawyers, and judges in order to understand precedents and to comply with existing laws. They also may need to note stray, incorrect, or missing punctuation in variant copies; these phenomena can alter meaning or affect public policy. The effect of such errors is real, as explained in a footnote to a passage in the *Territorial Papers of the United States*:

The governor, and judges or a majority of them shall adopt and publish in the district such laws of the original states criminal and civil as may be necessary and best suited to the circumstances of the

14 In the official printed copies of the ordinance, signed by Charles Thomson as Secretary of Congress, there appear certain variations in punctuation from the original Journals which were sufficient to leave the meaning of the former in doubt. The most significant of these differences occur in connection with the above sentence. In the printed copies no comma appears after the word “governor”, as in the Journal; on the other hand, a comma is inserted after the word “judges”, and after the word “them”. In a contemporary manuscript copy (CC Papers, no. 175, fol. 123) the comma is likewise omitted after “governor”. The punctuation in the printed copies, which were in the hands of territorial officials, provoked serious differences of opinion between Governor and Judges as to their respective legislative powers.

Fig. 4.12 Carter, *Territorial Papers*, 2:42

**Machine-Created Documents**

4.13 Although they offer the benefit of increased legibility, machine-created documents pose their own problems for transcribers. The conventions used by typesetters, telegraph operators, and typists need to be transferred into a readable format, and editors must decide which incidental marks bear information of use to readers. Editors create policies for transcribing machine-created documents, such as this one from the *Samuel Gompers Papers*:

Mechanically reproduced documents, such as articles printed in newspapers or typed letters, require a different treatment. Typographical and spelling errors seem to have been endemic to the labor bureaucracy and press of the late nineteenth and early twentieth centuries. For the most part it is impossible to determine which of these errors were the work of the documents’ authors and which were anonymously introduced by the typographers and secretaries of the day. Since their sheer numbers make them distracting, how many of them, with the following changes that would substantiate if they seem essential to the certain features of the document: capitalization, the spelling of usage, such as the use of “cigar,” the spelling of Pittsburgh or lowering the word “street” punctuation usually appears typographical error and ten.

The editors of the *Black Abolition* ments of nineteenth-century print.

The project adopted the following published sources (newspapers, nineteenth-century printed materials: quotation marks are converted to parentheses; audience separate sentence with parent: let stand certain nineteenth-century names or addresses in capital: visual character of the document is deleted from a printed document; material edited or deleted; but with material extraneous to the published.

4.14 Newspapers have employed typefaces of different styles and columns divided by subheadings transcribing newspaper documents stylistic devices they wish to re.
legal documents should be consulted by lawmakers, laws and to comply with contracts, or missing punctuation or affecting public policy and the footnotes to passages in the

them shall adopt and criminal and the circumstances of the

and by Charles Thomson as in punctuation from the meaning of the former in our in connection with the pears after the word “government” is inserted after the word temporary manuscript copy omitted after “governor” in the hands of territorial Governor and Judges as

Carter, Territorial Papers, 2:42

s eligibility, machine-created The conventions used by transferred into a readable marks bear information cribbing machine-created Papers:

Mechanically renewed newspapers or typographical and spelling error accuracy and presses. For the most parts were the work of hastily introduced by their sheer numbers

make them distracting, however, the editors have silently corrected many of them, with the following exceptions. There are no changes in the original syntax of mechanically produced documents and no changes that would substantively alter their meaning. Errors remain if they seem essential to the particular mood or style of a document. Certain features of the documents generally appear as found, including capitalization, the spelling of contemporary personal names, errors of usage, such as the use of a noun form for a verb, and common spelling conventions of the time, for instance, the use of “segar” for “cigar,” the spelling of Pittsburgh, Pennsylvania, without the final “h,” or lowercasing the word “street” as part of a street name. The original punctuation usually appears as found, except where it is an apparent typographical error and tends to garble the document’s meaning.

Fig. 4.13a Kaufman, Gompers, i:xxii–xxiv

The editors of the Black Abolitionist Papers also modernized many of the elements of nineteenth-century printed documents:

The project adopted the following principles for documents found in published sources (newspapers, pamphlets, annual reports, and other nineteenth-century printed material): redundant punctuation is eliminated; quotation marks are converted to modern usage; obvious misspellings and printer’s errors are corrected; printer’s brackets are converted to parentheses; audience reaction within a speech is treated as a separate sentence with parentheses, for example, (Hear, hear.). We have let stand certain nineteenth-century printing conventions such as setting names or addresses in capital or italic letters in order to maintain the visual character of the document. A line of asterisks signals that material is deleted from a printed document. In no instance is black abolitionist material edited or deleted; but if, for example, a speech was interrupted with material extraneous to the document, the irrelevant material is not published.

Fig. 4.13b Ripley, Black Abolitionist, 4:xxiv

4.14 Newspapers have employed a wide variety of printing styles, including typefaces of different styles and sizes, idiosyncratic use of italic and bold type, columns divided by subheadings, and extensive use of quotation marks. In transcribing newspaper documents, editors need to decide how many of these stylistic devices they wish to replicate and how best to transcribe them. The
editors of the Marcus Garvey and Universal Negro Improvement Association Papers transcribed a newspaper story in the following manner:

### Negroes Plan to Found Ship Line

**Project Will Be Discussed at Palace Casino Rally Tomorrow Night.**

The Universal Negro Improvement Association and African Communities League, recognizing, according to its president-general, Marcus Garvey, that this is a "selfish age," will hold a mass meeting at the Palace Casino, 135th street and Madison Avenue, tomorrow night to prove the Negro has caught the spirit.

According to Garvey, the Negroes are anxious to go back to Africa and the West Indies and create empires of their own as strong as that of the yellow and the white man. To go back, they need ships.

Therefore, at this mass meeting the founding of the "Black Star Line" will be attempted. The proposed steamship line will operate between American ports and those of Africa, the West Indies, Central and South America.

---

**Newspaper Report**

*New York Call, 27 April 1919*

**NEGROES PLAN TO FOUND SHIP LINE**

**PROJECT WILL BE DISCUSSED AT PALACE CASINO**

**RALLY TOMORROW NIGHT**

The Universal Negro Improvement Association and African Communities League, recognizing, according to its president-general, Marcus Garvey, that this is a "selfish age," will hold a mass meeting at the Palace Casino, 135th Street and Madison Avenue, tomorrow night to prove the Negro has caught the spirit.

According to Garvey, the Negroes are anxious to go back to Africa and the West Indies and create empires of their own as strong as that of the yellow and the white man. To go back, they need ships.

---

The editors of the Document modernized the stylized typewritten text to provide a natural reading experience.
Negro Improvement Association

Therefore, at this mass meeting the founding of the "Black Star Line" will be attempted. The proposed steamship line will operate between American ports and those of Africa, the West Indies, Central and South America.

Printed in the New York Call, Sunday, 27 April 1919.

1. The Negro World of 3 May 1919 reported that the UNIA "has opened a special account with the Corn Exchange Bank, 15th Street and Lenox Avenue, to receive lodgments of the Black Star Line" (Garvey v. United States, no. 837 [Ct. App. 3d Cir., Feb. 2, 1919], p. 1908). The same announcement reported that the secretary of the line was "W. T. Mitchell" and that the treasurer was George W. Tobias. The former was actually Uriah T. Mitchell, a Jamaican who at the time also held the position of UNIA secretary. However, a few weeks later Mitchell resigned from the UNIA, after temporarily serving as a member of a three-man committee auditing the UNIA's accounts. The committee also charged Garvey before New York Assistant District Attorney Edwin P. Kimbro, alleging financial mismanagement and collection of money from UNIA members under false pretenses.

The editors of the Documentary History of the Ratification of the Constitution modernized the stylized type and use of quotation marks when they transcribed the following newspaper passage:

13th. Ex post facto laws have ever been considered as abhorrent from liberty: necessity and public safety never can require them—"If laws do not punish an offender, let him go unpunished; let the legislature, admonished of the defect of the laws, provide against the commission of future crimes of the same sort—The escape of one delinquent can never produce so much harm to the community, as may arise from the infraction of a rule, upon which the purity of public justice, and the existence of civil liberty essentially depend"—Paley's Principles of Moral Philosophy, vol. 2, p. 234. Ov[el][o] Ed.

19th. Ex post facto laws have ever been considered as abhorrent from liberty: necessity and public safety never can require them—"If laws do not punish an offender, let him go unpunished; let the legislature, admonished of the defect of the laws, provide against the commission of future crimes of the same sort—The escape of one delinquent can never produce so much harm to the community, as may arise from the infraction of a rule, upon which the purity of public justice, and the existence of civil liberty essentially depend"—Paley's Principles of Moral Philosophy, vol. 2, p. 234. Ov[el][o] Ed.
4.15 Most telegrams or cables contain transmittal symbols in addition to a message. Editors must therefore decide which portions of a telegram warrant reproduction. The editors of the Family Letters of Victor and Meta Berger standardized the placement of the place, date, and message of telegrams, corrected keyboarding errors, converted the word period used by the telegraph operators to indicate the end of a sentence into the terminal punctuation itself, and presented the text in small capital letters to distinguish it from typed or handwritten letters:

Victor to Meta (telegram)

WASHINGTON, D.C.
JANUARY 10, 1920

CRUCIFIED ONCE MORE! THEY DID NOT EVEN HAVE THE DECENCY TO GRANT ME TEN MINUTES FOR MY DEFENSE. MAN N SHERWOOD AND VOIGHT WANTED ME SEALED AND SISSEN PROTESTED AGAINST THE INJUSTICE OF PROCEDURE. I WILL HAVE TO FIGHT AGAIN AND BE REELECTED ARRIVE HOME MONDAY NIGHT LOVINGLY.

VICTOR BERGER

Fig. 4.15a Stevens, Berger, 287

The Papers of Woodrow Wilson used the same typeface and headings for telegrams and letters but preceded the telegrams with the date and place from which they were sent and indicated that they were telegrams in the note following the document:

To Mary Allen Hulbert Peck

Trenton, N. J. Aug. 15, 1911

Ellen joins me in begging that you and Allen come to us Friday morning best train leaves New York at nine four morning.

Woodrow Wilson.

T telegram (WC, NJP).

Fig. 4.15b Link, Wilson, 23:267

The editor of the article "The Great Fire of 1871: A Nation Responds," which appeared in the magazine Voyager published by the Brown County [Wisconsin] Historical Society, used the following standardized format to present telegrams:

To: Gov. Fairchild
By Telegram from: J.B. Stickney, Mazomanie
Date: October 11, 1871

I have 11 bxs of bread & provisions brought here for the Chicago sufferers I have just rec’d notice from S.S. M—that it is not needed there Can it be sent to those in need in our own sate & how Answer

Fig. 4.15c Olson, Voyager, 13:16

4.16 The advent of the typewriter legible texts but also created typists and the malfunctions typographical errors, partial legibility of letters. The editor of the following irregular spacing and the handwriting when he transcribed the following:

August 15, 1895
Woodstock, Illinois

My dear Sir and Friend:—

I have received a very an inmate of the Vigo Co. of age, a mere child I show of larceny because of he nothing of the merits of it you well and would write or but half true, it seem wrong and yet in your n have escaped your person, I sure it is only necessary justice being done. With

Dict. E. V. D.

Fig. 4.16a Link, Wilson, 23:267

AMERICAN

Hon. Samuel Houston,
Prosecuting Att’y, Vlg,
Terre Haute

My Dear Sir and Friend:—

I Ellen M. Lupsin, now an im she is but 16 years of age.
4.16 The advent of the typewriter made possible the quick production of legible texts but also created a new set of concerns for transcribers. Slips by typists and the malfunctions of typewriters could create irregular spacing, typographical errors, partial letters, typed-over letters, erasures, and uneven or raised letters. The editor of the Letters of Eugene V. Debs (top) standardized irregular spacing and the hand-inserted corrections made by Debs or his typist when he transcribed the following typewritten letter (bottom):

August 15, 1895
Woodstock, Illinois

My dear Sir and Friend:—

I have received a very pitiable story from Ellen M. Lappin, now an inmate of the Vigo County Jail. She informs me she is but 16 years of age, a mere child I should judge and that she is in jail on a charge of larceny because of her failure to give bond. I, of course, know nothing of the merits of the case but have written her saying I knew you well and would write to you in her behalf. If her story is true, or but half true, it seems harsh to have imprisoned her under the circumstances. I know you will not suffer her to be subjected to a wrong and yet in your multifarious duties I thought the case might have escaped your personal attention and hence this letter. I am quite sure it is only necessary you should understand the case to insure justice being done. With cordial regards and best wishes, I am,

Yours Very Truly,
Eugene V. Debs

Dict. E. V. D.
she is but 16 years of age, a mere child I should judge and that she is in jail on a charge of larceny because of her failure to give bond. I, of course, know nothing of the merits of the case but have written her saying I knew you well and would write you in her behalf. If her story is true, or but half true, it seems harsh to have imprisoned her under the circumstances. I know you will not suffer her to be subjected to a wrong and yet in your multifarious duties I thought the case might have escaped your personal attention and hence this letter. I am quite sure it is only necessary you should understand the case to insure justice being done. With cordial regards and best wishes, I am,

Yours Very Truly,

Eugene F. Debs

Fig. 4.16b Constantine, Papers of Debs, reel 1:212

4.17 Editors transcribing printed forms or tables evaluate the preprinted information on these documents and decide how much of it should be reproduced either in the text or in footnotes. Printed forms include a wide array of documents such as tax forms; census forms; applications; business forms; birth, marriage, and death certificates; military and government documents; printed date books; and diaries. Editors must also decide if it is necessary to distinguish preprinted passages from those inserted at a later time.

4.18 A common form of preprinted matter is letterhead. When documents are written on letterhead, editors decide whether to reproduce the information from the letterhead. In most cases, the content of letterhead is not printed, but in rare instances it might be transcribed or reproduced as a facsimile to provide otherwise unavailable information about organizations, places, or people. The editors of the Marcus Garvey and Universal Negro Improvement Association Papers transcribed an example of the letterhead of the association, probably because it provided evidence about the organization's leaders and conveyed information about its public presentation.

"He created of one race a people of power to sway the world's destiny."

My dear Dr. Baldwin,

It is unfortunate that the public and other blacks as you will realize when there is no path. Thus, in keeping with meeting in your help by a small one. We feel sure to receive your kind respect fully.
I DO NOT LIKE TO SEE THE NEGROES SUFFER AT THE HANDS OF THE WHITE PEOPLE. I SHOULD BE DIRECTED TO CONSIDER THE CASES OF THE NEGROES AND NOT TO SEE THE WHITE PEOPLE. I AM SURE THAT THE NEGROES ARE HUMAN AND THAT THEY ARE CAPABLE OF GOOD DEEDS.

Very Truly,

[Signature]

5 Constandine, Papers of Debs, reel 1:212

or tables evaluate the preprinted forms; how much of it should be reproduced in the document; applications; business forms; and government documents; is also necessary to be inserted at a later time.

or letterhead. When documents, rather than in the case of letterhead, are not printed, but reproduced as a facsimile to promote organizations, places, or people.

One God! One Aim! One Destiny!

Universal Negro Improvement Association
AND
African Communities' League

Cable Address: "UNIANY", New York Telephone 7976 Morningside

IRMA MOORMAN-BLACKSTON
President Ladies' Division

IRENE W. WINGFIELD
Pres. Vice President

JANIE JENKINS
3rd Vice President

R. H. ROGERS
3rd Vice President

AMEE OUGHTON-CLARKE
General Secretary

CARRIE MERO
Associate Secretary

CLARENCE A. CARPENTER
Treasurer

LAFAYETTE BUILDING, 165 W. 131ST ST.
NEW YORK, APRIL 29TH 1918

"He created one blood all nations of men to dwell on the face of the earth."

My dear Dr. Butler:

It is unfortunate that we have to tell you at this time, when the public and other financial demands are so many on one of your position, but as you will realize sometimes, one has to appeal to the old friend for help, when there is no one else to ask.

Thus, in keeping with an unanimous vote of our members, at our last general meeting in the Lafayette Hall, I was instructed to write to you asking your help by a small donation to assist our building fund.

We feel sure that whatever you can do for us, will be done. Trusting to receive your kind help, with return of list, with best wishes, I am Yours respectfully,

UNIVERSAL NEGRO IMPROVEMENT ASSOCIATION
MARCUS GARVEY PRESIDENT

Fig. 4:18a Hill, Garvey, 1:242
The editors of *Mark Twain's Letters* also transcribed the letterhead used for some documents:

J. LANGDON, MINER & DEALER IN ANTHRACITE & BITUMINOUS COAL  OFFICE NO. 5 BALDWIN STREET  ELMIRA, N.Y. March 31—1869.

Dear Mother—

Bless you I don't want to go to California at all—and really I have not by any means determined to go, as yet. I know very well that I ought to go, but I haven't the slightest inclination to do it. Indeed, indeed,

Fig. 4.18b Branch, Twain, 3:184

4.19 The significant information in birth, marriage, and death certificates as well as census rolls usually consists of a few words of text written on a standard form. The editors of the *Collected Papers of Albert Einstein* presented his birth certificate, with the printed portion of the form in roman type and the inserted handwritten information in italics:

1. Birth Certificate

   *Ulm am 15. März 1879*

   Nr. 224.

   Vor dem unterzeichneten Standesbeamtten erschien heute, der
   Persönlichkeit nach bekannt,
   der Kaufmann Herman Einstein
   wohnt zu Ulm Bahnhofstraße B Nr. 135.[1]
   jüdischer Religion, und zeigte an, daß von der
   Pauline Einstein geb. Koch seiner Ehefrau
   jüdischer Religion, wohnt bei ihm
   zu Ulm in seiner Wohnung
   am vierzehnten März des Jahres
   tausend acht hundert siebzig und neun Vormittags
   um elf ein halb Uhr ein Kind
   Geschlechts geboren worden sei, welches
   erhalten habe
   Vorgelesen, gezeichnet und unterschrieben
   *Hermann Einstein*

   Der Standesbeamte.
   *Hartmann*

Fig. 4.19a Stachel, Einstein, 1:1

4.20 By contrast, the editors of the census return for the form and making no portions of the document.

**Entry in the**

<table>
<thead>
<tr>
<th>Liberty Old Antique Ground</th>
<th>Name of Street, Place, or Road, and Name or No. of House</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Householder's Schedule</td>
<td>118</td>
</tr>
</tbody>
</table>

Printed form, with pp. 24-25, PRO. 7, Blind, or Deaf-ar.

1. Actually Herm.
2. Fanny (Femme)
3. Clara (Mietze)
4. Catherine (G.
5. Sarah Gomp.
6. Simon Gom,
7. SC.

4.20 Editors nu...unication that lack classroom lecture, extemporaneously.
transcribed the letterhead used for

NER & DEALER IN ANTHRACITE & COAL OFFICE NO. 6 BALDWIN STREET MIRA, N.Y. March 31—1869.

California at all—and really I have yet, I know very well that I ought
ination to do it. Indeed, indeed,

Fig. 4.18b Branch, Twin, 3:184

marriage, and death certificates as
words of text written on a stan-
ers of Albert Einstein presented his
if the form in roman type and the

Ulm am 15. März 1879

r. 224.
beamten erschien heute, der
bekannt,

Vorstands

und zeigte an, daß von der
israelitische Religion,

des Jahres
sein Vormittags
männlichen
ches
der Vornamen Albert[2]
hrieben

Der Standesbeamte. Hartmann

Fig. 4.19a Stachel, Einstein, 1:1

by contrast, the editors of the Samuel Gompers Papers printed the 1851 London
census return for the Gompers family, omitting a column from the printed
form and making no clear distinction between the printed and handwritten
portions of the document:

Entries for the Gompers Household
in the 1851 Census Return for London

| No. of Householder's Schedule | Name of Street, Place, or Road, and Name or No. of House | Name and Surname of Each Person who
|                             |                                                      | Stood in the House on the Night of the
|                             |                                                      | 5th March, 1851 |
| 113                        | 24 Port S.                                          | Catherine Gompers' |
|                            |                                                    | Wife               |
|                            |                                                    | M                  | 46 | General Dealer | Holland |
| Solomon                   | Man                                                 | M                  | 22 | Cigar Maker   | do |
| Sarah                     | Son                                                 | M                  | 24 | do            | do |
| Fanny                     | Daughter                                            | UM                 | 14 | Cap Maker     | do |
| Mary                      | Daughter                                            | UM                 | 22 | do            | do |
| Catherine                 | Daughter                                            | UM                 | 10 | do            | do |
| Sarah                     | Daughter                                            | UM                 | 8  | do            | do |
| Simon                     | Son                                                 | M                  | 3  | Middendorf     | Spandau |
| Samuel                    | Child                                               | M                  | 3  | do            | do |
| May Duggar                | Servant                                             | UM                 | 20 | General Servant | Ireland |
| Anne [Rowe]               | Lodger                                              | M                  | 55 | General Dealer | Holland |
| Isaac Goldsmith           | Lodger                                              | UM                 | 17 | General Dealer | do |
| Joel [Rosenber]           | Lodger                                              | UM                 | 21 | Hawker        | do |
| Abraham Martin            | Lodger                                              | UM                 | 55 | Hawker        | Holland |

Printed form, with handwritten entries, Census of 1851, H.O. 107/1543, folio 46, pp. 24-25, PRO. The following column was left blank by the census taker: "Whether
Blind, or Deaf-and-Dumb."

1. Actually Henrietta (Jette) Salomme Harling GOMPERS (1807-79), SG's paternal grandmother.
2. Fanny (Fanny) Gompers Cohen, SG's aunt.
4. Catherine (Catherine) GOMPERS, SG's aunt.
5. Sarah Gompers Levy, SG's aunt.
6. Simon GOMPERS, SG's uncle and childhood companion, became a shoemaker.
7. SG.

Fig. 4.19b Kaufman, Gompers, 1:8

Oral Documents

4.20 Editors may be faced with the challenge of transcribing forms of commun-
ication that lack written sources. Spoken words—whether discussions, speeches,
classroom lectures, or sermons—may have been delivered from brief notes or
extemporaneously. To reproduce these works, editors employ the sources that
preserved the information, whether outlines, shorthand transcriptions, notes, or newspaper reports. Although these sources are incomplete and introduce inaccuracies, they are often the only means for recapturing such "lost" documents.

4.21 For sermons and speeches that either were delivered before the advent of modern recording technology or were not recorded, editors must rely on text or notes used by the speaker or accounts by audience members. The editors of the Papers of Martin Luther King, Jr. included King's outline notes to provide information about a sermon of his for which no other source existed:

"The Challenge of the Book of Jonah"

[1951-1955]
[Boston, Mass.]

The following sermon outlines were found among the hundreds of notecards King prepared for his courses at Boston. In the sermon King argues that God's love is universal and inclusive of all faiths and races.

No more delightful moments can be spent than those spent reading the book of Jonah. It is one of the greatest books of the Old Testament. Its themes is both arresting and electrifying. Its unknown author appears to have possessed the vision of a Saint Paul, the satiric power of a George Bernard Shaw, and the delicious humor of a G. K. Chesterton. This book does not represent an actual occurrence any more than the parable of the prodigal son. But who can doubt the accuracy of either as portraits of multitudes of human hearts.

To often have we spent our time arguing over the historicity of Biblical stories, while failing to grasp the underlying truths.

Let us look at this story for a moment and see what it has to offer us. Recall the story

I Tell the Story
II This story has within it two [three] fundamental truth that I would like to set forth

1. God's love is boundless and Universal
   (1) God loves the Ninevites
   (2) We are all involved in a single process and whatever effects one directly effects the other indirectly. So long as there is slavery in the world I can never be totally free.
   (3) Science has made this obviously true. We must have one World or none.
   (4) Quote John Donne

2. All men are their brothers keepers and dependent on each other
   (1) God's love to see this
   (2) We are all involved in a single process and whatever effects one directly effects the other indirectly. So long as there is slavery in the world I can never be totally free.
   (3) Science has made this obviously true. We must have one World or none.
   (4) Quote John Donne

Wilson and Hamilton moved

Mr. Madison supposed the Executive should be required to over answer the same purpose as happen that the Executive have firmness eno' to resist the body itself. The King would not be able to withstand the houses of Parliament. To get anxious to the temper of the

Some time after Eppes copied the concluding clause.

Yates's version:
"Mr. Madison against it—by exercise of this negative, and is of (Farrand, Records, I, 106).

King's version:
"Mad: I am opposed to the constitution the law was passed almost unanimously, and sufficient to do it" (ibid., I, 107).

Pierce's version:
"Mr. Madison was of opinion that had passed with the assent of

For oral presentations made use newspaper reports. The their methods for finding accurate challenges they posed:

Since newspapers furnishing procedures adopted problems caused by the specialization. Struggling to capture words spoken, nineteenth-century combination of italics, o
thand transcriptions, notes, or complete and introduce inaccuracies such "lost" documents. 

're delivered before the audience members, editors must rely on these written notes. The transcribed King's outline notes to which no other source existed:

of Jonah" [1951-1953]
[Boston, Mass.] of notecards King argues that God’s love is

n those spent reading the Old Testament. Its themes appear to have passed a George Bernard Shaw, as book does not represent the prodigal son. But who altitudes of human hearts, storied in Biblical stories,

it has to offer us. Recall

ai truth that I would like

is and the whole Jewish because God loves them.

jus significant as the silly is just as worthwhile to munition. dependent on each other. ss and whatever effects cally. So long as there is tally free.

ie. We must have one

21a Carson, King, 2:325-326

4.21 Transcription: Types of Sources

Editors may use notes taken by listeners to reconstruct oral documents, although such notes are often fragmentary, difficult to understand, or inaccurate because authors tend to write them for their personal reference. The notes taken by James Madison during the Constitutional Convention provide important information about that event, but as the alternative accounts in footnote 3 of the following excerpt indicate, his notes offer only one perspective on the events taking place:

Wilson and Hamilton moved to give the executive an absolute veto on laws.

Mr. Madison supposed that if a proper proportion of each branch should be required to overrule the objections of the Executive, it would answer the same purpose as an absolute negative. It would rarely if ever happen that the Executive constituted as ours is proposed to be would have firmness enou' to resist the legislature, unless backed by a certain part of the body itself. The King of G.B. with all his splendid attributes would not be able to withstand the unanimous and eager wishes of both houses of Parliament. To give such a prerogative would certainly be obnoxious to the temper of this Country, its present temper at least.

3Some time after Eppes copied the Debates, JM heavily deleted a sentence and interlined the concluding clause.

Yates’s version:

"Mr. Madison against it—because of the difficulty of an executive venturing on the exercise of this negative, and is therefore of opinion that the revisional authority is better" (Farrand, Records, 1, 106).

King’s version:

"Virtually: I am opposed to the complete negative, because no man will dare exercise it when the law was passed almost unanimously. I doubt whether the King of Eng. wd. have firmness sufficient to do it" (ibid., 1, 107).

Farey’s version:

"Mr. Madison was of opinion that no Man would be so daring as to place a veto on a Law that had passed with the asent of the Legislature" (ibid., 1, 109).

Fig. 4.21b Hutchinson, Madison, 10:24-25

For oral presentations made by noteworthy speakers, editors may be able to use newspaper reports. The editors of the Frederick Douglass Papers explained their methods for finding accounts of speeches in newspapers and some of the challenges they posed:

Since newspapers furnished most of Douglass’s speeches, the editorial procedures adopted represent attempts to solve some of the problems caused by the special characteristics of nineteenth-century journalism. Struggling to capture the flavor of an event as well as to record the words spoken, nineteenth-century reporters often used an unpredictable combination of italics, capital letters, and quotation marks to indicate
the words and phrases a speaker emphasized. Similarly, journalists' attempts to convey the rhythm of a speech frequently resulted in irregularly punctuated sentences. Usually reports included in parentheses notations of the speaker's gestures and audience reactions (cheers, hisses, interruptions). Long paragraphs often containing 2000 words were the rule. Journalists were inconsistent in capitalizing proper nouns and rarely included titles or date and place lines in their reports of speeches.

Nineteenth-century recording techniques were crude, especially in the United States, where editors had a limited number of trained stenographers upon whom they could rely. Many editors—recognizing this, and trying to protect themselves from potential libel suits—began each published speech with a phrase such as "the substance of" or converted stenographically recorded speeches into third person narratives. Harried by limited space in even the largest papers, editors often published only brief mentions of, or short extracts from, speeches.

* * *

Completeness, accuracy, and historical significance were the major criteria used in selecting the speeches included in this edition. In an effort to obtain the most accurate copies of Douglass's orations, we attempted to locate all of the extant versions of each speech. We have not, however, published each version we discovered. Extracts and brief mentions, lacking detail and context, are generally excluded. Summaries of speeches are included only when, in the editors' judgment, their topics are of special historical significance. The few extant handwritten copies of Douglass's speeches have been treated as pretexts and compared to all pamphlet copies and newspaper reports of them to determine whether he said what he intended to say.

Variants of stenographically recorded speeches posed special problems. When considering variants the editors tried to determine the "best copy." Fortunately, Douglass and his friends were quick to praise journals reporting his speeches accurately and to castigate those printing garbled versions of them. It was possible, in many instances, to determine the most accurate of two or more variant stenographic copies by comparing each to long narrative accounts of the same speech. The final test for the best of several versions of a speech was close textual analysis. Intuition and familiarity with Douglass's life and work inevitably played a role in our selections.

Fig. 4.21c Blassingame, Douglass, ser.1,1:4rev.book

The editors of the Papers of Woodrow Wilson explained in the introduction to volume 24 how they corrected obvious transcription errors introduced by Wilson's personal stenographer. The editors converted questionable passages back into
sharpered. Similarly, journalists' at- speech frequently resulted in ir-
ily reports included in parentheses and audience reactions (cheers, 
often containing 2000 words were 
7 in capitalizing proper nouns and 
lines in their reports of speeches. 
iques were crude, especially in 
limited number of trained stenog-
Many editors—recognizing this, 
potential libel suits—began each 
"the substance of" or converted 
to third person narratives. Ham-
largest papers, editors often pub-
e xtracts from, speeches.

The transcripts of oral interviews conducted during 1992 and 
1993 form the basis of the text printed here. We have tried to create 
a clean, readable text without sacrificing the original language of 
the interviews. Because written English differs from spoken English, 
we employed a number of conventions to deal with variations. We 
did not add or change any words, and we did not tamper with 
graham or sentence structure. Words added to clarify the text al-
ways appear in brackets. The transcripts omit false starts as well as 
filler words, such as "you know" or "um." We made no attempt to 
preserve dialect or pronunciation. The original tapes are available 
for those interested in such nuances.

The interviews ranged in length from one hour to more than 
three hours, and the texts presented here are excerpts. Questions 
and answers are presented in the order in which they occurred. 
When part of the answer to a question has been deleted, we note it 
with ellipsis points, although we have not used them to indicate 
omissions of entire questions and answers. The questions asked by 
the interviewers appear in italic type and have been edited for 
clarity.

In every case, we contacted the interviewees and asked them to 
review the text selected for inclusion in this book. In several cases 
they requested (and we made) minor changes that correct inaccura-
cies or clarify statements.

The editors of the Papers of Martin Luther King, Jr. transcribed tape-recorded 
sermons, including descriptions of audience responses in italics within square 
brackets, such as: [laughter]. Audience comments and interjections were 
inserted in italic type within parentheses (Go ahead), and passages emphasized 
by King were set in italics:

Now that isn't the only thing that convinces me that we've strayed away from 
this attitude, (Go ahead) this principle. The other thing is that we have adopted 
a sort of a pragmatic test for right and wrong—whatever works is right. (Yes) 
If it works, it's all right. Nothing is wrong but that which does not work. If 
you don't get caught, it's right. [Laughter] That's the attitude, isn't it? It's all 
right to disobey the Ten Commandments, but just don't disobey the Eleventh, 
Thou shall not get caught. [Laughter] That's the attitude. That's the prevailing
attitude in, in our culture. \textit{(Come on)} No matter what you do, just do it with a, with a bit of finesse. \textit{(All right)} You know, a sort of attitude of the survival of the slickest. Not the Darwinian survival of the fittest, but the survival of the slickest—who, whoever can be the slickest is, is the one who right. It's all right to lie, but lie with dignity. [Laughter] It's all right to steal and to rob and extort, but do it with a bit of finesse. \textit{(Yes)} It's even all right to hate, but just dress your hate up in the garments of love and make it appear that you are loving when you are actually hating. \textit{Just get by!} That's the thing that's right according to this new ethic. \textit{(Lord help him)}

Fig. 4.22b Carson, King. 2:252

The editors of George C. Marshall: Interviews and Reminiscences for Forrest C. Pogue transcribed a series of oral histories, including both questions and responses, and added headings to indicate the date and number of the original recording:

\textbf{Tape 8}

\textit{Recorded December 7, 1956}

1. \textit{What were the basic weaknesses of the National Guard and how did you attempt to eliminate these?}

As to the basic weaknesses: There were some of these that could not be eliminated, as they were inherent in the basic laws governing the National Guard. The principal trouble was the very short training period for the men, and as a few hours once a week would not suffice to replace the basic training which required us in the regular service on a twelve-hour day, very strenuous work, many months—six or more—in getting men ready for the war. Now it is only possible in a National Guard organization to get the men but once a week except for the two weeks in the summer. I did find it

Fig. 4.22c Bland, Marshall: Interviews, 255

Revised, Altered, and Collaborative Documents

4.23 Some editing projects may need to transcribe documents that have been altered or revised. Documents that have undergone heavy editing, either as part of the original composition process or after a period of time, require editors to decide how much of the revision should be made apparent to readers.

4.24 In an attempt to show the changes that occurred to bills during the process of legislative deliberation, the editors of the \textit{Documentary History of the First Federal Congress} printed documents with extensive notes to describe proposed amendments and alterations:

\textit{No appeal}

\textit{where the var}

\textit{shall any fact}

\textit{otherwise re}

\textit{On Septem}

\textit{insert the fol}

\textit{The lea}

\textit{the re}

\textit{Also on S}

\textit{On Ult}

\textit{On Septem}

\textit{8. Otis not}

4.25 Some of the changes might as well be compared with the original. To the end of 1787 original text of bill:}

\textit{(better)}

\textit{House}

\textit{might}

\textit{(This as D}

\textit{comp}

\textit{and o}

\textit{ture to t}

\textit{ate of bu}
ARTICLE THE ELEVENTH.

No appeal to the Supreme Court of the United States, shall be allowed, where the value in controversy shall not amount to one thousand dollars, nor shall any fact, triable by a Jury according to the course of the common law, be otherwise re-examinable, than according to the rules of common law. 16

16 On September 4, the Senate disagreed to a motion to strike out Article 11 and insert the following:

The Supreme Judicial Federal Court, shall have no jurisdiction of causes between citizens of different States, unless the matter in dispute, whether it concerns the real or personalty, be of the value of three thousand dollars, at the least: Nor shall the Federal Judicial Powers extend to any actions between citizens of different States, where the matter in dispute, whether it concerns the real or personalty is not of the value of fifteen hundred dollars, at the least—And no fact, triable by a Jury according to the course of the common law, shall be otherwise re-examinable, than according to the rules of common law.

Also on September 4, the Senate agreed to amend Article 11 to read as follows:

No fact, triable by a Jury according to the course of common law, shall be otherwise re-examinable in any court of the United States, than according to the rules of common law.

On September 9, the Senate struck out Article 11, after incorporating it into Article 8. Ons noted, "10th. and 11th incorporated."

4.25 Some authors revised documents years after their original composition. In his old age, James Madison significantly edited his papers but made many of the changes in writing that could be readily distinguished from that of the original. To note these changes, the editor of the Records of the Federal Convention of 1787 placed the revised portions in angled brackets and included the original text in footnotes:

of business in which he was to act, lamented his want of (better qualifications), & claimed the indulgence of the House towards the involuntary errors which his inexperience might occasion.

(The nomination came with particular grace from Penna., as Docr. Franklin alone could have been thought of (as a competitor). The Docr. was himself to have made the nomination (of General Washington, but the state of the weather and of his health confined him to his house.))

---

1 Crossed out "the requisites for it.
2 Crossed out "for the President ".
3 Crossed out "of the Genl. but the season of the rain did not permit him to venture to the Convention chamber."

---

22c Bland, Marshall: Interviews, 255

Documents

The documents that have been heavy editing, either as part of the daily routine, require editors to be alert to the subtleties of the language used. Some documents occur to bills during the Documentary History of the Senate, notes to describe process.

Fig. 4.24 De Pauw, Congress, 4:38

Fig. 4.25a Farrand, Convention, 1:4
Some documents are revised by both the author and third parties. The editors of the Papers of James Madison noted in a footnote both Madison's revisions and those made by his brother-in-law, John C. Payne:

"closed there. But the palpable urgency of the Ex. & its partizans to press war in proportion to the apparent chance of avoiding it, ought to open every eye to the hypocrisy which has hitherto deceived so many good"

2. JM originally wrote "hypocrisy & perfidy." Probably at a later time "& perfidy" was crossed out, and someone, probably John C. Payne, interlined "course" in place of "hypocrisy."

Fig. 4.25b Hutchinson, Madison, 17:142

4.26 Editors transcribe documents written by more than one author in such a way as to give both authors credit for their work. In a letter drafted by Dwight Eisenhower and edited by George Marshall, the final draft is used as the source for the version printed in the Papers of Dwight David Eisenhower, while the contribution of each author is identified in the footnotes:

"Marshall wrote the three preceding sentences (beginning "This weakness is now . . . . ") on Eisenhower's draft, replacing a sentence that read, "Since that time the enemy has not had to fear an attack involving a large scale sea campaign." The copy showing Marshall's handwritten corrections, given here and in the following notes, is in WPD 4639-31."

3 Marshall made several stylistic changes on Eisenhower's draft, as well as writing the two preceding sentences, replacing a sentence that read, "A fleet undertaking offensive operations through island areas must be preceded by land based aircraft firmly established on suitable fields from which a protective umbrella may be maintained." As the authors of the text of the West Point Atlas make clear, the Japanese fully recognized this point. Their pattern of advance was to select the objective, concentrate their land-based aircraft and wipe out the Allied air units there, and then land sufficient troops to overwhelm the local Allied ground force. They next made operational the airfield they had captured before moving on. "Japanese naval task forces seldom moved beyond the combat range of their supporting land-based planes" (Esposito, ed., West Point Atlas, II, map 128). Later in the war in the Pacific the Americans did, of course, make longer "leap-frogs," but by then they had a large enough superiority in aircraft carriers to cover the operation with fighter planes.

4 Marshall wrote this entire paragraph by hand.

5 In his #201 MacArthur had concluded: "If agreeable to you I would appreciate greatly the presentation of this view to the highest authority." In his draft Eisenhower had concluded: "I welcome and appreciate your views on these subjects and invariably submit them to the highest authority." Marshall penciled in the changes.

Fig. 4.26 Chandler, Eisenhower, 1:103

4.27 When a document is lost or destroyed, editors are faced with hard choices. The editors of the Papers of Thomas Jefferson have printed editions as their source, locating the original manuscripts whenever possible.

To Jane Franklin

Dear Sister,

I am highly pleased to hear from you. I always

4.28 In rare cases, reconstructed using drafts, the editors of Jefferson's The Life and Morals of the Prophet Muhammad and Biblical passages by which he cut out passages from the book and to verify its translation: Philosophy of Jesus," Princeton University Press, 1961.

4.29 Information on photographs, drawings, and historical evidence will be.

4.30 When documents are lost, editors devise solutions. When the Papers of Thomas Jefferson was edited, drawings along with
Original Missing or Destroyed

4.27 When a document has been copied and the original subsequently destroyed, editors are forced to rely on the copied version as their source text. The editors of the Papers of Benjamin Franklin, for example, used earlier printed editions as their source text for letters when the originals could not be located, indicating the circumstances and source text in the headnote:

To Jane Franklin⁴ ms not found; reprinted from Duane, Works, vi, 3.

Dear Sister,

Philadelphia, January 6, 1726–7

I am highly pleased with the account captain Freeman⁵ gives me of you. I always judged by your behaviour when a child that

Fig. 4.27 Labarre, Franklin, 1:100

4.28 In rare cases, missing or no longer extant documents may be reconstructed using drafts, outlines, or other sources of information. For example, the editors of Jefferson's Extracts from the Gospels: "The Philosophy of Jesus" and "The Life and Morals of Jesus" reconstructed Jefferson's vanished compendium of Biblical passages by using his notes and the original copies of the Bible from which he cut out passages. For more on the methods used to reconstruct the book and to verify its accuracy, consult the chapter "The Reconstruction of the Philosophy of Jesus," in Jefferson's Extracts from the Gospels (Princeton: Princeton University Press, 1983).

Other Types of Sources

4.29 Information can be acquired from sources other than verbal texts. Photographs, drawings, music, inscriptions, material objects, and maps all provide historical evidence but often defy easy transfer to the printed page.

4.30 When documents include both textual and non-textual elements, editors devise solutions for presenting them on the printed page. The editors of the Papers of Thomas A. Edison followed a policy of producing facsimiles of drawings along with the accompanying text.
In practice it may be found necessary to feed the paper drum by a separate magnet, which will take an intermittent work off from the Engine and perhaps add to the accuracy of the Unison of the two machine. It may be done in this manner. Fig 23

Fig 23

The editor of the Complete Works of Washington Irving: Journals and Notebooks separated drawings from text, indicating through footnotes in the text where drawings had appeared. Facsimiles of drawings appeared as illustrations in the middle of the volume:

385. All the writing on the page and also a small drawing of a profile (Figure 4) are in brown or black ink.

Figure 4. Small profile, mentioned on page 171, n. 385.

4.31 Editors present editorial cartoons through photoreproduction combined with enlargement or reduction if necessary. The editors of the Documentary History of the Ratification of the Constitution reproduced political cartoons as documents with appropriate annotation:

4.32 Maps mentioned in the text or in footnotes to the text are enclosed in the margins of the page.
necessary to feed the paper which will take an intermittent perhaps add to the accuracy of the It may be done in this

Fig. 4.30a Jenkins, Edison, 1:325

Fig. 4.30b Wright, Irving, 1:171, figure 4

Fig. 4.31 Jensen, Ratification, 15:566–567

4.32 Maps require special handling for inclusion in a documentary edition. The editors of the Expeditions of John Charles Frémont (Champaign: University of Illinois Press, 1970) produced a “Map Portfolio” as part of their edition. It contained five folded maps and a descriptive pamphlet, all held in a box placed in a binding identical to the preceding volumes. This format allowed readers to remove the facsimile maps and open them to their full size. The editors of the Atlas of the Lewis and Clark Expedition (Lincoln: University of Nebraska Press, 1983) created a separate oversized volume (14” × 20”) in which the large maps were printed as full-sized or reduced-size facsimiles.
The editors of the Documentary History of the Ratification of the Constitution included sheet music as a document:

Philadelphia, 4 July

Federal March

AS PERFORMED IN THE
GRAND PROCESSION
IN PHILADELPHIA THE 4TH OF JULY 1788
COMPOSED AND ADAPTED FOR THE
PIANO FORTE, VIOLIN OR GERMAN FLUTE
BY
ALEX. REINAGLE

1. Alexander Reinagle composed "a grand march" called the "Federal March" for the "Grand Federal Procession" that paraded in celebration of the Fourth of July in Philadelphia (Evans 21421). Reinagle's "Federal March" was played by a band which made up division XII of the procession. On 10, 12, 15, and 17 July the sheet music
4.34  

Three-dimensional artifacts may warrant reproduction in historical editions. Some of Thomas Edison’s inventions were presented in the Papers of Thomas A. Edison with a series of photographs from various angles:

[Newark or New York, Summer 1870]

And Franklin Pope, Production Model: Printing Telegraphy

M (historic photograph) (ext. 17 cm dia. x 17 cm), NIWOK, Cat. 541:4, 18, 30. See Doc. 54 tetwise. The machine frame is stamped “G. C.”

Fig. 4.34 Jenkins, Edison, 1:188
Conflated Documents

4.35 On occasion, editors may create new texts by combining together, or conflating, information contained in several documents. The editors of the Documentary History of the Ratification of the Constitution combined information from three different documents—the text of the draft constitution of August 6, the notes from the Journal of the Convention, and James Madison's notes of the debate—to reconstruct how the Constitutional Convention's debates altered the text of the draft constitution. The editors created an original synthetic document that contained information from the three sources, with the retained portions of the text of the original draft in roman type, portions deleted from the draft in strike-through type, and additions set in italics. The result was a document for which no original source text existed but that represented information about the transformation of the text of the U.S. Constitution:

We the People of the States of New-Hampshire, Massachusetts, Rhode-Island and Providence Plantations, Connecticut, New-York, New-Jersey, Pennsylvania, Delaware, Maryland, Virginia, North-Carolina, South Carolina, and Georgia, do ordain, declare and establish the following Constitution for the Government of Ourselves and our Posterity.

ARTICLE I. The stile of this Government shall be, "The United States of America."

II. The Government shall consist of supreme legislative, executive and judicial powers.

III. The legislative power shall be vested in a Congress, to consist of two separate and distinct bodies of men, a House of Representatives, and a Senate, each of which shall, in all cases, have a negative on the other. The Legislature shall meet at least once in every year; and such meeting shall be on the first Monday in December in every...
Chapter Five

Presenting the Text

Textual Issues

5.1 Editors who use expanded transcription must consider how they will present documents on the page, screen, or microform frame and decide how much they will emend documents to present them in a readable fashion. Most editors standardize the presentation of parts of their documents, creating datelines, placelines, signatures, salutations, or complimentary closings whose placement is consistent from document to document. Uniform presentation helps the reader locate important information such as the dates and recipients. It also helps editors avoid the difficulty and expense of reproducing the irregular physical layout of many documents, elements that are not important to most users and that are best studied in the original manuscripts. Within each document, editors also need to decide to what extent they will emend the text to facilitate its presentation. Some editors intervene only minimally, presenting a near-literal transcription of the text, while others standardize many parts of the document and make significant emendations to produce a text with a modern, standard appearance. The variety of possibilities within expanded transcription requires editors to explain to their readers exactly how they have standardized and emended the documents they publish.

Date and Placeline

5.2 If a document indicates date and place of composition, editors usually want to include this information in the published version. Most editors standardize the placement of this information regardless of where it appears in the document, although the exact format of this information varies among
editions. The editors of *Freedom: A Documentary History of Emancipation* standardized placement on a single line at the beginning of the document, but retained the original form:

207: Commander of the 2nd Division of the Army of Kentucky to the Commander of the Army of Kentucky, Enclosing an Order by the Former and a Letter to a Citizen of Ohio

near Lexington Ky. Dec 11th 1862
Sir I have never, until now, felt disposed to take notice of attacks

Fig. 5.2a Berlin, Freedom, ser.1.1:590

The editor of the *Letters of Eugene V. Debs* standardized both the placement and format of the date and place of composition. Each document's date was presented as a month spelled out, date, and year with the place appearing directly below it as a city and state. This information was placed one line above the salutation, regardless of its format or placement in the original document:

Katherine Metzel Debs to Grace D. Brewer

October 11, 1910
Terre Haute, Indiana

Dear Mrs. Brewer:
Your letter with enclosure is received. Yes, they are a long way

Fig. 5.2b Constantinople, Debs, 1:381

The editors of the *Documentary History of the First Federal Elections* placed the date and place of composition in a heading for each document. The year of composition did not appear in the heading but could be found in the running head at the top of each page:

To the German inhabitants of the State of Pennsylvania,
Philadelphia, 13 November

The character of the German (at least in Pennsylvania) has never been to demand privileges he is not intitled to. From natural diffidence, he rather steps

Fig. 5.2c Jensen, Elections, 1:339

The editor of *A Confederate Nurse: The Diary of Ada W. Bacot* placed the date above and to the right of each diary entry:
hope to awake in the morning to find the New Year coming in with a bright smile as the old has left us weeping.

Tuesday Jan 1st 1861

1861 yes little did I think four year ago, I would live to write the year 1861. But alive I am thank God, & far better than I was four years

Fig. 5.2d Berlin, Nurse, 23

The editor of the Diary of Elizabeth Koren placed the date in italic type on the first line of each new diary entry:

Monday, January 30. Just as I was about to drink my coffee this morning, Embret came very opportunely and gave

Fig. 5.2e Nelson, Koren, 151

5.3 When possible, editors supply datelines and placelines when the information is missing in the original document. When editors can provide this information with certainty, the supplied portion is placed in square brackets, as was done by the editors of Lydia Maria Child, Selected Letters:

Channcey St. Wednesday
[25 March 1857]

My Dearly Beloved David,

I am getting homesick already, and meditating how I can get through all

Fig. 5.3a Melitzer, Child, 308

When part of the supplied date was conjecture, the editors of the Letters of Jessie Benton Frémont followed that portion with a question mark:

Monday [Sept. 23, 1878]

Dear Nell,

I have been in bed three days—fatigue &c.—we go tomorrow and

Fig. 5.3b Herr, Frémont, 447

The editor of the Selected Letters of Charles Sumner placed editorially supplied information in italic type inside brackets. When supplied dates were conjectural, the editor marked them with a c. or ca. for circa. In this example, the
editor supplied both the date and place of composition, although only the date was conjecture:

[Washington]
[c. 24 January 1874]

Dear Mr Bolles,
I wonder if I deserve this.¹ The remarks I made were unprepared &

Fig. 5.3c Palmer, Summer, 2:628

Salutations and Closings

5.4 Most editors standardize the placement of salutations in letters. The editors of the Papers of John Adams employed a common method, placing all salutations at the top left of the body of the document, with no space between the salutation and the letter:

From Jonathan Trumbull

Gentlemen
Two accounts of loss by hostilities committed by the Ministerial

Lebanon 25th March 1776

The salutation was placed on the first line of the text by the editor of Dear General: Eisenhower’s Wartime Letters to Marshall:

8 LONDON

September 23, 1942

Dear General: Admiral Ramsay¹ brought to me a verbal message from you to the effect that you hoped I would not be compelled to make a visit

Fig. 5.4b Hobbs, Dear General, 43

The editors of the Documentary History of the Ratification of the Constitution did not print the common salutations of their documents (e.g. Dear Sir). They did print salutations that provide unusual information (e.g. Dear Friend):

Gov. Edmund Randolph to Meriwether Smith, Charles M. Thruston,
John H. Briggs, and Mann Page, Jr., Richmond, 10 December¹

Your favor of the second instant,² requesting permission to publish my letter on the new Constitution, gives me an opportunity of making

Fig. 5.4c Jensen, Ratification, 8:229
Editors generally standardize the placement of complimentary closings, regardless of their location in a document. The editors of the *Papers of Alexander Hamilton* ran together four closing lines in a single paragraph, separating the elements with space:

I am with great respect Sir your most humble and obedient Servant

John Foscin

Fig. 5.5a Syrett, *Hamilton*, 26:169

Other editions place the closing one line below the last line of the body of the letter, set off to the right, as done by the editors of the *Papers of Henry Laurens*:

by that Eternal God to whom all hearts are open & before whom the most Secret thoughts cannot be hid

I am Sir

Your most Obd¹ & most Hble Serv²

Alex³ Innes

Fig. 5.5b Hamer, *Laurens*, 10:332

The editors of the Documentary History of the Ratification of the Constitution did not print the standard complimentary closings of their documents:

particular, the Convention reciting the powers by which they were convened. I was exactly in time as the 1st of February was set down for taking up and entering the several ratifications, and I delivered ours before they began that business.

1. FC. Stevens Family Papers, NJHl. This letter, with minor differences in wording, is printed in Livingston Rutherfurd, *Family Records and Events: Com-

Fig. 5.5c Jensen, *Ratification*, 3:191

Editors transcribe authors' signatures in different ways. The editors of Abraham Lincoln's *Speeches and Writing* (New York: Library of America, 1989) omitted Lincoln's signature from his signed documents. The editors of the *Papers of Henry Laurens* transcribed signatures using upper- and lowercase letters:

Sir / Ye mo. obed¹ & very hble Serv²

P. Henry

Fig. 5.6a Hamer, *Laurens*, 14:529
The editors of the *Papers of James Madison* transcribed the same signature using large and small capital letters:

Sir Yr. Mo. obedt. & very hble Servt.

P. Henry

Fig. 5.6b Hutchinson, *Madison*, 1:267

If more than twelve signatures appeared at the end of a document, the editors of *Free at Last: A Documentary History of Slavery, Freedom, and the Civil War* did not print them but indicated the number of signers in brackets:

your faithful friends in all the perils and dangers which threaten our beloved country.

[59 signatures]

And many other colored citizens of Nashville

Fig. 5.6c Berlin, *Free*, 505

For documents with many signers, the editors of the *State Records of South Carolina: Journals of the House of Representatives* included the first signer and the number of additional signatures in the footnotes:

[The Petition of Sundry Inhabitants of the Village of Granby in the County of Lexington humbly sheweth That the present Road in the said County leading from Granby to Charleston was some time ago order'd to be laid off by the County Court of the said County and that shortly after the said Court was suspended and no Commissioners being appointed put it out of the power of your petitioners of Applying for the said road to be alter'd which is at present so bad as to render it almost impassable and is about four Miles further round your petitioners therefore pray that the old road may be open'd which is much nearer and easier kept in repair and that your Honble. House would take the Matter into Your Consideration and appoint Commissioners for the purposes aforesaid and your petitioners as in Duty bound will ever pray.]³⁶

³⁶Petitions, 1791, No. 107. The petition was signed by Alex. Blair and twenty-three other persons.

Fig. 5.6d Stevens, *House*: 1791, 388

The editors of *Free at Last: A Documentary History of Slavery, Freedom, and the Civil War* transcribed the mark (X) of illiterate men and women:

his

Archev X Vaughn

mark

Fig. 5.6e Berlin, *Free*, 113
Paragraphs

Editors decide how their editions will note irregular paragraphing, and many explain their procedure in their statement of editorial method. The editors of the Political Correspondence and Public Papers of Aaron Burr standardized the length of paragraph indentations but did not add indentations that did not appear in the original text:

Paragraph indentations are standardized to the customary five-character indentation. When an author has chosen not to indent the first line of a new paragraph, that format is respected.

The editors of Friends and Sisters: Letters between Lucy Stone and Antoinette Brown Blackwell interpreted the long dashes and elongated spaces of their subjects' letters as standard paragraphs:

Particularly in the early letters, the writers seemed reluctant to waste space, so they apparently used a long dash or an elongated space to separate the end of one paragraph and the beginning of the next on the same line. We have rendered these into standard paragraphs.

The editors of the Papers of Jefferson Davis modernized all paragraphing in their edition:

Standard paragraph indentation is used. It is commonplace in nineteenth-century manuscripts for the writer to end a paragraph in midline and return to the left margin without indentation. Whenever this occurs, extraneous punctuation, if any, is eliminated and the modern form of paragraphing is employed.

Capitalization

Editors usually retain the capitalization employed by the authors of documents, but a few editions standardize some elements of capitalization. When transcribing manuscript documents, it may be difficult to determine whether authors intended to use uppercase or lowercase letters, a situation made more difficult when transcribing seventeenth- and eighteenth-century manuscripts that include irregularly capitalized words. The editor of the
Complete Works of Washington Irving: Journals and Notebooks attempted to transcribe literally Irving's writings but explained the challenge posed by his penmanship and erratic use of capital letters:

The most frustrating part of transcribing Irving's handwriting stems from his whimsical and indiscriminate use of capital and lower-case letters, as much for common as for proper nouns, and occasionally for other parts of speech as well, so that the old-fashioned custom of capitalizing nouns does not explain Irving's peculiar use of capitals. His capital and lower-case formations of letters, particularly "a," "c," "g," "m," "n," "o," "s," "v," and "w" (and occasionally "d," "l," and "t") are often indistinguishable. Their size in relation to following or preceding comparable letters is seldom a clear index to whether a capital or lower-case form is intended, although the initial capitals "B," "F," "H," "I," "J," "K," and "Q," are seldom left in doubt. The conformation of the letter is often more helpful, but in letters like "a," "c," "m," "n," "o," "s," and "t," it is of little use because the formation of lower-case and capital letters is often nearly identical. Place names appear in varying forms, occasionally in successive lines and often on the same page. "Peter," "Paris," "French," and "Italian" are as likely to be lower-case as not. The initial "o" in "o'clock" (almost always without the apostrophe) appears in from pigmy to giant size and usually in the same conformation; whereas "Eve" occurs regularly and becomes routine form regardless of its location in the sentence. The initial flourish with which a capital is likely to begin is an unreliable index, for when it is used (especially in such a letter as "s"), it is used in varying forms – some long, some short, most of them in-between. It is the in-between cases, where the eye cannot decide, that the doubtful cases are resolved in Irving's favor, and the form that usage dictates is rendered. That is, doubtful words that are customarily capitalized are so transcribed, the only exception being those cases in which the letter is clearly a lower-case form – judged so by its distinctive conformation and by comparison with adjacent and comparable characters. In converse cases, where Irving seems to use capitals for normally lower-case words, he is given the benefit of the doubt by transcribing as capital only those letters which are clearly so – all intermediate forms being rendered lower-case.²

The editors of the Papers of Benjamin Franklin generally maintained Franklin's capitalization, but they chose to begin every sentence with a capital letter, regardless of the form of the original:

5.9"
b. **Capitalization.** Franklin himself was a life-long advocate of the use of initial capitals for all substantives, as he explained in a long letter to Noah Webster, December 29, 1789, and the editors, obedient to his views, will follow his practice in printing his manuscripts. We shall likewise retain all other capitalization as written, except that we shall begin every sentence with a capital letter. Some eighteenth-century writers, however, were so erratic and the size of their initial letters often varied so much in the same manuscript, that it is sometimes uncertain whether the writer meant to use a capital or a small letter, especially with such letters as "C" and "P." In these circumstances we shall render doubtful initial letters as if they were in the same manuscript or, such a guide failing, employ modern usage.

Fig. 5.8b Labaree, Franklin, 1:xxi

The editor of the *Correspondence of Roger Williams* retained original capitalization with the exception of proper nouns:

**Capitalization**

Capitalization is retained as found in the original manuscripts, except in the following instances. Personal names and geographical names have been capitalized. When it is not certain whether Williams or other writers intended to use a capital or lowercase letter, modern usage is followed.

Fig. 5.8c LaFantase, Williams, 1:xxi

**Interlineations, Marginalia, and Canceled Passages**

5.9 Editors treat interlineations in several ways. Many editors incorporate interlined passages silently into the text at the point indicated by the document’s author, as was done for most interlineations by the editors of the *Papers of Joseph Henry*:

- In a number of documents there are interlineations, canceled matter, variant texts, marginalia, and even footnotes by the original author. The first are silently brought into line unless there is some point in their position. In that event we generally use a footnote to elucidate the significance, retaining the original position only in exceptional cases.

Fig. 5.9a Reingold, Henry, 1:xxxvi
The editors of the *Journals of the Lewis and Clark Expedition* indicated interlinearations by inserting an editorial explanation set in italic type within brackets:

Thursday 29th Rained last night a violent wind from the N this morng with rain, Some hail we have a trial of John Shields, John Cotter & R Frasure which take up the greater part of the day, in the evening we [we written over I] walk to Higgens a blustering day all day, the blacksmiths return with part of their work finished, river Continue to rise, Cloudy Day

The editors of the *Papers of Zebulon Baird Vance* placed interlinearations between diagonal lines (\(\uparrow\)) at the point of insertion:

upon ourselves these ties of unity & common fellowship? With men who have slaughtered our sons & brothers in battle, murdered our citizens in cold blood, burnt our homes into cinders, stolen our property and inflicted upon our mothers /sisters/ and daughters the crowning outrage of humanity, and now send against

The editors of *Mark Twain’s Letters* inserted interlinearations into the text, marking single-word insertions with a single caret (\(^\uparrow\)) (not shown) and putting multiple-word interlinearations between two carets:

hold him in grateful remembrance if he did that which made it possible for you & me to—become all in all to each other. \(^4\) Oh, Livy! The clock has just struck 3! Another night without sleep! I am terrified. With kisses & blessings, good-bye my own darling.

Interlinearations may be placed at the point of insertion between up and down arrows (\(\uparrow\)), as did the editors of the *Autobiography of Benjamin Franklin: A Genetic Text*:

the Assembly; your equivocal Project would be just a Match for their Wheat or other Grain.

The editors of the *Papers of General Nathanael Greene* incorporated marginalia and other nonlinear passages into the body of the text. The editors of the *American Schoolmaster* indicated marginalia by placing arrows (\(\uparrow\) margina|al\(\downarrow\)):

scatter’d Counsels the greater Impression. This was copied in all the Nos.

The editor of the *Diary of the Neame Family* placed marginalia in his margins, such as the note below:

Mr. Noyes pray’d at the margin, Va, Væ, Væ, Væ, Væ, Væ,

With this relay chemically prepared perfectly; The soot and our discontent makes all the clouds that per minute but come.

"Marginalia written on the margin."

Editors establish points where passages that have been cut or added are significant, such as the note below:

I beg to be kind to Relations; [and am]
k Expedition indicated interlinealic type within brackets:

dent wind from the N this of John Shields. John Colter's the day, in the evening we ring day all day, the blacked, river Continue to rise.

9b Moulton, Lewis and Clark, 2:182

placed interlineations between

Iowship! With men who have ved our citizens in cold blood, and inflicted upon our mothers manity, and now send against

Fig. 5.9c Johnston, Vance, 2:244

notations into the text, mark-
(‘) (not shown) and putting

at which made it possible er.,“ Oh, Livy! The clock esp! I am terrified. With

Fig. 5.9d Branch, Twain, 3:154

ition between up and down shy of Benjamin Franklin: A

would be ↑just↓ a

may, Franklin: Genetic Text, 115

a document) is transcribed ipers of General Nathanael passages into the body of

5.11 Editors establish policies for determining if and how they will transcribe passages that have been canceled out or erased from original documents. The editor of the Papers of George Mason did not transcribe most canceled passages but included significant crossed-through passages in the notes:

I beg to be kindly remembered to Mrs. Mercer & my Young Relations; [and] am Dr Sir Yr affecte. Kinsman & obdt. Servt. G Mason

GM first wrote “my little Cousin,” but crossed it through and settled for MY YOUNG RELATIONS (Mercer and GM were first cousins).

Fig. 5.11a Ruland, Mason, 1:427
The editors of the *Papers of William Penn* indicated canceled passages with strikethrough type:

lead not out of the Sence & Unity of that living Word of God, For that is the defiled Bed & Dishonorable Marriage, where the Love & Affection that goe forth from either Male or Female, to each other, are not honoured with the unity, Blessings, & Holy Leadings of the Spirit, & Word of God in the Heart, & not preserved in the undefiled Life; So that this, the Lord God eternal requires at the hand of every Male & Female, that are convinced of his pure Light, & Way, &

Fig. 5.11b Durham, Penn, 1:233

The editors of the *Letters of Jessie Benton Frémont* transcribed canceled passages as italic type within angled brackets:

Mr. Pryor re-married a Miss Gray within *<a very short time>*—as soon as a *<month afterwards>*. Legislature in session granted divorce at once. woman housekeeper.

Fig. 5.11c Herr, Frémont, 1:16

The editors of the *Marcus Garvey and Universal Negro Improvement Association Papers* marked erased passages with the bracketed word *erasure*:

**WORLD, I do not see anything therein upon which this Department can base any action whatever. There is, [erasure], in a democratic form of government like ours the right to agitate to influence governmental policy, and so far as**

Fig. 5.11d Hill, Garvey, 1:389

The editors of the *Papers of Alexander Hamilton* used several different procedures for canceled passages:

**IV. CROSSED-OUT MATERIAL IN MANUSCRIPTS**

Words or sentences crossed out by a writer in a manuscript have been handled in one of the three following ways:

1. They have been ignored, and the document or letter has been printed in its final version.
2. Crossed-out words and insertions for the crossed-out words have been described in the notes.
3. When the significance of a manuscript seems to warrant it, the crossed-out words have been retained, and the document has been printed as it was written.

Fig. 5.11e Syrett, Hamilton, 26.xvi
that living Word of God, For -Marrige, where the Love & -ale or Female, to each other, -go, & Holy Lendings of the -ot preserved in the undeleted requires at the hand of every -his pure Light, & Way, &

Fig. 5.11b Dunn, Penn, 1233

ont transcribed canceled passages

sin (a very short time)—as in session granted divorce

Fig. 5.11c Herr, Fremont, 116

al Negro Improvement Association ted word erasure:

which this Department can base democratic form of government lenmental policy, and so far as

Fig. 5.11d Hill, Garvey, 1:389

n used several different proce-

IN MANUSCRIPTS

iter in a manuscript have ways:

ument or letter has been the crossed-out words seems to warrant it, the, and the document has

Fig. 5.11e Syrett, Hamilton, 20:xvi

Superscripts and Abbreviations

5.12 When documents contain superscript letters, editors may either transcribe them literally or bring them down to the line. The editors of the Papers of Henry Laurens literally reproduced the superscript letters found in the complimentary closing of this letter from Patrick Henry to Henry Laurens, while the editors of the Papers of James Madison brought them down to the line:

I beg to be presented to Congress in the most acceptable manner & in Terms expressive of that high Regard with which I have the Honour to be

Sir / Yr. mo. obedt. & very hble Servt.

P. Henry

Fig. 5.12a Hamer, Laurens, 14:529

I beg to be presented to Congress in the most acceptable manner & in Terms expressive of that high Regard with which I have the Honour to be

Sir Yr. Mo. obedt. & very hble Servt.

P. Henry

Fig. 5.12b Hutchinson, Madison, 1:267

5.13 When documents contain abbreviations and contractions, editors may transcribe them literally or expand them into full words to make them more easily understood. The editors of the Diary of Martha Ballard literally transcribed all abbreviations and contractions, while the editor of the version of the diary found in the History of Augusta expanded most abbreviations into the full words they represented:

Deacon Coney made ṣ 1st prayer, Esq' Petangil ṣ Last foren'. Col' Sewal ṣ 1st aft & red Doc' Dodridge's Discourses, Mr Liverm' Last prayer, but few attended.

Fig. 5.13a McCausland, Ballard, 176

Deacon Coney made the 1st prayer; Esqr. Petengil the last, forenoon; Colonel Sewall the 1st, afternoon, and red Doctor Dodridge's Discourses; Mr. Livermore, last prayer; but few attended.

Fig. 5.13b Nash, Augusta, 296
The editors of the Papers of Ulysses S. Grant lowered superscript letters used in abbreviations but otherwise transcribed abbreviations as they were written:

The undersigned having been duly appointed Colonel of the 7th Congl Dist Regt. of Ills Volts. Militia by order of Govr Richard Yates, duly promulgated hereby assumes command.

Fig. 5.13c Simon, Grant, 2:45

Abbreviations were transcribed literally by the editor of A Plantation Mistress on the Eve of the Civil War: The Diary of Keziah Goodwyn Hopkins Brevard, although missing letters in brackets were supplied if an abbreviation was unclear:

$13.35 cts. in my purse after paying D[ick], N[ed] & John. Paid this (87½ cts.) for lard (Ned) leaf lard. This is right, 7 lb. of [leaf] lard at 12½ cts. per lb.

Fig. 5.13d Moore, Plantation Mistress, 59

The editors of the Correspondence and Miscellaneous Papers of Benjamin Henry Latrobe noted in their statement of editorial method that they expanded abbreviations that they regarded as difficult to understand:

Ampersands have been changed to “and” except in the names of business firms and in the abbreviation “&c.” Superscript letters have been lowered to the line of type, and the resulting abbreviation silently expanded if it is not easily recognizable (e.g., while “Pha.” has been rendered “Philadelphia,” “Phila.” has been let stand).

Fig. 5.13e Van Horne, Latrobe, 1:xii

The editors of the Papers of Thomas Jefferson altered the abbreviated portions of their texts significantly, expanding all abbreviations into full words:

As for contractions and abbreviations, these will normally be expanded in ordinary documents, and raised letters at the close of such contractions or abbreviations will in all cases be lowered. In many instances, however, particularly in such a document as Jefferson’s notes and outlines of his argument in behalf of religious freedom in 1776, the text will be presented quite literally even though many of the contractions present a difficult problem of decipherment. A single example of the kind of expansion that the editors have in mind in adopting this conventionalized practice will suffice. If presented literally, what Jefferson wrote as he took down hasty notes in a congressional investigating committee in

the busy summer

lard, yet we were

condol to wch sm

having heard that that, was] so ter

had he not been

to which small pre

presents the kind of

would have used

in print. It makes

esting of Jefferson’s

be expanded, ex-

literal presentation

ought the ampu

the form “&c.”

eighteenth-century

object to the atten-

ployed in the case
given with liter-

expanded are the

comme (com-

plaint, plaintiff), and

when employed

5.14  Editors decline to reproduce the sections and styles of typography used in Documentary History, why they did not reproduce.
lowered superscript letters used in
vocations as they were written:

ly appointed Colonel of the
Miltia by order of Govr
reby assumes command.

Fig. 5.13c Simon, Grant, 2:45

: editor of A Plantation Mistress on
vodwyn Hopkins Brevard, although
abbreviation was unclear:

of lard. This is right, 7

: j. 5.13d Moore, Plantation Mistress, 39

incus Papers of Benjamin Henry
thod that they expanded abbre-
stand:

and” except in the names of
ce.” Superscript letters have
resulting abbreviation silently
(e.g., while “Pha.” has been
let stand).

Fig. 5.13c Van Horne, Latrobe, 1:xxii

red the abbreviated portions of
ons into full words:

these will normally be
sed letters at the close of
in all cases be lowered.

: in such a document as
ent in behalf of religious
nted quite literally even
a difficult problem of
nd of expansion that the
convensionalized practice
fferson wrote as he took
estigating committee in

the busy summer of 1776 would read as follows: “Carleton havg
hrd yt we were returning with considble reinfrnt, so terrifd, yt
wd hve retird immedly h d. nt bn infmd by spies of deplorble
condn to wch sm pocx hd redds.” Such a passage, by the con
ventionalization to be followed in these volumes, will read: “Carleton
having heard that we were returning with considerable rein-
forcement, [was] so terrifed, that [he] would have retired immediately
had he not been informed by spies of [the] deplorable con-
ditions to which small pox had reduced us.” This expanded text rep-
resents the kind of clear and readable form that Jefferson himself
would have used for a document intended for formal presentation
in print. It makes for clarity and readability and yet sacrifices no-
thing of Jefferson’s words or meaning. Among the typical forms to
be expanded, except in those outstanding documents in which a
literal presentation is adopted, are the ampersand and the thorn,
though the ampersand will be retained in the names of firms,
the form “&c.” will be employed since it was widely used in
eighteenth-century printing, and the thorn, much as the editors
object to the attempt to represent it by the letter “y,” will be em-
ployed in the case of a few highly important documents which are
given with literal exactness. Among other typical forms to be
expanded are the following: agst (against), cos (companies),
commee (committee), dft (defendant), execs (executors), plt
(plaintiff), and pr (per; though the symbol @ will be retained
when employed for per, pre, and pro).

Fig. 5.13f Boyd, Jefferson, 1:xxx

Typefaces

5.14 Editors decide how literally they wish to represent the different sizes
and styles of typefaces used in original documents. In the first volume of the
Documentary History of the Ratification of the Constitution, the editors explained
why they did not reproduce the typefaces of original documents:

Reproduction of Newspaper, Pamphlet, and Broadside Material

Eighteenth century printers sometimes used several varieties of type
in a single item—large capitals, small capitals, and italics, as well
as ordinary type. No attempt is made to reproduce varieties of type
except when capital letters and italics were evidently used for emphasis
by the author or the printer. In a few cases we have reproduced, so
far as possible, the format of newspaper items.

Fig. 5.14a Jensen, Ratification, 3:19
The editors of the *Papers of Benjamin Franklin* described their method for transcribing the various typefaces of printed documents:

b. Proper nouns, such as the names of persons and places, which were conventionally printed in italics in the eighteenth century, will be printed in roman type.

c. Prefaces to pamphlets or other publications and passages of substantial length in other pieces were often printed in italics for typographical reasons that were cogent in the eighteenth century but no longer seem persuasive. These also will appear in roman type. Italics will be preserved, however, for words or phrases of emphasis and in other special cases as, for example, in some instances of dialogue or conversation.

d. Single words originally printed in full capitals for emphasis or other similar reasons will usually be printed in small capitals in harmony with modern typographical taste.

---

*Fig. 5.14b Laharee, Franklin, 1:xi-xl*  

**Underlined Passages**

5.15 Editors transcribe underlined passages by representing them literally with underlined type or, more often, by representing them with italics. The editors of the *Papers of Thomas A. Edison* transcribed underlined passages literally with underscored type and indicated double-underlined passages, such as the words *sixty-thousand* in the example below, by printing them in underscored type and then indicating the second underlining in the textual note marked with a superscript letter:

A girl of ordinary intelligence and a few hours or days of practice, works either machine with perfect accuracy.

We found that where a single wire was tumbling messages into the office at the rate of *sixty thousand* words per hour, that we must have something more rapid than the hand-pen

\[\text{Underlined twice.}\]

*Fig. 5.15a Jenkins, Edison, 1:251-252*

The editors of *Lafayette in the Age of the American Revolution* presented underlined passages in italic type:

Before settling any thing the french generals want to hear from theyr Second division. *Don't fear By any means* theyr acting *Rashly* and Be assur'd that you may very far depend on theyr *Caution*.\(^6\) But our

*Fig. 5.15b Idzerda, Lafayette, 3:114*
The editor of the Selected Letters of Charles Sumner placed underlined passages in italics but indicated double-underlined passages by presenting them in italicized, underscored type:

This proposition, if passed, to a vote, will split our party—as no other party ever has been split. *Seward will not vote for it*. He will use it as a lure; *but not vote for it*. I deplore its introduction, & think of it constantly with tears. My constant hope is that events in their rapid

single-underlined passages were placed in italics (not shown), while double-underlined passages were placed in small capital letters by the editors of the *Salmon P. Chase Papers*:

those positions of the Slave Power which are most exposed. Our first and chief effort should be directed to the *denationalization of slavery*—to the absolute and complete divorce of the General

Slips of the Pen and Other Errors

5.16 Editors need to establish policies on how to treat writing, typing, or printing errors. The editors of the *Samuel Gompers Papers* silently corrected many of the errors created by typesetters and typists (see Section 4.13), as did the editors of the *Frederick Douglass Papers*:

Obvious printer’s errors have been corrected. The most common of these were the transposing of letters within a word, the repetition of words and punctuation marks, and transposition of numbers within dates. Misspelled names have been treated as printer’s errors when it was apparently impossible for anyone to pronounce them as transcribed or when the correct spelling appeared elsewhere in the same document.

The editors of the *Papers of Woodrow Wilson* printed all documents exactly as they appeared, without correcting errors, because of the insights such errors could provide:

We think that it is very important for several reasons to follow the rule of *verbatim et literatim*. Most important, a document has its own integrity and power, particularly when it is not written in perfect literary form. There is something very moving in seeing a
Texas dirt farmer struggling to express his feelings in words, or a semiliterate former slave doing the same thing. Second, in Wilson's case it is crucially important to reproduce his errors in letters which he typed himself, since he usually typed badly when he was in an agitated state. Third, since style is the essence of the person, we would never correct grammar or make tenses consistent, as one correspondent has urged us to do. Fourth, we think that it is very important that we print exact transcripts of Charles L. Swem's copies of Wilson's letters. Swem made many mistakes (we correct them in footnotes from a reading of his shorthand books), and Wilson let them pass. We thus have to assume that Wilson did not read his letters before signing them, and this, we think, is a significant fact. Finally, printing typed letters and documents verbatim et literatim tells us a great deal about the educational level of the stenographic profession in the United States during Wilson's time.

Fig. 5.16b Link. Wilson, 51:xx

The editor of the Correspondence of Mother Jones marked slips of the pen using the word sic in brackets:

would have to suffer the result. I want to say right here, I may never see him again, but one thing one thing [sic] certain, I will fight to death

Fig. 5.16c Steel, Jones, 52

5.17 If an edition has erratic or missing punctuation, an editor may literally transcribe those passages to maintain the character of the original document or emend the passages to produce a document that may be easier to read. Following the revision of the editorial policy in volume 10, the editors of the Papers of Henry Laurens retained all original punctuation. The editors of Witness to the Young Republic: A Yankee's Journal emended some forms of punctuation to make the documents easier to read:

Our intent throughout has been to present the text as French wrote it; the words are all his except for the few that have been bracketed. In the interests of consistency and clarity, however, we have made a few changes in punctuation and spelling. Some of the commas and dashes, which French used much too liberally, have been removed or replaced silently by other marks of punctuation. To avoid incomplete or run-on sentences we have occasionally created new ones, and sentences without periods have had them added, in each case without mention. Otherwise, when marks of punctuation have been added, brackets have been used. The punctuation of dates and addresses has been regularized silently, and the datelines at the start of each entry have been made uniform in style. The place at which an entry was written is cited only when it differs from the previous entry.

Fig. 5.17a Cole, Witness, xi-xii

Misleading punctuation can confuse dates, and after about a comma, and superfluous parenthesis omitted are dashed. The other superfluous parenthesis omitted is silently supplied. The periods used to enclose paragraph headings are used to avoid confusion with dates.

5.18 Editors must be careful with documents. The editor of the Diary original text quite significantly.

Der Mother Jones, I write to inform you that were able to have a telegram and you would have had the news about Christmas. W Va til after the Holler and for you as usual.

The editors of the journals were able to:

went to the post office and got the Bank where I attempted to devoured the place.
The editor of the Diary of Isaac Bachus altered the erratic punctuation of the original text quite significantly:

Misleading punctuation is corrected where the meaning is not in doubt. Punctuation is regularized between the members of a series, in dates, and after abbreviations. A period used as a comma is replaced by a comma, and superfluous interior periods are omitted. Also omitted are dashes used to fill out manuscript lines and dashes or other superfluous marks at the ends of sentences. A quotation mark or parenthesis omitted at the end of a quoted or parenthetical passage is silently supplied if there is no question as to its placement. Brackets used to enclose parenthetical phrases are replaced by parentheses to avoid confusion with the brackets that enclose editorial insertions.

Fig. 5.17b McLoughlin, Bachus, i.xxxvi

3.18 Editors must decide how to present words misspelled by the authors of documents. The editor of the Correspondence of Mother Jones retained all misspellings:

Reynoldsville Pa
Dec the 9 1901

Der Mother Jones
I write to inform you that I had hoped by this tim you would have been able to have made your way to Reynoldsville according to both telegrams and letters. I made arrangements for you to speak and you would have had a rousing meeting but unfortenly you was Debared from coming. now Mother Jones you will not be able to do much until after the Holladays and I hope you will see your way clear to com here about Christmas time if so let me know. I don think I will return to W Va til after new years. Com if you can as these People is anishtius for you ashiphely the textile workers

Fig. 5.18a Steel, Jones, 18

The editors of the Journals of the Lewis and Clark Expedition retained the authors’ original spelling, clarifying confusing orthography in brackets:

I took Collins & went to the place he found a Hog Skined & Hung up, the Crows had devoured the meet, Killed Pray fowl and went across a Pray to a 2nd Bank where I discovered an Indian Fortification, near the Second bank I attempted to cross a Bond [pond] of about 400 yds wide on the Ice &

Fig. 5.18b Moulton, Lewis and Clark, 2:153
All spelling (and many other features) was normalized by the editor of *Children of Pride*, as explained in the volume’s introduction:

The Joneses and their friends wrote extraordinarily careful English, but for ease of reading it has seemed wise to reduce a varying text to a consistent standard. To this end I have normalized all spelling, capitalization, punctuation, and paragraphing.

Fig. 5.18c Myers, *Children*, xiv

**Nontextual Elements**

5.19 Editors decide how to handle symbols, drawings, and other unusual elements that appear in documents. The editors of the *Joseph Henry Papers* reproduced drawings, doodles, and sketches found in Henry’s documents as facsimiles (see Section 4.10). The editors of the *Documentary History of the First Federal Congress* reproduced some elements, such as the hand symbol used by William Maclay in his diary:

> favour of Titles from these Motives, but that in conformity to the Practice of the other House, for the present they resolv’d to address the President without Title
>

(permit Ye. Meh. Yesterday G(enera)l M. (Muhlenberg) accosted me with the wish, Your highness of the Senate. on my passing he said Wynkoop had

Fig. 5.19a De Pauw, Congress, 9:37

The editors of the *Journals of the Lewis and Clark Expedition* represented the sun, moon, and star symbols with typographical equivalents:

> Note—The Longitude of the mouth of the River Dubois was calculated from four sets of observations of the ☯ & ☩, in which the ☯ was twice West, and twice East; two sets with Aldebaran, ☆ East in one, and W. in the other; and one set with Spica, ☆☆ East. the Longt. above stated is

Fig. 5.19b Moulton, *Lewis and Clark*, 2:228

Flourishes and decorative marks were not transcribed by the editors of the *Documentary History of the Supreme Court of the United States*:

> We delete from our transcription all textual marks that are of strictly decorative or design function. Thus, end-of-line dashes used to justify texts to the right margin are deleted as unnecessary since our typeset lines do not break at the same point as the document lines. Dashes employed in the same fashion within braces are not reproduced either. Any dash, however, that might serve

5.20 Because some purpose was to reproduce the style of handwriting in the original documents as faithfully as possible, the style of handwriting in the original documents is reproduced as faithfully as possible. The experience of a typeset edition is as different as the word appearance of the original documents.

The editors of the *La Salle* have had 20 years in high Life.

Fig. 5.19d Bache, *La Salle*, 5:422

But I must not

If a sentence of 10 words I must write:

Give my love to Andrew Johnson.

Plea...
s normalized by the editor of *Children*.

Ordinary care in English, but for a varying text to a consistent spellling, capitalization, punctuation.

Fig. 5.18c Myers, *Children*, xiv

elements

abol, drawings, and other unusual editors of the *Joseph Henry Papers* has found in Henry's documents as the *Documentary History of the Firsts*, such as the hand symbol used by

at in conformity to the Practice of d to address the President without for.

4. (Muhlenberg) accosted me with my pausing he said Wynkoop had

Fig. 5.19a De Pauw, *Congress*, 9:37

d Clark Expedition represented the nical equivalents:

the River DuBois was calculated & D, in which the ○ was twice paran, ☆ East in one, and W. in East. the Longtd. above stated is

. 5.19b Moulton, *Lewis and Clark*, 2:228 

transcribed by the editors of the United States:

nal marks that are of strictly dec-

dashes used to justify texts to the our typeset lines do not break hes employed in the same fashion dash, however, that might serve

some purpose as punctuation is retained. Elaborately self-conscious changes in the style of handwriting—textual enlargement, ornate capital letters—are not reproduced as such; nor are idiosyncratic decorative or aesthetic marks whether meant to balance or serve some other purpose inherent in the layout of the original document page. To reproduce such idiosyncratic embellishment would be technologically difficult and prohibitively expensive. We make this decision fully aware that the meaning of a document inheres as much in its appearance as in the words it contains. But the meaning that derives from the physical appearance of a document can be discerned only by examining it in person. The experience of personal contact with a document cannot be duplicated in a typeset edition.

Fig. 5.19c Marcus, *Supreme Court*, vol.1, 1:lvii–lvi

Terminal Punctuation

5.20 Because authors end sentences in a variety of ways, editors need to decide how to indicate terminal punctuation. The editors of *Jonathan Carpenter*’s *Journal* literally transcribed the long dashes and extra spaces Carpenter used to punctuate and terminate his sentences:

Declt ye 10 Last night 5 of our men made their escape joy go with ’em They tell us we shall soon be exchanged but I supose they lie as they used to do—our Money is Reduced to 1s pr week we have had 2 heretofore ever since I have bin in this troublesome scene in high Life below stairs—

Fig. 5.20a Herwig, *Carpenter*, 51

The editors of the Book of *Abigail and John: Selected Letters of the Adams Family* literally transcribed long dashes, even when they followed a period as terminal punctuation:

But I must bid you adieu or the post will go of without my Letter.—

Fig. 5.20b Butterfield, *Abigail and John*, 191

If a sentence did not have terminal punctuation, the editors of the *Papers of Andrew Johnson* inserted the appropriate punctuation in brackets:

Give my best respects to Messrs Browns & Dossier, tell dossier he must write to me and send me all the news[.] Please accept for your Self my best wishes[.] A. Johnson

Fig. 5.20c Graf, *Johnson*, 1:34
The editors of the Naval War of 1812: A Documentary History changed long dashes into modern forms of punctuation:

The punctuation used is that of the original document, with one exception. In the early 19th century, dashes were often used in place of periods, commas, semi-colons, and question marks; most of these dashes will be replaced by appropriate modern punctuation.

Fig. 5.20c Dudley, 1812, I.xii

**Interpolation and Missing Text**

**5.21** In cases where words or passages are missing from a document, editors sometimes use their knowledge of the topic and the surrounding words to interpolate the meaning of the missing portions. The editors of the Papers of James Madison printed a letter with several pieces missing, supplying the lost portions in brackets; information about which the editors were uncertain is followed by question marks:

[adverse to?] the American Cause, Independ[ent] of the Recommendation* [to detain the?] person and papers of Govr. Eden and to form a New Gove[rn]mt. [representative?] of the people, they have re-

Fig. 5.21a Hutchinson, Madison, I:182

When the editors of the Papers of Alexander Hamilton supplied missing or indecipherable words or letters, they placed them in angled brackets (< >):

The answer to this is, that however true it may be, that th(e)
right of the Legislature to declare wa(t) includes the right of judging whether the N(ation) be under obligations to make War or not

Fig. 5.21b Syrett, Hamilton, 15:40

The editors of Mark Twain's Letters provided their interpolation as part of the document's text, without brackets, but noted the reconstruction and interpolations of missing information at the back of the volume:

- 12 September 1861 · To Mary E. (Mollie) Clemens · Carson City, Nev. Terr. · UCCL 02715

**Copy-text:** MS, Jean Webster McKinney Family Papers, Vassar College Library (NPV). Only the last leaf of the MS survives. As the illustration shows, a piece has been torn out of the leaf at the upper right corner, affecting seven lines; the missing text has been conjecturally supplied by emendation. Accompanying the MS facsimile is a type facsimile of the same lines with the emended readings
original document, with ones were often used in place of esion marks; most of these modern punctuation.

Fig. 5.20d Dudley, 1812, 1:33

Text

missing from a document, editors and the surrounding words to ons. The editors of the Papers of pieces missing, supplying the lost which the editors were uncertain is

end[ent] of the Recomm-
i of Govr. Eden and to form if the people, they have re-
fig. 5.21a Hutchinson, Madison, 1:182

milton supplied missing or inde-
angled brackets (<>):

true it may be, that th(e) includes the right of judg-
atations to make War or not

Fig. 5.21b Syrett, Hamilton, 15:40

heir interpolation as part of the e reconstruction and interpola-
tion:

c) Clemens · Carson City,

ly Papers, Vassar College Li-
s. As the illustration shows, a it corner, affecting seven lines; y emendation. Accompanying ces with the emended readings

5.22 If a source text is defaced or damaged, an editor may want to indicate this information to the reader. The editors of the Collected Works of Abraham Lincoln marked missing words or letters with dots in brackets ([...]) and indicated the extent of the missing portion in the footnotes. The editors of the Papers of Ulysses S. Grant marked lost letters and words with dots and dashes in brackets ([...]) or [- - -]), with the number of such marks approximating the number of lost letters or words. Two sets of approximately four lost letters are indicated by the bracketed dots in the following passage:

I would have been able to call you Wife. Dearest Julia if you have been just as constant in your love it shall not [...] long until I will be entitled to call you by the [...] affectionate title. You

Fig. 5.22a Simon, Grant, 1:113

The editor of the Papers of Josiah Bartlett marked missing portions of documents with a set of empty brackets ([ ]):

Cold that I took on my Journey. The Small Pox is in the City. Some of the members of the Congress are now under Innocation & some have taken [...] as hitherto to Escape it. Which I Shall Do I am not fully

Fig. 5.22b Meviers, Bartlett, 18
The editors of the Salmon P. Chase Papers indicated missing portions of documents with the word torn italicized and in brackets:

\[
\text{instant destruction. All [torn] things I wished you to see and enjoy with me. And besides I wanted [torn] society and to feel your heart.}
\]

Fig. 5.22c Niven, Chase, 2:231

5.23 When editors transcribe documents, they often find words or passages that are illegible. The editor of the Correspondence of Mother Jones indicated illegible passages in brackets (i.e. [three words illegible]):

I am scratched this off in a hurry. It goes out underground. I am watched on all sides of my room. My old [three words illegible] this off. I am writing it blindly for I have to watch the windo I'll take no

Fig. 5.23a Steel, Jones, 109

The editors of the Letters of Delegates to Congress described the system used in the edition to signify missing or illegible words and numbers. Three bracketed ellipses points [...] were used for missing words within sentences and four bracketed ellipses points for words missing from the end of sentences:

The following devices will be used in this work to clarify the text.

\[
[\ldots],[\ldots].
\]

One or two words missing and not conjecturable.

\[
[\ldots],[\ldots].
\]

More than two words missing; subjoined footnote estimates amount of material missing.

\[
[\ ]
\]

Number or part of a number missing or illegible.

\[
[\ ]
\]

Blank space in manuscript; explanation in subjoined footnote.

\[
[\text{roman}]
\]

Conjectural reading for missing or illegible matter; question mark inserted if reading is doubtful.

Fig. 5.23b Smith, Delegates, 1:xii

Text with Original Annotation

5.24 When a document contains footnotes, editors must devise a system to prevent the reader from confusing those notes written by the document's author with those written by the editor. The editors of John Franklin Jameson and the Development of Humanistic Scholarship in America distinguished Jameson's footnotes (top) from the editor's (bottom) by placing the number of the editor's notes in braces:

1. Lecture before the American Historical Association, September 12, 1912.


The editors of the Bates letters with a series of notes for and against the problem of fine in favor of the body. Tillich concludes that "man's finitude, existential nonbeing..."

* Tillich, ST, I, 185.
† Tillich, ST, I, 186.
‡ Tillich, ST, I, 186.
§ Tillich, ST, I, 186.

46. Roberts, "The problem of the man who sought to understand God, but it affirms in some respects and perverts the idea that man is given by God alone..."

47. Tillich, Systematics, the problem of fine...
I wished you to see and enjoy society and to feel your heart:

Fig. 5.22c Niven, Chase, 2:23)

They often find words or passages (ynodence of Mother Jones indicated rds illegible):

It goes out underground. I am old (three words illegible) this to watch the wind I’ll take no

Fig. 5.23a Steel, Jones, 109

gress described the system used in rds and numbers. Three bracketed words within sentences and four on the end of sentences:

I this work to clarify the text.

issing and not conjecturable. s missing; subjoined footnote material missing.

umber missing or illegible. uscript; explanation in sub-

for missing or illegible material if reading is doubtful.

Fig. 5.23b Smith, Delegates, Lxii

otation

etors must devise a system to notes written by the document’s editors of John Franklin Jameson (ship in America distinguished atom) by placing the number of

1. Lecture before the Trustees of the Carnegie Institution of Washington, December 12, 1912.

(3.) Edward Waterman Townsend (1855–1942) created the character of Chimmie Fadden in a series of stories that first appeared in the New York Sun and were collected and published in 1895 as Chimmie Fadden, Major Max and Other Stories and Chimmie Fadden Explains, Major Max Expounds.

Fig. 5.24a Rothberg, Jameson, 1:229,304

The editors of the Papers of Martin Luther King, Jr. replaced King’s numbered footnotes with a series of symbols and numbered the editor’s notes sequentially:

nal life. And this is why Christianity rejects the doctrine of natural immortality in favor of the belief that eternal life is given by God alone. 46

Tillich concludes that the dialectical problem of nonbeing is inescapable. It is a problem of finitude. Finitude involves a mixture of being and nonbeing. 47 “Man’s finitude, or creaturiness, is unintelligible without the concept of dialectical nonbeing.”

* Tillich, ST, I. 188.
† Tillich, ST, I. 189.
§ Tillich, ST, I. 189.
¶ Tillich, ST, I. 190.

46. Roberts, “Tillich’s Doctrine of Man,” pp. 119–120: “The Christian doctrine of creatio ex nihilo attempts to solve the problem by denying that there is a second principle co-eternal with God; but it affirms that there is an element of nonbeing in all finite existence. Tillich denies that when Augustine attributes sin to nonbeing he is following a purely privative theory; rather, Augustine is asserting that although sin has no positive ontological status it nevertheless actively resists and perverts being. Indeed, since anything created originates out of nothing, it must return to nothing. This is why any view which regards the Son as a creature (Arianism) had to be rejected by the Church on the ground that a creature cannot bring eternal life. And this is why Christianity rejects the doctrine of natural immortality in favor of the belief that eternal life is given by God alone.”

47. Tillich, Systematic Theology, p. 189: “The dialectical problem of nonbeing is inescapable. It is the problem of finitude. Finitude unites being with dialectical nonbeing.”

Fig. 5.24b Carson, King, 2:397
The editor of the *Papers of Chief John Ross* incorporated the footnotes of the original documents sequentially into the editor's notes, distinguishing them within the notes themselves:

1 A footnote here reads: "The meaning here probably is that the names of beloved towns, which are exchanged between nations of indians when they made peace, were in this instance speedily exchanged back again; and this sudden return to each others' names is considered as indicating a sudden breaking of friendship—like the return of love tokens when courtships are broken off."
2 A footnote here reads: "Mohawks & Senecas both bore this name."

Enclosures, Attachments, and Envelopes

5.25 Editors should evaluate information found in enclosures, attachments, and envelopes and should determine if and how this material should be noted in the edition.

5.26 Frequently the author of a letter encloses it in other documents. These may include not only textual documents such as orders, newspaper clippings, wills, receipts, contracts, and the like but also such non- textual items as drawings, pictures, and maps. The editors of the *Papers of Woodrow Wilson* marked such items with the word enclosure in small capital letters and printed the full text of the enclosed document:

ENCLOSURE

Vira Boarman Whitehouse to Joseph Patrick Tumulty

Dear Mr. Tumulty:


Your letter of October 12th has brought rejoicing to the suffrage camp. You say the President asks what would be the best

The editors of the *Booker T. Washington Papers* chose not to reproduce an article enclosed in a letter, but instead provided a brief summary of the work and a full citation:

I enclose a very bad proof of an article which will appear in the July issue of the A.M.E. Church Review. It is an abstract of a fragment of my talks to the Hampton and Tuskegee students a year ago. My recol-

---

1 Durham's article, "Three Growths," *A.M.E. Church Review*, 14 (July 1897), 121-30, concluded that the establishment of trade schools and the movement to open factories, workshops, and businesses were in harmony with natural economic
orated the footnotes of the notes, distinguishing them

... is that the names of be-

of Indians when they made

again; and this sudden re-

sudden breaking of friend-

are broken off...".

Fig. 5.24c Mouton, Ross, 1:121

velopes

a enclosures, attachments,

: material should be noted

it other documents. These

lers, newspaper clippings,

ontextual items as draw-

f Woodrow Wilson marked

letters and printed the full

ick Tumulty

k Oct. 13th, 1917;

ocing to the suf-

would be the best

g. 5.26a Link, Wilson, 44:384

se not to reproduce an

copy of the work

ll appear in the July

act of a fragment of

ear ago. My recol-

revised, 14 (July 1897),

and the movement to

with natural economic

development. Durham discussed the different economic roles of slaves and how

they evolved after the Civil War. He had high praise for the ex-slave craftsmen

and suggested that schools such as Tuskegee continued that tradition.

Fig. 5.26b Harlan, Washington, 4:310

It is not unusual for a letter to mention an enclosure that cannot be located.
The editors of the Papers of Robert Morris speculated on the content of a miss-

ing enclosure in a footnote:

Office of Finance 23. Sept. 1782

Sir,

I enclose you Copy of a Letter of the thirteenth Instant from Mr

Webb Receiver of Taxes for Virginia.1 In Consequence of which I also

1. George Webb's letter has not been found, but apparently suggested purchase of

public tobacco from Virginia with Morris's notes in order to enable the state to make

payment on the congressional specie requisition for 1782. Although the arrangements

Fig. 5.26c Ferguson, Morris, 6:419

in some cases, enclosures can be found but not their cover letters. The Political

Correspondence and Public Papers of Aaron Burr printed an enclosure for which a

cover letter could not be located, alerting the reader in the document's endnote:

ALS (NHi:Burr), enclosed in an unlabeled letter to AB of the same date.

Fig. 5.26d Kline, Burr, 1:52–53

5.27 When documents contain address information on their reverse side, on

envelopes, or on wrappers, editors may wish to transcribe that information.
The Papers of George Mason quoted that information in a note following the

document's text, after the word Addressed:

RG (Morristown National Historical Park, N.J.). Addressed: "To Mr. John

Craig Merch. in Port Tobacco."

Fig. 5.27a Rutland, Mason, 1:146

The editors of Mark Twain's Letters used a small envelope icon to separate the

body of the letter from information taken from its envelope:

War H. Clagett, Esq. | Unionville | Humboldt Co. | N. T. [partly boxed, 
lower left:] Via Carson [postmaster's hand:] Esmeralda Cal. | April 21st 
1862 [bracketed] [three-cent U.S. postage stamp, canceled with a pen]

Fig. 5.27b Branch, Twain, 1:193
5.28 Editors often preserve endorsements—that is, notations written on the back of a document or a wrapper. The editors of the Political Correspondence and Public Papers of Aaron Burr placed endorsements in an unnumbered note at the end of the document following the provenance note:


Fig. 5.28a Kline, Burr, 1:224

The editors of the Papers of Robert Morris marked endorsements with the word ‘endorsed’ in small capital letters and placed it before the provenance note:

ENDORSED: 26th Augt 1782/Office of Finance/de Jno Bradford Esqr/received 16th Sept. seq. enclosed

Fig. 5.28b Ferguson, Morris, 6:259

Endorsements are particularly common on military and government documents, with each subsequent recipient writing a response or order directly onto the original document or its wrapper. Such endorsements occasionally take up several pages of text. The editors of Freedom: A Documentary History of Emancipation transcribed sequential endorsements in the following manner:

99: Commander of U.S. Forces at Fort Donelson, Tennessee,
    to the Headquarters of the 3rd Division, Reserve Corps,
    Army of the Cumberland

Fort Donelson Tenn. October 14th 1863.

There are at this Post about one hundred and twenty negroes who have been at work for the Government of the U.S. without pay since the first of June last. In order to get paid they must be mustered into the service as Infantry soldiers and detailed in the Engineer Department. Their services will be needed here for some time. Can there not be an officer sent here to muster them at once?

(sgd) E. C. Broctt.

[Endorsement] Hd. Qur’ Com’ Org. U.S.C.T. Nashville Nov 5th 1863. Respectfully referred to Major Gen’l Thomas Comdg Dep’l of Cumberland, Chattanooga Tenn. There is a large amount of money due negroes for work on the fortifications in the Department and I am informed by Lieut Burroughs (Eng. Corps) that there is money in his hands to pay them. Many families are suffering for want of money earned from six to twelve months since. The 1st Reg’l U.S. Col’l Vol’ of Tenn. earned by labor on the fortifications previous to Aug 15 Eighteen to twenty thousand dollars which is still due them. (sgd) Geo. L. Stearns Maj. & A.A.G. USV. Com’ Org U.S.C. Troops

5.31 From the language of the letter one can see the sensitivity to the issue at hand.
its—that is, notations written on the editors of the Political Correspondence of the editors of the Political Correspondence overseers in an unnumbered note at the end of the text.

"5. July 1795 / A Burr."

Fig. 5.28a Kline, Burr, 1:224

marked endorsements with the word "A" before the provenance note:

/de Joa Bradford Esqr; received 16th

Fig. 5.28b Ferguson, Morris, 6:259

in military and government documen
ting a response or order directly:

Such endorsements occasionally.

Freedom: A Documentary History of the editors of the Papers of James Madison placed docketing information in an unnumbered note following the document, after the letter code indicating the location of the document:

RC (DNA: RG 59, DD, Netherlands, vol. 4). Marked duplicate; docketed by Wagner as received 21 Nov.

Fig. 5.29a Bruggisser, Madison: State, 1:417

The editors of the Papers of John Adams also placed docketing in the note following the document, following the word "docketed:

RC (NH: Misc. MSS, Adams); docketed: "Mr J A Lettr June 1775."

Fig. 5.29b Taylor, Adams, 3:43

Previously Published Documents

5.30 In the case of new editions of previously published works or manuscripts, it may be important for editors to preserve references to the original pagination and page breaks. Readers may want to use a modern edition but still be able to find references or citations to the original work made in previous scholarship. The editor of the Complete Works of Captain John Smith placed the page number from the source text in the margin, in bold type and brackets, and indicated the exact location of page breaks in the original with two vertical lines (II):

your knowledge: now though some bee hurt by your misprision, yet he is your friend, and so will continue: and since the ice is open hee [65]

would have you send a: I waie your corne; and if you would have his companie send also your armes, which so affrighteth his people, that they dare not come to you, as he hath promised they should.

Fig. 5.30 Barbour, Smith, 1:250

Foreign Languages and Encoded Documents

5.31 Editors of documents written in languages other than the primary language of the edition's audience establish translation policies that balance fidelity to the original work with the desire to provide readers access to the content of the documents.
5.32 Editors of English-language documents containing frequently used foreign-language phrases may create a list of these phrases and their meanings. The editors of the *Diary of Elizabeth Drinker* provided at the beginning of the edition a list of French words and phrases in Drinker’s phonetic spelling:

**Glossary of French Words and Phrases**

Accouchement: delivery; childbirth

allez cet matin chez un person a mon desire: went this morning to the home of a person at my request

allez de: left; went out

a mon Chere: to my dear

a point du purpose: to no purpose

Fig. 5.32 Crane, *Drinker*, 1:iii

5.33 Editors may choose to translate documents originally written in a foreign language entirely into English. For example, the editor of *A Documentary History of the Indiana Decade of the Harmony Society* translated all documents into English but indicated that a document had been translated with *Tr.* in an unnumbered note at the end of the document:

and wish you continuous well-being while with friendship and respect I remain your sincere friend

**CASIMIR KURTZ**

[Endorsed:] Casimir Kurtz July 25th 1817

**ALS.** *Tr.* Since there is no postmark the letter was probably delivered by an immigrant. Kurtz was a reliable contact in Amsterdam for the Harmony Society. However, in 1805 he had been involved in an emigration scheme so questionable that the Duke of Württemberg on May 20, 1805, issued an order for the arrest of anyone connected with it and a similar order was issued by the Free City of Frankfurt am Main.

Fig. 5.33a Arnzt, *Harmony*, 1:369

The editors of the *Papers of Daniel Webster: Diplomatic Papers* preceded translated documents with the italicized word *Translation*:
COUNT NESSELRODE TO ALEXANDER DE BODISCO

Translation.

Copy of a despatch from the Vice Chancellor Count [Karl Robert] de Nesselrode to M. [Alexander] de Bodisco dated Saint Petersburg 28/30th March 1841.1

Fig. 5.33b Shewmaker, Webster: Diplomatic, 1:51

5.34 Editors may choose to print foreign-language documents in their original language. Some editors believe that such a decision preserves the exact text of the original, while others maintain that it denies access to those readers who are unfamiliar with the language of the document. Although the editors of the Papers of Benjamin Franklin did preface some non-English documents with abstracts in English, they printed the text of all foreign-language documents without translation:

From Paul-Pierre le Mercier de la Rivière

Ls: American Philosophical Society

This odd letter seems to be Franklin's only contact with a noted French philosophe. Le Mercier de la Rivière (1730–93) had been a councilor at the Parlement of Paris, intendant at Martinique, guest of Catherine the Great to advise her on a new legal code for Russia, and popularizer of Quesnay's economic doctrines.1 The accession of Louis XVI in May, 1774, and his appointment of Turgot were beginning a new if short-lived era of economic reform. An edict of Sept. 13, which permitted the importation of foreign grain and its free circulation in France,2 was doubtless the background for Mercier's proposal to buy five thousand tons of Philadelphia flour, in return if possible for wine. But how did he expect Franklin to arrange such a massive purchase in Pennsylvania? or to take French wine in trade, despite the restrictions of the Navigation Acts and the colonists' well developed partiality for port and madeira?3 We cannot answer these questions, and presume that Franklin could not either. If he ever troubled to respond to the letter, his reply has been lost.

De Paris ce 21 ÿbre 1774.

Je me flatte, Monsieur, que mon nom ne vous est point inconnu, puisque vous avez à Paris pour amis ceux qui sont les miens. C’est aussi parce que je vous connais très bien de réputation, que je n'hésite point à m'adresser à vous, par préférence à tous autres, pour l'affaire dont je vais vous entretenir.

Je voudrois, Monsieur, tirer de vos Colonies Angloises pour Paris 100 mille quincaux, et même plus, de vos plus belles fleurs

Fig. 5.34a Labarre, Franklin, 21:309
The editors of the *Letters of Henry Adams* translated difficult phrases but assumed a basic foreign-language literacy among the readers of the edition:

We have translated difficult and deceptive foreign words and phrases, but tried to avoid the obvious. We have assumed the modest acquaintance with French, German, Italian, and Spanish that tourists, schoolboys, or stamp-collectors might have. Three letters from Paris, which Adams wrote in the summer of 1860 and which we have left untranslated, provide a rough measure of what we call understandable French (even with his errors of spelling and usage intact). Our rule was not to gloss words, English or foreign, which can be found in a standard desk dictionary.

Fig. 5.34b Levenson, *Adams*, 1:xliv

The editors of the microfilm edition of the *Papers of General Friedrich Wilhelm von Steuben* did not translate foreign-language documents but preceded them with brief abstracts:

Abstract

*Paris Sept. 10, 1777*

Beaumarchais recommends St. to Robert Morris. He describes St.'s former duties as aide-de-camp to Frederick II in the Prussian Army and as court Chamberlain to the Prince of Hohenzollern; that he served 22 years under the Prussian king and will be honored to serve under Gen. Washington in the cause of liberty.

Fig. 5.34c von Zemskov, *von Steuben*, reel 1:10

5.35 Some editions contain both the text in its original language and a translation. The editors of the *Collected Papers of Albert Einstein* presented French and German documents in their original languages and produced a companion volume with translations of these documents into English. The editors of *Lafayette in the Age of the American Revolution* printed documents in their original languages with translations in an appendix. The editors of the *Papers of George Washington* translated foreign-language documents into English but provided the full text in the original language in the footnotes.

4. By

5.36 Editors of historical documents, aware of the need to bring my young and labors to the united and the splendor of the present day.
From Ségur

Sir,

St Petersburg [Russia] 24th August 1789.

I take the advantage of the departure of Mr Paul Jones to bring myself to the recollection of your Excellency, and to congratulate you upon the ratification of the new Constitution by the United States and electing you their President—They will not be deceived in their experience—Your wisdom will maintain the splendor of that liberty which your courage established.

2. The original letter reads: "Je veux profiter du départ de Monsieur Paul Jones pour me rappeler au Souvenir de Votre Excellence; Et pour la féliciter sur le choix que les Etats unis ont fait d'Elle lorsqu'ils ont adopté la nouvelle Constitution Et nommé un président du Congrès. Ils ne seront point trompés dans leurs Espérances Et Votre Sagesse Saura maintenir dans toute sa Splendeur la liberté que l'Ais Établie votre courage. Je désire infiniment que les circonstances me mettent à portée d'aller un jour contempler la prospérité d'un Pays pour le quel je m'estime heureux d'avoir Combattu, Et la gloire d'un général célèbre Sous le quel je regrette de n'avoir pas plus longtemps Servi.

The editors of the Family Letters of Victor and Meta Berger retained German phrases in English-language letters, providing translations in the footnotes:

That's all I can tell you to night. I must take a walk—because I feel that I don't get enough exercise.

Im übrigen hab' ich Dich sehr lieb. Kiss the babies for me, and remember I am ever

Your
V.L.B.

4. By the way, I love you very much.

5.36 Editors decode documents written in code to make them accessible to readers. The editors of John Jay presented decoded passages in small capital letters:
The editors of the *Papers of James Madison* presented decoded passages in italics with an indication in the unnumbered source note following the document that the passage was written in code:

> to 32 & 30/ on James River & 28/ on Rappahannock. The scarcity of cash is one cause. *Harrison the late governor was elected in Surry whither he previously removed with his family. A contest for the chair will no doubt ensue. Should he fail it he will be for Congress.*

---

RC (DLC). Cover missing. Docketed by JM and in an unknown hand. Also headed “No. 8.” Italicized words, unless otherwise noted, were encoded by JM in the code first used by Jefferson on 14 Apr. 1781

---

5.38 Editor's exceptions among doc. aband. volumes of nos. sued a more Literal Office Confid. A few also reg. such docs (script)

---

**5.37** Many editors convert documents written in shorthand into standard written English. The editor of *An Education in Psychology: James McKeen Cattell's Journal and Letters from Germany and England* preceded documents originally written in shorthand with a note explaining the condition of the original text:

1. **Journal, Paris, 7 June 1881**

   (Transliterated from Pitman shorthand except for proper names, titles, and place names, which were originally written in longhand.)

   *The evening before leaving London I heard Patti sing in Rossini's Semiramide.*

   *Wednesday the 25th we came over via Dover & Cal.*

---

The editors of the *Papers of Woodrow Wilson* presented his shorthand diary in standard English but alerted readers in the heading that the text was written in shorthand:

---

**From Wilson's Shorthand Diary**

> [July] 19th Thursday. Spent most of the day playing croquet[,] riding etc. Made satisfactory progress on my speech and read the introduction of Fox's speech on parliamentary reform—his great effort in support of Mr. Gray's bill. Went with Jessie in buggy to

---

**Fig. 5.37a Sokol, Cattell, 33**

**Fig. 5.37b Link, Wilson, 1:285**
Exceptions to Transcription Policies

Editors may make exceptions to their transcription policies as long as exceptions are clearly defined. Exceptions should be made to reflect differences among documents or to highlight specific, unusual cases for the reader. Before abandoning their policy of modernizing transcription, the editors of the first three volumes of the Documentary History of the Ratification of the Constitution also pursued a more literal policy for documents of exceptional significance:

**Literal Reproduction of Official Documents**

Official documents such as the Constitution, resolutions of the Confederation Congress, state acts calling conventions, forms of ratification, and proclamations are reproduced as literally as possible. A few other documents, because of their character or importance, are also reproduced as literally as possible. The literal reproduction of such documents is indicated by the symbol “LT” (i.e., literal transcript) in the footnote citation to the source.

Fig. 5.37a, Sokal, Cattell, 33

Presented his shorthand diary in that the text was written in

Fig. 5.37b, Link, Wilson, 1:285

He day playing croquet[,] my speech and read the

Fig. 5.38, Jensen, Ratification, 2:14

and in an unknown hand. Also

Fig. 5.38a, Madison, 8:346-347

ppahannock. The scarcity of

eror was elected in Surry

imility. A contest for the chair
ll be for Congress.

[In shorthand into standard

n in Psychology: James McKeen

d England preceded documents
explaining the condition of the

...for proper names, titles, and

a longhand.)

[In Rosini’s Semiramida.]

il. We went to the Hotel de la

Fig. 5.37a, Sokal, Cattell, 33

Presented his shorthand diary in that the text was written in
Chapter Six
Principles of Annotation

Reasons for Annotation

6.1 Annotation, the information added by editors to improve readers' understanding of historical documents, can serve several different functions. Editors may explain the history of documents, supply missing parts of the text such as a date, place, or word; or offer editorial commentary that helps the reader understand the text. Annotation may appear as bracketed insertions in the text, footnotes, endnotes, headnotes, microform targets, or supplemental materials such as tables, illustrations, charts, glossaries, directories, and introductory essays. Annotation makes the text of documents more readable, clarifies unusual terms, offers background on events and people, supplies missing information, and provides readers with historical context.

6.2 The quantity and specificity of annotation will be determined by the needs of the audience, the characteristics of the documents being edited, the resources available for researching and printing annotation, and the judgment of the editors. Editors of complex documents may need to provide extensive annotation to assist their readers, while easily understood documents may require little annotation. Editions produced for academic audiences may assume a high level of background knowledge and thus provide readers with more technical information, while editions for general audiences may assume their readers have little historical background and thus use their annotation to contextualize and clarify documents. Editors balance the value of providing useful information that will enhance the accessibility of documents with the cost of producing and printing notes. Furthermore, excessive annotation may preclude printing additional documents and can diminish an edition's readability by cluttering the page.
6.3 The editors of *Freedom: A Documentary History of Emancipation*, which is intended primarily for an academic audience, provided narrative introductions to their volumes and chapters but offered no introduction to individual documents:

153. Missouri Unionist to the Commander of the Department of the West, and the Latter's Reply

Saint Louis Mo. May 14, 1861.

Sir: In common with thousands who have perused your admirable proclamation of this morning,* I return you the thanks of a citizen of Missouri for its patriotic tone and tranquillizing assurances.

Fig. 6.3a Berlin, *Freedom*, ser. 1:1:413

When the editors published the same document in *Free at Last: A Documentary History of Slavery, Freedom, and the Civil War*, an edition for a general audience, they introduced the document with a narrative headnote that linked it to the preceding document and provided the background and context that a general reader might need:

While some Confederate partisans foresaw a threat to slavery; President Abraham Lincoln and the U.S. Congress reiterated that the North sought only to preserve the Union. Desiring to reassure unionists in the slave states—particularly the border states of Maryland, Kentucky, and Missouri, which had not seceded—Lincoln promised that the army would respect the property rights of slaveholders. In the early months of the war, federal field commanders hewed closely to the president's policy. Writing to General William S. Harney, the commander at St. Louis, a Missouri unionist sought assurances that the government would protect slavery.

Saint Louis Mo. May 14, 1861.

Sir: In common with thousands who have perused your admirable proclamation of this morning,* I return you the thanks of a citizen of Missouri for its patriotic tone and tranquillizing assurances.

Fig. 6.3b Berlin, *Freedom*, 6

Types of Annotation

6.4 Editors create three general categories of notes: provenance notes, textual notes, and contextual/informational notes. Provenance notes report the current location of source documents and, in some instances, their publication
The Command of the
and the Latter's Reply

The Missouri, May 14, 1861.

revered your admirable
and the thanks of a citizen of

The Missouri, May 14, 1861.

the thanks of a citizen of

Provenance Notes

Editors frequently include provenance notes to indicate documents' location, genre, and publication history. Editions drawn from documents at a single repository or from within a single collection may be able to provide this information concisely in their introductions unless the documents cannot be located in the repository or collection without more detailed guidance. For editions containing documents drawn from multiple archives and manuscript repositories, editors provide this information for each document.


For documents taken from newspapers, the editors of the Documentary History of the Ratification of the Constitution provided the name of the newspaper and the date when the article or letter appeared, as well as contextual information about the item's circulation:

1. Charleston Columbia Herald, 22 October. Reprinted twelve times from Richmond to Boston by 26 November.

Some editions are more specific about the location of manuscripts, providing box numbers for archival collections. While this information can be helpful for readers, box numbers may change if collections are moved, processed, or reorganized. In large repositories, record group and file numbers are essential provenance information. The Marcus Garvey and Universal Negro Improvement Association Papers provided information on the repository, record group, and specific file where documents were located:

DNA, RG 165, File 10218-77/5 2-1.

Fig. 6.6c Hill, Garvey, 1.337
Editors of documents that have been microfilmed may cite the microfilm reel as part of their provenance information. The Samuel Gompers Papers led readers to both reel 9 of the microform Letterbooks of the Presidents of the American Federation of Labor and the original manuscripts (SG Letterbooks) housed at the Library of Congress (DLC):


Fig. 6.6d Kaufman, Gompers, 4:140

The Collected Papers of Albert Einstein provided a note citing the source for Einstein’s published writings:


[Einstein 1910a]

PUBLISHED 15 January and 15 February 1910

IN: Archives des sciences physiques et naturelles 29 (1910): 5-28, 125-144

Translated by Edouard Guillaume

Fig. 6.6e Staehel, Einstein, 3:130

The editors of Mark Twain’s Letters provided a section at the end of the volume describing the provenance and publication history of all documents included in the volume:

• 30 January 1862 • To Jane Lampton Clemens • Carson City, Nev.
  Terr. • UCCL 60034
  • Copy-text: PH, “Model Letter from Nevada,” Keokuk Gate City, 6 Mar 62, 4.
  Newsprint of the Gate City is in the Iowa State Historical Department, Division of
  the State Historical Society, Iowa City (IaHi), and the Hildreth Library, The
  State Historical Society of Iowa, Des Moines. We have not been able to use the
  original newspaper as copy-text. • Previous publication: Paine, 419-20, and
  MTB, 1:183-84, paraphrase and excerpts; Lorch 1938, 345-49; Rogers 1961,
  29-34. • Provenance: Paine cites the Keokuk Gate City as his source for the
  excerpts published in MTB; he evidently did not see the MS, which is not known
  to survive.

Fig. 6.6f Branch, Twain, 1:505
The provenance of documents may be noted in an editorial statement or introduction, as was done by the editors of the *Family Letters of Victor and Meta Berger*:

All the letters printed here are recipients' copies found in the Victor L. Berger Papers at the State Historical Society of Wisconsin, with the exception of Victor's letter of July 28, 1926, to Jan Edelman, which is a retained carbon copy from the same collection.

The *Papers of Joseph Henry* provided provenance information in document headings (see Section 7.2).

### Textual Notes

6.7 Textual notes provide information about the physical appearance of a document that is not conveyed through transcription. Common textual notes describe details such as deletions, insertions, interlineations, tears, amendments, evidence of multiple authorship, enclosures, and marginalia. They are also used to compare variant copies of the same text. Some textual notes appear as editorial insertions within the text. Others are placed in footnotes or headnotes.

6.8 The editors of *John Jay* used a footnote to describe William Laight's unusual method of deleting and inserting passages:

> Laight did not actually cross out the word "address" when he added "short[ ] Epistle" above the line as a substitute. This peculiarity is repeated several times throughout the draft, and such incomplete deletions and substitutions have been transcribed as though the material to be omitted has been struck out.

The editors of the *Papers of George Washington: Colonial Series* used a footnote to explain the mutilated condition of the original manuscript and the method used by the editors to transcribe it:

> LS, DLC:GW; MH: Sparks Papers. The letter was written on both sides of a sheet of paper, and the sheet was torn approximately in half. The top part of both pages of the letter are in DLC; the bottom parts are in MH. The letter appears to be in the hand of John Kirkpatrick. The second postscript, dated "29th," is in Dinwiddie's shaky hand, a rare sample of his handwriting from his years in Virginia.

1. The portion of the first page, at DLC, ends here; and the portion at MH begins with the lower half of the words taken to be "a detachment." The left-hand margin of the lower part of the first page has been torn off. The letters in angle brackets in this and the next paragraph are the letters that are missing as a consequence of the tear. In his *Writings of Washington*, Stanislaus
Hamilton prints only the top, or DLC, portion of the letter, but he does supply the letters missing from words on the lower half of the page at MH, indicating that the scrap torn from the lower left margin remained with the upper half of the letter in DLC. Hamilton also supplies words or parts of words that he mistakenly supposed came at the end of the lines that were missing to him.

Fig. 6.8b Abbot, Washington: Colonial, 4:138

6.9 Notes may be used to describe documents or enclosures that were not selected for publication but may contain information of value. Such notes may take the form of abstracts that quote or excerpt the most significant passages of documents (see Section 2.23), or they may be footnotes or headnotes that describe the document.

Contextual/Informational Notes

6.10 Contextual notes include any type of editorially supplied information that helps readers more fully understand and appreciate the content of a document. Contextual notes generally amplify or clarify the information in a document by providing historical background, biographical data, fuller descriptions of mentioned events, clarification of ambiguous passages or words, unstated outcomes, corrections of erroneous information in the text, or other information about people and events that could not be found by reading the documents alone. Contextual notes allow editors to share their expertise with the reader, who can then more fully appreciate the meaning, background, and content of a document.

6.11 Biographical notes present information about authors, recipients, or individuals named or mentioned in a document, providing readers with sufficient background to understand the role of an individual within, and occasionally beyond, the context of a document. Not all individuals need to be identified, and those who are identified do not need to be described with equal detail. Moreover, information cannot always be found about every person mentioned in a document. Editors choose whom they will identify and how much information they will provide. Basic biographical information should be provided only once, usually the first time an individual is mentioned or within a biographical directory placed at the beginning of the edition or in an appendix. The editors of the *Papers of Andrew Johnson* provided full biographical notes for most individuals mentioned:

1. Alexander Outlaw Anderson (1794–1869) at the time of this letter was unsuccessfully seeking the position of American minister to Mexico. An East Tennessean who was with Jackson at New Orleans, he served briefly as U.S. senator (1840–41), filling out the unexpired term of Hugh Lawson White. Later he participated in the gold rush and briefly held office in California. During the Civil War he lived in Alabama, but returned to Knoxville before his death. *BDAC*, 477.

Fig. 6.11a Graf, Johnson, 1:153
The editor of the *Correspondence of Mother Jones* provided brief biographical descriptions that place individuals in the context of the document in which they are mentioned:

**ALS (William B. Wilson Papers)**
1. William Warner, a UMWA organizer, had led the strike of the Maryland miners in 1900, when Mother Jones was an associate in the field. Later in life he became a mine superintendent.
2. Albert Manka, like Warner an international organizer, had worked in West Virginia the preceding year, when Warner was still involved in the Maryland strike.

The editors of the *Black Abolitionist Papers* wanted to improve understanding of the contributions of members of the black community whose lives had received little attention from historians. Consequently, they provided longer biographical entries for people and events that could not be found in standard reference sources:

**People and events**
that are covered in standard biographical directories, reference books, or textbooks are treated in brief notes. We have given more space to subjects on which there is little or no readily available information, particularly black individuals and significant events and institutions in the black community. A full note on each item is presented at the first appropriate point in the volume.

---

**6.12** Editors seldom record information that is common knowledge to their audience in the annotation. For example, most editions aimed at American audiences would not need to note that George Washington served as the first president of the United States, but they might include lesser-known information about him if necessary to understand the documents.

**6.13** Editors will find individuals who merit identification but about whom they cannot locate authoritative sources. Most editors pass over these people without comment, some acknowledge the lack of information with a note such as that employed by the *Papers of Daniel Webster: Correspondence*:

1. Not identified.

---

*Graf, Johnson, 1:153*
When a small quantity of information is available, an editor may provide it but acknowledge other areas of uncertainty, as did the editor of the Letters of Eugene V. Debs:

2. Samuel Howell is listed as a carpenter in the Terre Haute city directories of 1871-73. The basis of his notoriety in Terre Haute has not been discovered.

Fig. 6.13b Constantine, Debs, 14

6.14 Literary references, Biblical references, and foreign phrases can be identified and explained in an edition's notes. Passages that would have been readily understood by a document's intended recipient or audience may otherwise be difficult for a contemporary reader to grasp. For example, the editors of the Documentary History of the Ratification of the Constitution provided notes locating literary quotations or misquotations and explaining their relevance:

   the very idea of which, is enough to make every honest citizen exclaim in the language of Cato, O liberty, O my country!—

3. The phrase, taken from Joseph Addison's play, Cato. A Tragedy (1713), actually reads: "O liberty! O virtue! O my country!" (act 4, scene 4). First performed in London, this play was about Marcus Porcius Cato Uticensis (Cato the Younger), a republican opponent of Julius Caesar, who committed suicide in 46 B.C. rather than accept the triumph of Caesar over Pompey and his followers, of whom he was one. (See "Cato Uticensis," Virginia Independent Chronicle, 17 October [above] which defended Mason.)

Fig. 6.14a Jensen, Ratification, 8:211

The editors of the Papers of Martin Luther King, Jr. provided sources and more complete quotations for fragmentary literary references:

This universe hinges on moral foundations. (Yeah) There is something in this universe that justifies Carlyle in saying,

   No lie can live forever.¹

¹ King may have been paraphrasing Carlyle's French Revolution (1837), part 1, book 3, chapter 1: "No lie you can speak or act but it will come, after longer or shorter circulation, like a bill drawn on Nature's Reality, and be presented there for payment—with the answer, No effects."

Fig. 6.14b Carson, King, 2:253

6.15 Geographical descriptions that were evident to the author or recipient of a document may be meaningless to the readers of a documentary edition without editorial assistance. Editors help their readers place geographical references in
6.14 Principles of Annotation

context by providing maps or descriptions. For example, the editors of the Diary of Elizabeth Drinker used a note to identify the location of a bridge:

—WD. and self, walk'd as far as the draw Bridge this evening.—The name of Drawbridge is continued, tho there has not been one there since my memory.*

47. By 1740 a stone bridge had replaced a drawbridge erected in the early 1690s to extend Front Street over Dock Creek. During the course of the eighteenth century the creek was filled in and paved to form Dock Street (Joseph Jackson, Encyclopedia of Philadelphia (Harrisburg, Pa.: The National Historical Association, 1934), 2:661, 591).

The editors of the Papers of General Nathanael Greene also used annotation to help readers understand cursory references to geography:

Our purpose in crossing Santee was to fight [Col. John] Watson but unluckily We was one day to Late. He crossed at Buckenham's Watson yesterday & is gone toward Camden with about two hundred men & two field p[iece].$$

3. On Watson's earlier movements, see note at Wade to NG, 19 April, above. From Georgetown, where he had retreated after breaking off pursuit of Marion, Watson marched toward the Santee River; his detachment was "much reduced in Number, thro' Casualties, Sickness and a Reinforcement which he had left to strengthen the Garrison at Georgetown." (Rawlins to Cornwallis, 24 May, PRO 30/11/6) He crossed the Santee at Lenard's Ferry, near the river's mouth, and waited for a day or two by order of Nisbet Balfour to cover the possible return of Lord Cornwallis or Banastre Tarleton to South Carolina. Watson's detachment of about 500 men then moved toward Nelson's Ferry, on the Santee. (Balfour to Germain, 1 May, PRO 30/11/109; Balfour to Clinton, 6 May, ibid.; Rankin, Swamp Fox, p. 200) Henry Lee later stated that NG's position prevented Watson from reaching Camden "on the usual route from Motte's post." Watson, who was moving up the south side of the Santee, was thus left with two other options, according to Lee: he could cross the Congaree at Motte's and the Wateree below the High Hills of Santee or else cross the Santee below Motte's and move up the north side of the Santee by way of the High Hills. (Lee, Memoirs, 2: 70) Guessing that Watson would take the former route, Marion and Lee crossed to the south side of the Santee at Scott's Lake and moved "with celerity" toward Motte's. (Lee, Memoirs, 2: 70-71; Rankin, Swamp Fox, p. 201) Watson, however, who was now ahead of them, crossed the Santee at Buckenham's—or Buchanan's—a small ferry near the confluence of the Congaree and Wateree, which, he later wrote, had been "left unguarded" because it was supposedly "impracticable" to cross. His detachment then proceeded to Camden "without molestation," though not without difficulty, by a route that required the troops to wade across six creeks, build a sixty-foot bridge over a seventh, and cut "for about a Mile and half through the canes that grow in those swamps."

—WD. and self, walk'd as far as the draw Bridge this evening.—The name of Drawbridge is continued, tho there has not been one there since my memory.*

47. By 1740 a stone bridge had replaced a drawbridge erected in the early 1690s to extend Front Street over Dock Creek. During the course of the eighteenth century the creek was filled in and paved to form Dock Street (Joseph Jackson, Encyclopedia of Philadelphia (Harrisburg, Pa.: The National Historical Association, 1934), 2:661, 591).

The editors of the Papers of General Nathanael Greene also used annotation to help readers understand cursory references to geography:

Our purpose in crossing Santee was to fight [Col. John] Watson but unluckily We was one day to Late. He crossed at Buckenham's Watson yesterday & is gone toward Camden with about two hundred men & two field p[iece].$$

3. On Watson's earlier movements, see note at Wade to NG, 19 April, above. From Georgetown, where he had retreated after breaking off pursuit of Marion, Watson marched toward the Santee River; his detachment was "much reduced in Number, thro' Casualties, Sickness and a Reinforcement which he had left to strengthen the Garrison at Georgetown." (Rawlins to Cornwallis, 24 May, PRO 30/11/6) He crossed the Santee at Lenard's Ferry, near the river's mouth, and waited for a day or two by order of Nisbet Balfour to cover the possible return of Lord Cornwallis or Banastre Tarleton to South Carolina. Watson's detachment of about 500 men then moved toward Nelson's Ferry, on the Santee. (Balfour to Germain, 1 May, PRO 30/11/109; Balfour to Clinton, 6 May, ibid.; Rankin, Swamp Fox, p. 200) Henry Lee later stated that NG's position prevented Watson from reaching Camden "on the usual route from Motte's post." Watson, who was moving up the south side of the Santee, was thus left with two other options, according to Lee: he could cross the Congaree at Motte's and the Wateree below the High Hills of Santee or else cross the Santee below Motte's and move up the north side of the Santee by way of the High Hills. (Lee, Memoirs, 2: 70) Guessing that Watson would take the former route, Marion and Lee crossed to the south side of the Santee at Scott's Lake and moved "with celerity" toward Motte's. (Lee, Memoirs, 2: 70-71; Rankin, Swamp Fox, p. 201) Watson, however, who was now ahead of them, crossed the Santee at Buckenham's—or Buchanan's—a small ferry near the confluence of the Congaree and Wateree, which, he later wrote, had been "left unguarded" because it was supposedly "impracticable" to cross. His detachment then proceeded to Camden "without molestation," though not without difficulty, by a route that required the troops to wade across six creeks, build a sixty-foot bridge over a seventh, and cut "for about a Mile and half through the canes that grow in those swamps."

—WD. and self, walk'd as far as the draw Bridge this evening.—The name of Drawbridge is continued, tho there has not been one there since my memory.*

47. By 1740 a stone bridge had replaced a drawbridge erected in the early 1690s to extend Front Street over Dock Creek. During the course of the eighteenth century the creek was filled in and paved to form Dock Street (Joseph Jackson, Encyclopedia of Philadelphia (Harrisburg, Pa.: The National Historical Association, 1934), 2:661, 591).

The editors of the Papers of General Nathanael Greene also used annotation to help readers understand cursory references to geography:

Our purpose in crossing Santee was to fight [Col. John] Watson but unluckily We was one day to Late. He crossed at Buckenham's Watson yesterday & is gone toward Camden with about two hundred men & two field p[iece].$$

3. On Watson's earlier movements, see note at Wade to NG, 19 April, above. From Georgetown, where he had retreated after breaking off pursuit of Marion, Watson marched toward the Santee River; his detachment was "much reduced in Number, thro' Casualties, Sickness and a Reinforcement which he had left to strengthen the Garrison at Georgetown." (Rawlins to Cornwallis, 24 May, PRO 30/11/6) He crossed the Santee at Lenard's Ferry, near the river's mouth, and waited for a day or two by order of Nisbet Balfour to cover the possible return of Lord Cornwallis or Banastre Tarleton to South Carolina. Watson's detachment of about 500 men then moved toward Nelson's Ferry, on the Santee. (Balfour to Germain, 1 May, PRO 30/11/109; Balfour to Clinton, 6 May, ibid.; Rankin, Swamp Fox, p. 200) Henry Lee later stated that NG's position prevented Watson from reaching Camden "on the usual route from Motte's post." Watson, who was moving up the south side of the Santee, was thus left with two other options, according to Lee: he could cross the Congaree at Motte's and the Wateree below the High Hills of Santee or else cross the Santee below Motte's and move up the north side of the Santee by way of the High Hills. (Lee, Memoirs, 2: 70) Guessing that Watson would take the former route, Marion and Lee crossed to the south side of the Santee at Scott's Lake and moved "with celerity" toward Motte's. (Lee, Memoirs, 2: 70-71; Rankin, Swamp Fox, p. 201) Watson, however, who was now ahead of them, crossed the Santee at Buckenham's—or Buchanan's—a small ferry near the confluence of the Congaree and Wateree, which, he later wrote, had been "left unguarded" because it was supposedly "impracticable" to cross. His detachment then proceeded to Camden "without molestation," though not without difficulty, by a route that required the troops to wade across six creeks, build a sixty-foot bridge over a seventh, and cut "for about a Mile and half through the canes that grow in those swamps."
The editor of "As to the People: Thomas and Laura Randell's Observations on Life and Labor in Early Middle Florida," published in the Florida Historical Quarterly, used a footnote to provide historical information about Tallahassee in the 1820s:

5. Tallahassee had been established as the territorial capital in 1823. The village that Randell visited on business was the home of about a thousand people, and claimed a church, school, several hotels, and various businesses. Mary Louise Ellis and William Warren Rogers, Favor, Land Tallahassee: A History of Tallahassee and Leon County (Norfolk/Virginia Beach, 1988), 33-36.

6.16 Collected documents often provide a fragmentary narrative because some documents are lost or not printed and because authors frequently fail to follow up on events discussed in their documents. Editors can bridge the narrative of documents by providing readers with the missing parts of the story in the notes. Such notes are particularly useful in explaining responses when editors choose to publish only a single side of a correspondence, such as Felix Frankfurter's responses to Louis Brandeis's letters provided by the editors of "Half Brother, Half Son": The Letters of Louis D. Brandeis to Felix Frankfurter:

It occurred to me that you might know of some New York (or other) lawyers, who would be glad to join in making this gift. 3

3 On 27 January FF replied: "I have gladly written to several New York lawyers of the opportunity to ease their swollen fortunes."

6.17 Notes can clarify passages that are confusing, ambiguous, or unclear. For example, the editors of the Papers of Ulysses S. Grant used a footnote to explain the meaning behind Grant's cryptic phrase "urgent family reasons":

I am induced to make this application at this time for the reason that my services can probably be better dispensed with at present than at any future time, there being at this Post, with one company, a Commanding officer, Adjutant and three Company officers besides myself. Urgent family reasons also induce me to respectfully submit this application.

The "urgent family reasons" mentioned by USG probably involve the birth of USG's son Frederick Dent Grant on May 30, 1850, in St. Louis. Julia Dent Grant had been advised by Maj. Charles Stuart Tripler, surgeon at Detroit, to return to her parents for the birth of her first child. Unpublished memoirs of Julia Dent Grant.

Fig. 6.17a Simon, Grant, 1:194
The editors of the Letters of Delegates to Congress used contextual notes to clarify oblique references, such as this one made by John Adams:

"The People of England, have thought that the Opposition in America, was wholly owing to Dr. Franklin: and I suppose their scribblers will attribute the Temper, and Proceedings of this Congress to him; but there cannot be a greater Mistake. He has had but little share farther than to co operate and assist. He is however a great and good Man. I wish his Colleagues from this City were All like him, particularly one, whose Abilities and Virtues, formerly trumpeted so much in America, have been found wanting."

*That is, John Dickinson. Adams' disparaging remarks about Dickinson were extended in a letter the next day to James Warren, which was subsequently captured and published by the British. John Adams to James Warren, July 24, 1775.*

The editors of the Papers of William Livingston used notes to define unfamiliar terms such as draught:

"The Draught¹ from my Battalion serving on the Woodbridge Department is not yet returned, as they have been out six Times since"

ALS, MHi.
1. draught: a group of soldiers posted from a larger body to some special duty.

Editors can use notes to clarify misstatements of fact in original documents, such as in this passage from Letters of Delegates to Congress:

"The Secretary for Foreign Affairs yesterday read to us a passage in a letter from a correspondent at Boston which positively asserts that General Sullivan had marched with 1000 men to reduce to obedience to their government the revolters on the East side of Connecticut River."

³ Although the New Hampshire legislature had authorized the raising of two regiments commanded by John Sullivan "to be sent to the western part of the State" to subdue those border towns attempting to join Vermont, the governor did not call the force into service because the legislature subsequently voted to pardon those residents who had previously revolted. Sec. N. H. State Papers, 10:476-78, and Charles P. Whittmore, A General of the Revolution: John Sullivan of New Hampshire (New York: Columbia University Press, 1901), p. 184."
Editors may use contextual notes to provide follow-up information about events, such as when the editors of the Papers of George Washington: Revolutionary War Series used a note to tell their readers the outcome of a request by the Six Nations to have an audience with the Continental Congress in Philadelphia:

These Indians reached Philadelphia by 25 May when they requested an audience with Congress. That audience was held at 11:00 A.M. on 27 May, and preceding it at 9:00 A.M. Congress staged a military review "in order," as one delegate candidly remarked, "to give those savages some idea of our strength and importance" (Joseph Hewes to Samuel Johnston, 26 May 1776, in Smith, Letters of Delegates, 4:77-78; JCC, 4:392-93, 396-97). "Coll. Dickinson's, Coll. Robertdoo's, Coll. Cadwallader's, Coll. Mclean's and Coll. Matlock's Battalions [of associators], three companies of Artillery & the Light-horse of the Militia; and Coll. Shee's and Coll. Magaw's Battalion of the Continental Troops Were all Reviewed," Caesar Rodney wrote Thomas Rodney on 29 May, "the day before Yesterday, on the Common, by the Congress, Generals Washington, Gates and Mifflin accompanied by a great number of other officers, most of the Assembly, the Presbyterian Clergy who were here at the Sidewalk—and 21 Indians of the Six Nations, who gave the Congress a War-dance yesterday" (Smith, Letters of Delegates, 4:99-100). The Indians appeared in Congress again on 11 June when a speech pledging American friendship and various gifts were delivered to them (JCC, 4:410, 5:421, 430-31, 471).

Fig. 6.17e Chase, Washington: Revolutionary, 4:320

Editors may use contextual notes to alert readers when documents they may expect to find in the edition are not included, such as when the editors of the Papers of James Madison chose not to publish two documents previously attributed to Madison that they believed he did not write:

EDITORIAL NOTE

Few editors and scholars would question that the drafting of the Virginia Resolutions was JM's most important contribution to the proceedings of the 1798-99 session of the Virginia General Assembly. Earlier generations of editors have maintained, however, that JM was responsible not only for the resolutions of 21 December 1798 but also for three resolutions criticizing the foreign policy of the Adams administration that were approved by the General Assembly on 10 January 1799; and more importantly they have claimed that he wrote another attack on the Alien and Sedition Acts in the form of the "Address of the General Assembly to the People of the Commonwealth of Virginia" of 23 January 1799 as well. Both the compilers of the congressional edition and Gaillard Hunt included these last two documents in their canons of JM's papers, and several of JM's biographers, from John Quincy...
follow-up information about events during George Washington: Revolutionary War sometime of a request by the Six Nations Congress in Philadelphia:

5 May when they requested an audience at 11:00 A.M. on 27 May, and military review "in order," as one savages some idea of our strength Johnston, 26 May 1776, in Smith, 93, 396-97. "Coll. Dickinson's, J. McKean's and Coll. Matlock's of Artillery & the Light-horse of w's Battalion of the Continental y wrote Thomas Rodney on 29 mon. by the Congress, Generals by a great number of other off. Clergyman who were here at the Si. gave the Congress a War-dance. The Indians appeared in asking American friendship and 410, 5:421. 450-51, 471).

So will those when documents they may such as when the editor of the two documents previously attribu-

8

The drafting of the Virginia resolution to the proceedings of the assembly. Earlier generations of historians responsible not only for the three resolutions criticizing the Articles of Confederation, the compilers of the Congress last two documents in their collection, from John Quincy Adams to Irving Brant, have concurred with this attribution, at least to the extent of asserting that the "Address" of 23 January 1799 was the product of JM's pen (Madison, Letters [Cong. ed.], 4:308-14; Madison, Writings [Hunt ed.], 5:331-40; John Quincy Adams, An Eulogy on the Life and Character of James Madison, Fourth President of the United States; Delivered at the Request of the Mayor, Alderman, and Common Council of the City of Boston, September 27, 1836 [Boston, 1836], p. 57, later republished as The Lives of James Madison and James Monroe [Buffalo, 1850], see p. 70; Brant, Madison, 3:464). The present editors do not believe that JM composed these two documents, and for that reason they have decided to exclude them from this edition of his papers. This decision was not taken lightly, nor without an awareness of its significance for the body of important scholarship that exists in relation to the contents of these two documents, particularly the "Address" of 23 January 1799. If JM did not write this last document, historians will have to reconsider some aspects of the lengthy and complicated debate about the development of thinking on the freedom of the press in the United States, especially with respect to the claim that JM changed his views markedly on this subject between January and December 1799. The author of the "Address" of 23 January 1799 denied that the federal government could institute prosecutions for sedition libel, but at the same time he explicitly left open the possibility of punishing "libellous writing or expression" in state courts by juries summoned by an officer, who does not receive his appointment from the President. The power of state courts over matters of sedition libel, however, was denied by JM barely twelve months later when he wrote his Report of 1800 on the Virginia Resolutions, a document that has become justly celebrated as one of the most important expositions of the meaning of the freedom of the press clause in the First Amendment (Madison, Letters [Cong. ed.], 4:390; see Walter Berns, "Freedom of the Press and the Alien and Sedition Laws: A Reappraisal," in 1970: The Supreme Court Review [Chicago, 1970], ed. Philip B. Kurland, pp. 119-30; Levy, Emergence of a Free Press, pp. 319-26; Jeffrey A. Smith, Printers and Press Freedom: The Ideology of Early American Journalism [New York, 1988], pp. 71-73).

It is, unfortunately, not possible to address and resolve all these significant issues in an edition of JM's papers. As it is, little enough is known about JM's activities at this time, and historians do not understand a great deal more about the proceedings in the Virginia General Assembly as it ratified these two documents of protest in January 1799. The shreds of information that are available, however, cannot be adduced to support the contention that JM was responsible for either of them. In this context, it should be stressed above all else that there exists no good contemporary evidence of any sort, direct or circumstantial, to substantiate the claim that JM had either composed the two documents in question or contributed in any way to their adoption by the General Assembly.

Fig. 6.17f Hutchinson, Madison, 17:199-200
6.18 Editors may bridge gaps between documents by writing transitional passages, such as the one below, written by the editors of the *Papers of Daniel Webster: Correspondence*, that took the reader from international relations to domestic politics and provided the necessary context for understanding the documents that followed:

Although Webster's energies during the winter of 1851–1852 were largely absorbed in his conduct of foreign affairs, including delicate negotiations regarding a canal across Central America, his attention was never far from politics. Despite his age, and the practical obstacles to a nomination in 1852, Webster's presidential ambitions burned as strongly as ever—perhaps more so, since he well knew this would be his last chance for the elusive prize.

Unfortunately, a long list of prominent backers in the major cities of the Eastern Seaboard did not translate into popular enthusiasm for his cause, or add up to delegate strength for the upcoming Whig National Convention. Moreover, although Webster was loath to admit the fact, southern support for his candidacy was almost entirely lacking. With the decline of fears for the Union, his efforts in 1850 made less of a difference to southern moderates, many of whom looked to Millard Fillmore as their best hope in 1852. For his part, the President, avowedly not a candidate, nonetheless declined to withdraw from the race, a "nondecision" that was announced on January 24 by the Republican, Fillmore's organ in Washington.

Under the circumstances, Webster's best chance for the nomination lay in a brokered convention, in which Winfield Scott and Fillmore supporters, failing to gain the necessary majority, would each turn to Webster as their second choice. For such an event to occur, however, Webster had to demonstrate substantial delegate support of his own. As the correspondence below suggests, this proved a frustrating quest.

Fig. 6.18 Wiltse, *Webster: Correspondence*, 7:301
Chapter Seven
Forms of Annotation

Choosing Annotation Forms

7.1 Editors have tremendous latitude in deciding how to annotate documents. In selecting a system of annotation, they employ methods that provide the greatest clarity in the most economical fashion. They consider the characteristics of their documents and the needs of their audience in order to create annotation that is appropriate for the content and format of the edition.

Headings

7.2 Many editors precede each document with a heading. The information included in headings varies from edition to edition. The editors of the Papers of Ulysses S. Grant listed only the name of the author or recipient of correspondence, because most letters in the edition were written by or to Grant:

To Julia Dent

Corpus Christi Texas
Jan. 12th 1846

My Dear Julia
I have just been delighted by the receiving a long and inter-

Fig. 7.2a Simon, Grant, 1:69
Many editors use headings to note both the author and the recipient of letters, as in those created by the editor of *Religious Philanthropy and Colonial Slavery: The American Correspondence of the Associates of Dr. Bray*:

**Robert Carter Nicholas to Rev. John Waring**

The editor of *A Documentary History of the Indiana Decade of the Harmony Society* included the author, the recipient, and an abstract of the document:

**JEREMIAH WARDER, JR., TO GEORGE FLOWER**: Requests information on the Harmonists; questions George Rapp’s motives; expresses concern for redemptioners purchased by the Society.

The editors of the *Naval War of 1812: A Documentary History* used larger type to introduce new topics and smaller type, in small capitals, to identify the author and recipient of each document:

**Diplomatic Arrangements for Prisoner Exchanges**

**ANTHONY ST. JOHN BAKER TO SECRETARY OF STATE JAMES MONROE**

The editors of the *Papers of Joseph Henry* included in the heading both the title of the document and its provenance:

**"RECORD OF EXPERIMENTS"**

*Henry Papers, Smithsonian Archives*

The editor of *German-American Pioneers in Wisconsin and Michigan: The Frank Kerler Letters* placed the year of composition in the heading of all letters:

**1852**

**Louise Foerster to August Frank.**

7.3 To avoid unnecessary repetition, editors may use headings to note the author and recipient of letters. This is particularly valuable in the context of a series of letters. For example, in *Religious Philanthropy and Colonial Slavery: The American Correspondence of the Associates of Dr. Bray*, the editor included headings that noted both the author and the recipient of letters. This can be seen in the following example:

**Robert Carter Nicholas to Rev. John Waring**

The editor of *A Documentary History of the Indiana Decade of the Harmony Society* included the author, the recipient, and an abstract of the document:

**JEREMIAH WARDER, JR., TO GEORGE FLOWER**: Requests information on the Harmonists; questions George Rapp’s motives; expresses concern for redemptioners purchased by the Society.

The editors of the *Naval War of 1812: A Documentary History* used larger type to introduce new topics and smaller type, in small capitals, to identify the author and recipient of each document:

**Diplomatic Arrangements for Prisoner Exchanges**

**ANTHONY ST. JOHN BAKER TO SECRETARY OF STATE JAMES MONROE**

The editors of the *Papers of Joseph Henry* included in the heading both the title of the document and its provenance:

**"RECORD OF EXPERIMENTS"**

*Henry Papers, Smithsonian Archives*

The editor of *German-American Pioneers in Wisconsin and Michigan: The Frank Kerler Letters* placed the year of composition in the heading of all letters:

**1852**

**Louise Foerster to August Frank.**
Introduction to Annotated Documents

7.3 To provide readers with more context for understanding the documents, editors may include historical essays in their editions. While such essays can be valuable in guiding readers through unfamiliar historical terrain, editors must remember that pages occupied by interpretive essays deny space to a greater number of documents. The editors of the Samuel Gompers Papers included short essays on Gompers’s life and the labor movement in their edition:

Samuel Gompers’ Participation in Workingmen’s Organizations in the 1870s

Gompers began attending meetings of the International Workingmen’s Association (IWA) in 1873 and soon became a participant in an inner circle of the Association that called itself “Die Zehn Philosophen” or the ten philosophers. By the mid-1870s “Die Zehn Philosophen” had evolved into the loosely organized Economic and Sociological Club, whose members included Gompers and many of his early associates, including David Kronburg, Ferdinand Laurrell, Fred Bloote, Louis Berliner, Henry Baer, Hugh McGregor, and J. P. McDonnell. George E. McNeill and Ira Steward, leaders of the Boston eight-hour movement, were also associated with the club.

The Economic and Sociological Club rejected what Gompers, in his memoirs, called “Socialist partyism,” and favored “trade unions, amalgamated trades unions, and national or international amalgam-

7.4 The Papers of George Catlett Marshall introduced each chapter with a brief introductory essay:

Operations Officer with the first division sent to France was the job Marshall “particularly desired.” General Sibert’s telegram meant duty in France, but Marshall did not know that the general was to command the First Division, and he did not know in what general staff capacity he was to serve. On June 8, when Sibert formally took command of the division, Marshall learned that he was to be the assistant chief of staff for Operations. (Memoirs, p. 5.)

The troops began boarding their hastily and inadequately prepared ships on the night of June 10. The next day, the division’s staff assembled for the first time aboard the United States Army Transport Tenadores, formerly a United Fruit Company banana boat. Marshall shared a stateroom with the division’s
Headnotes

7.5 Editors may use headnotes to explain or provide context for documents, as did the editors of the *Papers of Martin Luther King, Jr.*:

“*A Study of Mithraism*”

[13 September—23 November 1949]
[Chester, Pa.]

During the first semester of his second year at Crozer, King wrote this paper for Enslen’s course on Greek religion. Mithraism, a sect of Zoroastrianism characterized by the worship of Mithra as the defender of the truth, was a monotheistic mystery religion prevalent in the Roman empire before the acceptance of Christianity in the fourth century. Followers of Mithra became less common after the Roman emperors banned their cults, and Christianity gained the popularity that once belonged to Mithraism. Enslen gave the essay an A and wrote: “This is an exceedingly good paper. You have given a very complete picture of the essential details and you have presented this in a balanced and restrained way. And furthermore you know how to write. You should go a long way if you continue to pay the price.”

Fig. 7.5a Carson, King, 1:211

Editors of popular editions often use headnotes to provide readers with basic historical and contextual information that will enable them to understand and appreciate the documents. The editors of *Voices from Vietnam* preceded a letter describing the United States withdrawal from Vietnam with a headnote noting the date of the Paris peace accords and the rapid removal of troops that followed, giving readers a context for the events described in the letter:

On January 25, 1973, the Paris peace accords were signed, officially ending direct U.S. participation in Vietnam as well as military assistance to South Vietnam. Over the next sixty days, all American combat troops were withdrawn from the country. In the following letter, Weidner describes the final months of the war.

January 29, 1973

Dear Mom & Dad,

The war has never been so bad as since the peace was declared. You’ve heard already—TSN was rocketed Sunday A.M. Also Bien Hoa, Pleiku,

Fig. 7.5b Stevens, Voices, 130

7.6 Editors may place endnotes at the bottom of the page or in a separate bibliography at the end of the work.

7.7 More common in numbered sequen-
tions of the Letters and manuscripts and textual in-

7.8 When an essay has been divided into sections, place endnotes at

end.
Footnotes and Endnotes

7.6 Editors may place annotation in footnotes—that is, notes placed at the bottom of the page—such as those employed by the editors of the Colonial Records of North Carolina: Second Series:

7GO 116.1 here includes the following: "At a Council held at Brunswick the third day of June 1766, Present His Excellency the Governor, James Hasset, John Buteford, Lewis Henry DeBouet, William Dey, Robert Palmer, Benjamin Heman, [Blank]."

7FPs. 542, 549-546, below.

Fig. 7.6 Parker, North Carolina, 9.161

7.7 More commonly, notes directly follow the text of a document and are numbered sequentially, with new numbering for each document's notes. The editors of the Letters of Delegates to Congress used an unnumbered note for provenance and textual information and numbered notes for contextual information:

LB (DNA: PCC, item 16). There is no indication in Hanson's presidential letterbook why this appeal to the states was addressed only to the governor of New York.

1. The enclosed resolution recommending that the state legislatures "make suitable provision for staying all suits" instituted against Continental officers "for debts contracted by them for supplies furnished or services rendered to the United States," was adopted in response to a letter of February 25 from Quartermaster General Timothy Pickering. See CCD2, 22:138–39; and PCC, item 19, 5:171–73, item 192, fols. 89–98. Pickering's letter had been referred the same day to a committee consisting of Elias Boudinot, Daniel Carroll, Thomas McKean, Edmund Randolph, and John Morin Scott, whose report was in turn referred on March 7 to a second committee consisting of Abraham Clark, Samuel Livermore, and James Madison. For the work of these committees and Madison's authorship of the committee recommendation eventually adopted by Congress, see Madison, Papers (Hutchinson), 4:91–92.

Fig. 7.7 Smith, Delegates, 18:416

7.8 When an edition consists of a single document, such as the Diaries of David Lawrence Gregg: An American Diplomat in Hawaii, editors sometimes place endnotes at the back of a volume.

Notes

NOTES

INTRODUCTION

1. Biographical material on Gregg's family was collected from the letters of David L. Gregg, Jr., of Glendale, Calif., to Ralph S. Kuykendall, 1923–1939, in Gregg Folder, Hawaiian Collection Letters File, University of Hawaii; hereafter cited as Gregg Folder; St. Patrick's Church records, La Salle, Ill., in Florence Clarke to Pauline King Joergen, November 8, 1973.
D. L. Gregg to William Gregg, August 26, 1856, in Gregg Letter Books, Gregg Collection, hereafter cited as Gregg LB and Gregg Coll.

2. The Gregg Collection is made up of diaries, letter books, and letter files and is on deposit at the Archives of Hawaii, hereafter cited as AH.

Fig. 7.8a King, Gregg, 503

Editors who want to keep their annotation unobtrusive may also choose to use endnotes, as did the editors of Abraham Lincoln: Speeches and Writings, who referred readers to page and line numbers rather than using numbered notes:

**Notes**

In the notes that follow, the reference numbers denote page and line of this volume (the line count includes item headings). No notes are made for material included in a standard desk-reference book. Correspondents and names mentioned by Lincoln are identified only when they are essential to an understanding of the text. Prefatory and end notes within the text are Lincoln’s own. For more detailed notes and references to other studies, see The Collected Works of Abraham Lincoln, 8 volumes plus index (New Brunswick: Rutgers University Press, 1953–55), edited by Roy P. Basler, Marion Dolores Pratt, and Lloyd A. Dunlap; The Collected Works of Abraham Lincoln: Supplement 1842–1865 (Westport, Connecticut: Greenwood Press, 1974), edited by Roy P. Basler; Lincoln Day by Day: A Chronology, 3 volumes (Washington, D.C.: Lincoln Sesquicentennial Commission, 1960), edited by Earl Schenck Miers, William E. Baringer, and C. Percy Powell; and Mark E. Neely, Jr., The Abraham Lincoln Encyclopedia (New York: McGraw-Hill, 1982).

7.1 Mary S. Owens (1808–77), the only woman besides Mary Todd whom Lincoln is known to have courted, came to New Salem from Kentucky in 1836 for an extended visit with her sister, Elizabeth Owens Abell.

9.5 Mr. Linder] Usher F. Linder (1800–76), an attorney and Democratic legislator. In 1858 Linder became a Whig, but his hatred of abolitionism caused him to return to the Democrats in the 1860s. He appeared with and against Lincoln in many law cases.

Fig. 7.8b Lincoln, Speeches and Writings, 858

Endnotes may also be placed at the end of a short magazine article, such as was done by the editor of "A Thousand Mile Motor Trip" published in Nebraska History:
O.K. here. So end[s] the Thirteenth Day and the Thousand Mile Motor Trip. We traveled about 800—eight hundred miles. We had seen a lot of country; had enough new experiences to suit anyone. But like little Pollyanna we are Glad; the Buick is battle-scarred; we are motor veterans!!!

Notes

2 Ibid.
4 Polk's Lincoln City Directory, 1925 (Kansas City: R. L. Polk and Co., 1925), 53.
5 The 1910 Cornhusker Yearbook lists Maggie May Gehrie as a member of the Y.W.C.A., English

Fig. 7.8c Koelling, Nebraska. 78:27

Intratextual Annotation

7.9 Editors may also choose to place annotation within the text of documents. Editorial interpolations and brief textual comments may be indicated through words or symbols and placed within sentences (see Chapter 5), but extended annotation may also appear within the text of documents, as it did in the Papers of John C. Calhoun, whose editors placed intratextual editorial notes in italic type in brackets:

received there, the abolitionists had gained all they wanted; and so long as they were permitted to come there he would take the liberty to speak of them in the terms they deserved.

[Bedford Brown of N.C. praised the resolutions as showing the weakness of abolition sentiment in Me. and remarked that it was impracticable "to expect a State to eradicate every folly or infatuation from the minds of all its citizens." He further remarked that, while the language of the petitions "was bad enough," he did not believe any had used the words "pirates, robbers, and murderers" for
the people of the South. Calhoun thereupon asked Thomas Morris of Ohio for the petitions that he had presented and then withdrawn earlier in the session. Morris gave Calhoun some petitions but said he could not consent for them to be used "at that time."

Mr. Calhoun said he was utterly astonished at the remarks of the gentleman from North Carolina. These charges were made when the Ohio petitions were presented and read, and in the gentleman's

Fig. 7.9 Meriwether, Calhoun, 13:145

Specialized Directories

7.10 Editors sometimes find that a glossary or directory is more convenient for readers and requires less space than separate notes, especially when the persons, places, or terms to be described recur throughout the edition. The editors of Circular Letters of Congressmen to Their Constituents created biographical entries that provide brief identification of each congressman whose letters appear in the volumes:

DIRECTORY OF CONGRESSMEN

The following directory contains identifications of the members of Congress represented in these volumes and a listing, under each congressman, of the circular letters here published. More detailed biographical information can be found in the Biographical Directory of the American Congress, 1774–1971 (Washington, D.C., 1971). A party identification as a Republican refers to the Jeffersonian Republican party. Since the old party labels of Federalist and Republican were not very meaningful for some members who entered Congress after 1815, party identification has not been made for all members.

ALLEN, ROBERT (1778–1844). Republican representative from Tennessee in 16th–19th Congresses, 1819–1827; not a candidate for reelection.


Fig. 7.10a Cunningham, Circular Letters, 1:111

Editors may explain in an introductory headnote how they produced specialized directories, and may detail the problems they encountered. The editors of the Diary of Elizabeth Drinker: The Life Cycle of an Eighteenth-Century Woman preceded their biographical directory with such an introduction:

Elizabeth Drinker, nearly fifty-year pyramid of a woman, with her husband and children, and many only spoken of identifying as a quantity of biographical amount and natural Whose feasible as occupation, proper dates are prefaced and in ambiguous. They're a "Cong. age," and his birth, death, and information obtained attempted to any birth or death details.

Because of the certain problems, the family members, even within a single name in a family, and matters, ED k's, and similar names. The sometimes unclear has in mind. (The fact makes it possible that three equally strong Jacob Taylor is in the directory positive identification of Northern D District supporting evidence.)

Spelling variation under the current included as given in contemporary form.

Since the dictionary has two names: the
Biographical Directory

Elizabeth Drinker recorded the names of several thousand people over the nearly fifty-year period during which she penned her entries. Some, such as her husband and children, were mentioned almost daily, others less frequently, and many only sporadically. The directory has been compiled for the purpose of identifying as many people as possible in all three categories. A surprising quantity of biographical information has surfaced in the attempt, although the amount and nature of that information vary from person to person.

Where feasible, birth, death, and marriage dates have been included, as well as occupation, public roles, religion, and relationship to the Drinkers. Some dates are prefaced by a c., which means that the genealogical source was ambiguous. Thomas Coombe, for example, "d. 1799 in the 79th year of his age," and his birthdate, therefore, is listed as "c. 1720." Although some of the information obtained from the various sources is suspect, the editor has attempted to amend the reference material only when ED herself inscribes a birth or death date that conflicts with secondary evidence.

Because of the length of the diary and its eighteenth-century qualities, certain problems arise of which the reader should be aware. Most important, family members with the same last name frequently had the same given name even within a single generation, and as Drinker herself realized, "several of one name in a family occasions confusion and mistake—oftimes." To compound matters, ED knew a number of distantly related or unrelated people with similar names. Thus, when she refers simply to Ann Emlen, for example, it is sometimes unclear whether she has Ann Emlen Mifflin or Ann Emlen Pleasant in mind. (The former married in 1788, the latter in 1796.) Where the context makes it possible to narrow the choice to a specific person, or where two or three equally strong candidates seem likely, the alternatives have been included in the directory. Where, however, there are several possibilities but no clue to a positive identification, the name has been omitted from the directory altogether. Jacob Taylor is a case in point: three men of that name were members of the Northern District Monthly Meeting during the late eighteenth century, but no supporting evidence exists to indicate which one appears in the diary.

Spelling variations have been cross-indexed. The main entry may be found under the currently accepted spelling of the name; ED's departures are included as guides to the main entry, since some bear little similarity to contemporary references.

Since the diary covers nearly five decades, many women appear in it under two names: their name at birth and their name after marriage. In the directory,
the full identification may be found under the name by which a subject first appears in the diary. A cross-reference under the married name will direct the reader to the name at birth if the full entry appears at the latter location. A word of caution, however: given the relative similarity of first names (the profusion of Sarahs, Marys, and Anns is bewildering), and the frequency of intermarriage among people who answered to no more than six different surnames, it is possible that in a few cases the cross-reference may not be to the same person, as intended, but rather to a mother or cousin with the same name.

The directory entry for each person always refers to the subject's formal name. ED, however, referred to many acquaintances by their diminutives. The following list, therefore, provides the reader with the nicknames that appear most often in the diary:

Ann
Nancy
Catherine
Caty, Katy

Fig. 7.10b Crane, Drinker. Life Cycle, 305

The editors of the Samuel Gompers Papers included a biographical glossary at the back of each volume of their edition:

**INDIVIDUALS**

**Appel, George W.** (b. 1860?), served as the general secretary of the Metal Workers' National Union of North America from 1886 to 1889. Born in Maryland, he worked as a silver plater and brass finisher in Baltimore.

**Arch, Joseph** (1826-1919), was a founder (1872) and leader of the National Agricultural Labourers' Union and a Member of Parliament (1885-86, 1892-1900).

Fig. 7.10c Kaufman, Gompers, 1:473

The editor of The Children of Pride provided a list of "principal characters" at the front of his volume to simplify identification of the individuals who wrote and received letters or who were discussed in the documents:

**The Principal Characters**

**At Maybank, Montevideo, and Arcadia**

Rev. Dr. Charles Colcock Jones, a retired Presbyterian clergyman
Mary (Jones) Jones, his wife
In Savannah

Charles Colcock Jones, Jr., a lawyer, elder son of Dr. and Mrs. C. C. Jones
Ruth Berrien (Whitehead) Jones, his wife
Julia Berrien Jones, their daughter

Fig. 7.10d Myers, Children, 3

7.11 The editor of the Papers of George Mason included a joint biographical-geographical glossary in his edition:

FAIRFAX PARISH: Fairfax County; divided from Truro Parish in 1764; contained Alexandria and vicinity; built Falls (1767) and Alexandria (Christ's) churches.

FALCONER, Nathaniel: Capt. of ship Friendship on Philadelphia-to-London run, 1760; sought and may have obtained command in U.S. navy during Revolution.

FALLS WAREHOUSE: (Falls of the Potomac), Fairfax County; about one mile north of Difficult Run; built by Thomas Lee in 1742 as a public tobacco warehouse, it had little success.

FARGOSON, John: Fairfax County freeholder; Truro Parish vestry, 1733-1744.

Fig. 7.11 Rutland, Mason, 1.11

7.12 For editions that contain numerous technical terms, a glossary may be a useful addition. The editors of the Papers of General Nathanael Greene included a glossary of eighteenth-century military terms:

A GLOSSARY OF MILITARY TERMS

ABATIS
A barrier of felled trees, with limbs pointing toward the enemy, usually temporary.

ACCOUNT
A financial record.

ACCOUNTREMENTS
Soldiers' outfits, usually not including clothing or weapons.

ARTILLERY PARK
An encampment for artillery.

ARTILLERY TRAIN
An army's collection of cannon and material for firing them.

Fig. 7.12a Showman, Greene, 8:xxxvii
The editors of the *Samuel Gompers Papers* included a glossary of labor organizations:

The *Journeymen Bakers’* National Union of the United States was organized in 1886, participated in the formation of the AFL that year, and was chartered by the AFL in 1887. In 1890 it adopted the name *Journeymen Bakers’ and Confectioners’* International Union of America and, in 1903, *Bakery and Confectionery Workers’ International Union of America.*

The *Bricklayers’ and Masons’* International Union of America was organized in 1865. It did not affiliate with the AFL until 1916.

Fig. 7.12b Kaufman, *Gompers,* 1:507

The editors of the *Documentary History of the First Federal Congress* included a glossary of legislative and legal terms:

**GLOSSARY**

**ACT:** (1) A bill that has been passed by one or both Houses of Congress. (2) A bill that has been passed by both Houses of Congress and signed by the President or passed over his veto.

**AMENDMENT:** (1) A proposal to alter a bill, act, amendment, or resolution after it has been formally introduced. Amendments vary in importance from slight word changes to major substantive alterations. To be adopted, an amendment must be agreed to by a majority of the Members voting. (2) A change in the Constitution. Such an amendment is usually proposed in the form of a joint resolution of Congress, which may originate in either House. If passed, it does not go to the President for his approval but is submitted directly to the states for ratification.

**BILL:** A proposal of specific legislation presented to Congress for enactment into law. Bills may originate in either House; two exceptions, however, are those for raising revenue, which, according to the Constitution, must originate

Fig. 7.12c De Pauw, *Congress,* lxvii

Tables, Lists, and Calendars

7.13 Tables and charts may be created to distill important information, as does this table created by the editors of the *Documentary History of the*
Ratification of the Constitution to show the attendance of the delegates to the Georgia Ratification Convention:

Convention Roster and Attendance Record  
25 December 1787-5 January 1788

<table>
<thead>
<tr>
<th>Name</th>
<th>Dec 25</th>
<th>Dec 26</th>
<th>Dec 27</th>
<th>Dec 28</th>
<th>Dec 29</th>
<th>Jan 1</th>
<th>Jan 2</th>
<th>Jan 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHATHAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>William Stephens</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Joseph Habersham</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Thomas Gibbons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPWORTH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jenkin Davis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nathan Brownson</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Caleb Howell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BUTLER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edward Telfair</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Henry Todd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>George Walton</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 7.13 Jensen, Ratification, 3:270

7.14 Genealogical tables provide readers with a quick and easy method of understanding complex family relations, as does this chart included in The Secret Eye: The Journal of Ella Gertrude Clanton Thomas:

The Thomas Family

Joseph and Rebecca Thomas were parents of John Thomas, who settled in Lunenburg County, Virginia, in 1632. Joseph married Elizabeth Kidd of Lunenburg County, Virginia, a descendant of Ed. George Edwards, owner and planter of George Washington's middle, Mary Hall. Their children and descent were

Fig. 7.14a Burr, Secret Eye, xviii
The editor of "Dear Master": Letters of a Slave Family used another format for showing genealogical relationships:

**Skipwith Family**

Children of Lucy Nicholas, who cohabited with Jesse Skipwith
Peyton Skipwith (d. 1849)
  m. Lydia Randall (1804?–1834)
  Diana (1822–1844?) m. Moore James (b. 1820?)
  Matilda (b. 1824)
  m. Samuel Lomax (1823?–1850)
  infant (died)
  Eliza Adala
  Lydia Ann
  child (died)

Fig. 7.14b Miller, Dear Master, 20

7.15 A calendar of the years covered by the documents in an edition assists readers in following the chronology of events, especially when many of the documents refer regularly to days of the week without specifying dates. This one was included in the *Documentary History of the First Federal Elections*:

**CALENDAR FOR THE YEARS 1788–1790**

<table>
<thead>
<tr>
<th>JANUARY</th>
<th>FEBRUARY</th>
<th>MARCH</th>
<th>APRIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3</td>
<td>1 2 3</td>
</tr>
<tr>
<td>5 6 7 8</td>
<td>9 10 11</td>
<td>12 13 14 15</td>
<td>16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31</td>
</tr>
<tr>
<td>MARCH</td>
<td>JUNE</td>
<td>JULY</td>
<td>AUGUST</td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEPTEMBER</td>
<td>OCTOBER</td>
<td>NOVEMBER</td>
<td>DECEMBER</td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 7.15 Jensen, Elections, l:xxx

7.16 Chronologies help readers understand the relationship of documents to the broader sweep of events. The editors of the *Papers of Thomas A. Edison* provided a chronology of Edison's inventions for the period covered in each volume:
<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1811</td>
<td>6 January. Born to Charles Pinckney and Relief Jacob Sumner in Boston.</td>
</tr>
<tr>
<td>1821-1826</td>
<td>Attends Boston Latin School.</td>
</tr>
<tr>
<td>1835-1837</td>
<td>Lectures at Harvard Law School; edits Sumner's Reports.</td>
</tr>
<tr>
<td>1837-1840</td>
<td>European tour.</td>
</tr>
<tr>
<td>1839</td>
<td>24 April. Charles Pinckney Sumner dies.</td>
</tr>
<tr>
<td>1840</td>
<td>3 May. Returns to United States, resumes Boston law practice.</td>
</tr>
</tbody>
</table>

The editor of the two-volume Selected Letters of Charles Sumner printed a chronology spanning Sumner's entire life, focusing on his education, orations, and political career.

Fig. 7.16a Jenkins, Edison, 2:xxx

Fig. 7.15 Jensen, Elections, 1:xxx

A relationship of documents to papers of Thomas A. Edison period covered in each volume:
Maps and Illustrations

7.17 Editors include maps to provide readers with a better understanding of the geographical areas discussed in documents. The editors of the Papers of Ulysses S. Grant provided a map to explain the location of Union and Confederate units around Richmond and Petersburg:

![Map of Richmond and Petersburg]

*Area of Operations around Richmond and Petersburg*

*From Horace Greeley, The American Conflict: A History of the Great Rebellion (Hartford, Conn., 1868), II, 594*

Fig. 7.17 Simon, Grant, 12.115
Editors may provide readers with illustrations as a way to enhance the information provided within documents. Information about the medium, artist, and institution holding copyright on the image should be provided in a caption or in a note. Commonly used illustrations include portraits and photographs of individuals, facsimiles of significant documents, and pictures of buildings or scenes associated with the topic. When presenting illustrations to complement documents, editors should strive to include only relevant materials.

Annotation in Microform and Electronic Editions

Microform editions may be annotated with many of the same types of notes found in a book edition, although the microform format often makes it difficult to refer to tables, charts, or other reference tools located in the first or last frames of an edition. As this example from the microfiche supplement to the Documentary History of the Ratification of the Constitution demonstrates, a document in microform can be accompanied by headings and footnotes containing provenance notes, just like a document in a book edition:

5 Williamsburg Meeting, 6 October, 1787

At a meeting of the Proprietors and other England to the City of Williamsburg at the Courthouses of said City, the 6th day of October 1787.

The Proprietors of the City of Williamsburg. To the Proprietors and other England to the City of Williamsburg.

RESOLVED, that this is the opinion of this meeting of the Proprietors, and that the same is true and acceptable to this Body.

RESOLVED, that this is the opinion of the Proprietors, and that the same is true and acceptable to this Body.

RESOLVED, that this is the opinion of the Proprietors, and that the same is true and acceptable to this Body.

RESOLVED, that this is the opinion of the Proprietors, and that the same is true and acceptable to this Body.

Fig. 7.17 Simon, Grant, 12:115
Microform targets—cards containing annotation that are filmed alongside a document—are more common. The editors of the Guide to the Thomas A. Edison Papers: A Selective Microfilm Edition described different ways their edition employed targets:

**Targets**

Targets are editorial aids that appear on the microfilm to assist the reader in using the materials in the collection. Targets introduce each series, each major subseries, and the individual volumes and folders that appear in the edition. The series and subseries targets are essentially more detailed versions of the Series Notes on pages 19-30 of this guide. The target preceding each bound volume provides information about authorship, inclusive dates, and number of pages, as well as a brief description of the character and contents of the volume and, wherever appropriate, cross-references to related materials. The target preceding each folder in the Document File Series lists the types of documents in the folder, any variations from chronological order, and cross-references to related materials. The target also provides a brief description of materials in the folder that were not selected for filming. Where the contents of a particular folder were not filmed, an explanatory target appears on the microfilm.

Occasionally, explanatory targets accompany individual documents in order to: (1) explain relationships among documents (for example, the target ENCLOSURE; (2) describe multiple versions of a document (for example, the targets TRANSLATION FOLLOWS and TRANSLATION; (3) supply missing information (for example, the date or author of a letter); and (4) note the existence of missing information (for example, the target INCOMPLETE).

Fig. 7.19b Jeffrey, Edison, part 1:15

The editors of the Papers of Eugene V. Debs filed a rectangular card at the top of each document; it provided the date, author and recipient, type of document, length of document, and provenance:

1905 March 17

EVD to Marguerite Bettrich Debs
ALS 4pp

Int'l, Debs Coll.

Fig. 7.19c Corrandaht, Papers of Debs, reel 1:445

The editors of the microfilm edition of the Emma Goldman Papers: A Microfilm Edition also used targets describing the type of document and its date, place, author, recipient, number of pages, and physical size, as well as its provenance and copyright holder:

**The Emma Goldman Papers**


Permission to reproduce or quote in any form must be obtained from the Schlesinger Library, Radcliffe College.

Fig. 7.19d Falk, Goldman, reel 2:14
7.20 Hypertext links make special types of annotation possible in electronic editions. By selecting certain highlighted areas on the screen, readers may bring forth text, pictures, sounds, or visual images related to the particular image or passage. For further information on the creation and annotation of electronic editions, editors should consult the home page of the Model Editions Partnership (URL: http://mep.cla.sc.edu/), a consortium of seven historical editing projects working in conjunction with the Text Encoding Initiative and the Center for Electronic Texts in the Humanities to establish guidelines for electronic documentary editing and to develop model electronic editions. For links to home pages of various historical editions, see the Web site of the Association for Documentary Editing (URL: http://etext.virginia.edu/ade/).

Abbreviations

7.21 Editors use many symbols and abbreviations to convey information about documents. Abbreviations are commonly used to refer to the archival repository of documents, the form of the source text, and frequently used reference works. Because there are no standardized abbreviations for provenance or short titles, editors must borrow these symbols from other editions or create their own. When referring to provenance, modern editions often use the symbols established by the Library of Congress and found in the most current edition of Symbols of American Libraries (Washington, D.C.: Cataloguing Distribution Services). Editions that began to be published before these abbreviations were standardized developed...
their own systems or used one of numerous other systems established regionally
or locally by libraries or bibliographers.

7.22 Documentary editions commonly describe the form and genre of documents by using abbreviations devised for individual projects. For example, the editors of the *Papers of Thomas Jefferson*, the *Papers of Woodrow Wilson*, and the *Papers of Thomas A. Edison* employed different sets of abbreviations, highlighting different aspects of documents:

- **Dft** draft (Usually a composition or rough draft; later drafts, when identifiable as such, are designated "2d Dft," &c.)
- **Dupl** duplicate
- **MS** manuscript (arbitrarily applied to most documents other than letters)
- **N** note, notes (memoranda, fragments, &c.)
- **PoC** polygraph copy
- **PrC** press copy
- **RC** recipient's copy
- **SC** cylograph copy
- **Tripl** triplicate
- **FC** file copy (Applied to all forms of retained copies, such as letter-book copies, clerks' copies, &c.)
- **Tr** transcript (Applied to both contemporary and later copies; period of transcription, unless clear by implication, will be given when known.)

Fig. 7.22a Boyd, *Jefferson*, 1:xxxv–xl

### ABBREVIATIONS

- **AL** autograph letter
- **ALS** autograph letter signed
- **att.** attached
- **enc.** enclosure, enclosed
- **env.** envelope
- **hw** handwritten, handwriting
- **JRW** Joseph Ruggles Wilson
- **JWW** Janet Woodrow Wilson
- **misc.** miscellaneous
- **sh** shorthand
- **shLS** shorthand letter signed
- **shP(S)** shorthand postal (signed)
- **WW** Woodrow Wilson
- **WW(hw)(S)** Woodrow Wilson handwriting or handwritten (signed)
- **WWsh** Woodrow Wilson shorthand
- **WWshLS** Woodrow Wilson shorthand letter (signed)
- **WWshP(S)** Woodrow Wilson shorthand postal (signed)

Fig. 7.22b Link, *Wilson*, 1:xxxi

Symbols used to describe documents:

- **A** autograph
- **H** handwriting
- **P** printed
- **T** typed

The first describes the content; the second, its physical characteristics. The third describes the document. The fourth, if any, describes its current state.
7.23 The editors of Freedom: A Documentary History of Emancipation listed the different types of abbreviations used throughout the edition, including those describing the physical characteristics of documents, the provenance of collections, general and military abbreviations, and short-title references to published sources.

**SYMBOLS AND ABBREVIATIONS**

**SYMBOLS USED TO DESCRIBE MANUSCRIPTS**

Symbols used to describe the handwriting, form, and signature of each document appear at the end of each document.

The first capital letter describes the handwriting of the document:
- A autograph (written in the author’s hand)
- H handwritten by other than the author (for example, by a clerk)
- P printed
- T typed

The second capital letter, with lower-case modifier when appropriate, describes the form of the document:
- L letter
- D document
- E endorsement
- W wire (telegram)
- C copy
- P press copy
- D draft
- F fragment

Fig. 7.22c Jenkins, *Edison*, 1:ixix
The third capital letter describes the signature:
S  signed by the author
Sr signed with a representation of the author's name
I  initialed by the author
     no signature or representation

For example, among the more common symbols are: ALS (autograph letter, signed by author), HLS (handwritten letter, signed by author), HLSr (handwritten letter, signed with a representation), HLcSr (handwritten copy of a letter, signed with a representation), HD (handwritten document, no signature).

ABBREVIATIONS FOR RECORD GROUPS IN THE NATIONAL ARCHIVES OF THE UNITED STATES

RG 11 General Records of the United States Government
RG 15 Records of the Veterans Administration
RG 21 Records of District Courts of the United States

* * * *

SHORT TITLES

**Freedom**


**Navy Official Records**


**Official Records**

Military and Other Abbreviations That Appear Frequently in the Documents

A.A.A.G. Acting Assistant Adjutant General
A.A.G. Assistant Adjutant General
A.C. Army Corps
A.C. Assistant Commissioner (Freedmen's Bureau)
Act., Actg. Acting
A.D. African Descent
A.D.C. Aide-de-Camp
Adjt. Adjutant
Agt. Agent
A.G.O. Adjutant General's Office
A.Q.M. Assistant Quartermaster
Asst. Assistant
A.S.A.C. Assistant Subassistant Commissioner (Freedmen's Bureau)

Fig. 7.23 Berlin, Freedom, ser. 1,1,xxxiv–xxxvi

Cross-References

7.24 A convenient way to save space and avoid excessive duplication of information is to refer readers to documents or notes provided in an earlier edition, especially a microform edition. Cross-references to microform editions in particular allow editors to lead readers to original documents and additional sources that they may not have published. By providing cross-references to earlier volumes, editors lead readers to their previous research and annotation, avoid repetition, and focus efforts on providing new information.

7.25 The editors of the Papers of Robert Morris regularly referred readers to earlier documents and notes:

1. RM had taken a weekend "excursion into the Country." See Diary, July 19, 1782.
2. On the congressional committee to investigate the Office of Finance, see Diary, June 12, 1782, and notes. James Duane's letter is dated July 19.
3. See Diary, February 14, 1782, and notes.
4. Arthur Lee was on the committee mentioned in note 2 above.
5. See RM to Duer, July 23, and notes.
6. On Crawford's case, see Diary, July 19, 1782, and notes.
7. On Campbell's application, see Diary, May 8, 1782, and notes.

Fig. 7.25a Ferguson, Morris, 6:4
In the following example from the Papers of Daniel Webster: Correspondence, the footnote following the reference to the Crafts’ Case directed the reader to an earlier note in the same volume that had described the circumstances of the case:

also to state, that I knew the correct opinions of Mr Sanger, on this subject, when the “Crafts’ case” came up. And I cannot but trust that

6. See above, headnote preceding
DW to Fillmore, November 5, 1850,
p. 177.

Headnote, p. 177:

Threatened and actual resistance to the new Fugitive Slave Law was not restricted to Pennsylvania. In Boston a “vigilance” committee led by Theodore Parker sprang to the aid of a fugitive couple, William and Ellen Craft, whose Macon, Georgia, master, Robert Collins, sought to reclaim them. Collins’s two agents were harassed in Boston and briefly clapped in jail, until they left the city empty-handed. On November 7, prior to the Crafts’ leaving for England to circumvent the law, they were married, in a highly charged ceremony conducted by Reverend Parker. See John Weiss, ed., Life and Correspondence of Theodore Parker (New York, 1864), 2: 94–102, and below, pp. 199–200, Theodore Parker to DW, December 12, 1859.

Fig. 7.25b Wiltsie, Webster: Correspondence, 7:177,223,224

The editors of the Papers of Andrew Johnson used cross-references to guide readers to earlier volumes in which individuals had already been identified in biographical notes:

2. For Thomas, see Johnson Papers, V, 367n.

Fig. 7.25c Graf, Johnson, 6:141

Researching and Citing Annotation Sources

Although editors possess a wealth of knowledge, most projects still entail considerable research. Editors will want to provide their readers with some indication of the sources used to compile their notes, but the exact method depends on the audience of the edition, the number and type of sources used, and the judgment of the editors.

Editors may provide full citations listing the sources used to write their notes. For example, the notes of the Papers of George Calhoun Marshall provided citations for books, newspapers, and primary documents:
...new Fugitive Slave Law was a "vigilance" committee led a fugitive couple, William and F. Robert Collins, sought to re-assessed in Boston and briefly handcuffed. On November 7, d to circumvent the law, they forcibly conducted by Reverend Correspondence of Theodore below, pp. 189–190, Theodore Webster: Correspondence, 7:177,223,224

1 used cross-references to guide us had already been identified in papers, V, 367n.

Fig. 7.25c Graf, Johnson, 6:141

Citation Sources

Knowledge, most projects still strive to provide their readers with adequate notes, but the exact type, the number and type of sources used to write their George Catlett Marshall provided documents:

7.28 Editors decide on the format of citations, including the level of detail they will include. The Papers of Joseph Henry cited secondary works as follows:


Fig. 7.28a Reingold, Henry, 1:277

The Collected Papers of Albert Einstein borrowed their citation format from scientific texts, using author-date citations, with a list of full citations at the back of the volume:

[^2]: See Thomson 1851. As in Doc. 37, \( \kappa \) denotes the thermal conductivity, \( T \) the temperature, Marshall's Diplomarbeit, which she hoped to use as the basis of her doctoral dissertation, was on heat conduction (see Docs. 63, 75, note 5, and 119).

LITERATURE CITED


Fig. 7.28b Stachel, Einstein, 1:258,405
7.29 When frequent reference will be made to a small number of published and unpublished works, editors often devise short titles or abbreviations for such works. At the front or back of the volume, they then provide a list of short titles and abbreviations, along with the corresponding full citations. For example, the Collected Papers of Charles Willson Peale and His Family used abbreviations to indicate secondary works used in the notes:

1. Robert Eden (1741–84) was Maryland’s last colonial governor. DAB.
2. A new state house was nearing completion at this time, and the ostensible reason for CWP’s presentation of his first painting of Pitt was to contribute to the building’s adornment. The painting now hangs in the State House, Annapolis. CWP, p. 110.

Fig. 7.29a Miller, Peale, 1:132.

Short Titles of Sources

CWP

DAB

Fig. 7.29b Miller, Peale, 1:x.

7.30 Editions may develop methods that address the specific circumstances of their subject or the available primary and secondary sources. For example, information about many of the individuals named in the Booker T. Washington Papers is accessible only in the Tuskegee Institute archives. Therefore, source notes do not list the specific locations of sources but instead mention the years when individuals were associated with the Tuskegee Institute:

Fig. 7.30 Harlan, Washington, 4:227

7.31 Editors may choose to provide no citations for the information contained within their notes. For example, the Documentary History of the Ratification of the Constitution attached no secondary source citations to its notes because the large array of sources used by the editors would have made the notes unwieldy and added considerably to the space required for annotation:
small number of published titles or abbreviations for they then provide a list of spounding full citations. For Peale and His Family used he notes:

overner. DAB.
and the ostensible reason for iluse to the building’s adorn-
dis. CWP, p. 110.

Fig. 7.29a Miller, Peale, 1:132


 здоровь American Biogra-
8–96).

Fig. 7.29b Miller, Peale, 1:131

he specific circumstances try sources. For example, the Booker T. Washington chives. Therefore, source

Cleveland dentist, taught e in 1878. After her hus-
ital at Tuskegee Institute,

instead mention the years institute:
10 Harlan, Washington, 4:227

or the information con-
re History of the Ratifica-
citations to its notes s would have made the quired for annotation:

1. RC, Gratz Collection, History. Printed: CC-537 (a longer excerpt). In an omitted portion of the letter, Griffin speculated about the prospects of ratification in several states. Griffin (1748–1810), a lawyer, represented Lancaster in the House of Delegates, 1777–78, 1786–87. He was a delegate to Congress, 1778–80, 1787–88 (president, 1788), and a member of the Continental Court of Appeals in Cases of Capture, 1780–87. FitzSimons (1741–1811), a Philadelphia merchant and a member of the Pennsylvania Assembly, signed the Constitution in the Constitutional Convention.

2. James Madison, Griffin's fellow Virginia delegate to Congress, probably supplied him with this information about Lee and Page. (See letters to Madison from George Washington, 10 January; Archibald Stuart, 14 January; and Tench Coxe, 23 January, all above.)

Fig. 7.31 Jensen, Ratification, 8:382
Electronic Textual Editing: Documentary Editing
[Bob Rosenberg]

Contents
- Database
- Images
- Text

The most important point to be made about any digital documentary edition is that the editors' fundamental intellectual work is unchanged. Editors must devote the profession's characteristic, meticulous attention to selection, transcription, and annotation if the resulting electronic publication is to deserve the respect given to modern microfilm and print publications. At the same time, it is abundantly clear that a digital edition presents opportunities well beyond the possibilities of film and paper. In the case of the Edison Papers, both the microfilm and print editions were well under way, and the electronic edition was seen as a means to combine and extend the work done without significantly altering the established editorial principles.

There are a number of principal considerations facing anyone planning to create an electronic edition of historical documents. If the documents are presented as images, the primary concerns will be construction of a database and creation of those images; if the documents are all transcribed, then preparation and presentation of the text will be foremost. The Edison Papers is working to combine images and text, and I hope that a careful examination of some avenues explored and lessons learned in that process will be helpful to anyone fortunate enough and bold enough to undertake such a task.

The Edison Papers is unusual in several respects, the first and most striking being the size of the archive from which it draws. When the project was launched at the end of the 1970s, it was estimated that the collection at the Edison National Historic Site comprised about 1.5 million pages. Within a few years that estimate had grown to approximately five million pages, which made
the tens of thousands of Edison pages in other archives, libraries, and repositories seem easily manageable. It was clear that the two editions projected at that time—microfilm and print—would be selective, with the microfilm including about ten percent of the holdings, and the printed volumes including perhaps two or three percent of that ten percent. A second unusual aspect of the Edison corpus is the central importance of drawings and even physical artifacts to an understanding of its subject's work, which is a direct consequence of Edison's being an inventor, fresh territory for documentary editing. Nevertheless, despite these differences and others more subtle, we decided at the outset to hew as closely as possible to the standard practices of documentary editing.

The original plan was to have the microfilm proceed ahead of the print edition, since the microfilm could capture years of Edison's life while the book editors were dissecting it one day at a time. This worked admirably, even if at first it had the book editors champing at the bit, teasing insight out of reams of photocopies. Organizing, comprehending, selecting, and filming the documents was—and remains—a truly Herculean task. When the first part of the collection was published on film in 1985, the book editors had Edison's early work at their fingertips. At the same time, two crucial pieces for the foundation for the electronic edition had been unwittingly put in place. First was the structure of the descriptive data recorded about each document; second was the high quality of the microfilm itself, which would later allow the creation of excellent digital images.

**Database**

It had been impressed on the project organizers that the only way to control a collection of this size was with an electronic database. Fortunately, the Joseph Henry Papers had already started blazing a trail into that mysterious territory. Using that experience and knowledge as a foundation, the Edison Papers created a database that would prove two decades later to be the heart of their electronic edition. At its inception, the database served two functions: it was the raw material from which was created the detailed, item-level printed index that accompanied the microfilm, and it also contained information about the organization and contents of the documents which was used for in-house research.

The first incarnation of the database lived on a university mainframe and was written by a hired programmer in Fortran 77. The main table had 24 data fields:

**Group**

A two-character field identifying the record group to which the document belonged ('A' for accounts,'PN' for pocket notebooks, etc.)

**Location**

An eight-character field specifying the position of the document within its record group ('7204A' would indicate the first document in the fourth folder for 1872)
Type
A two-digit field that held a code indicating which of the many types of documents this document was (there are more than 50 types in the edition). A separate table held the codes and their full meaning (‘01’ = Accounts, ‘33’ = Test Reports, ‘79’ = Interviews, etc.)

Date
Three two-character fields: month, day, year. Dating turned out to be very complex. Although it would be nice to use the date function that is built into most databases, it cannot be done in any straightforward way because documents are frequently partially dated, and the date function will not allow dates such as ‘May 1875.’

Author
Two three-character fields containing codes. As in the Type field, there was a separate table holding the codes and the names they represent. There were two author fields (and two Recipient fields) to allow for situations such as an individual writing on behalf of a company.

Recipient
Two three-character fields containing codes.

Name Mention
Two three-character fields containing codes. Limiting name mentions to two fields meant that many names appearing in documents could not be included, which was unfortunate. However, computational power and storage were much dearer when the project began.

Subjects
Three three-character subject codes

Status Codes
Six single-character fields that flagged information about the documents and their data: What language is it in? Is it a fragment? Is it a photocopy? Is it an attachment or enclosure? Is some part of the date conjectured?

Reel, Frame, Addframe
These three fields recorded the reel and frame numbers of the document on the microfilm.

As a database structure, this is far from what would now be considered optimal. However, it did contain almost all the information needed for searching and retrieval. Over time, the database migrated to three new programs: first to a different mainframe program, then to a desktop PC, and finally to a newer desktop program. Once on the desktop the number of name mentions and subjects was increased to 16 each, the Group and Location fields were merged into a single Document ID, and a field was added that held a code for the folder or volume containing the document. With the most recent migration (to Microsoft Access) the data was ‘normalized’ which means it was broken up into a larger number of interlinked tables, most of which only contain a few fields. This makes the data easier to manipulate; it also means there is no limit to the number of names or subjects recorded for a document.
This is not meant to be simply a long and inordinately detailed technical discussion. Any digital documentary edition that provides images of the original documents, with or without transcriptions, can be no better than its database. At the very least the database must have dates and authors' and recipients' names. It should also have information about the organization of the edition. One of the great strengths of a microfilm edition is that once a user finds the first page of a letter, an account book, or a legal proceeding, the successive pages will usually appear on successive frames. Moreover, documents in a given editorial grouping usually appear together on the film. Digital images, however, must be ordered for the user, and beginnings and ends of documents must be flagged somehow. We will return to the issue of the database after discussing images.

Images

In the mid-1990s, after the first graphical browsers had awakened everybody to the Web's potential, a foundation program officer was arguing against the work of entering documents' information into the Edison Papers' database. 'You don't have to do all that indexing,' he said. 'Just scan the documents and put them on the Internet!' Not only was he wrong about the indexing; he was wrong about the 'just scan.' At the time of that conversation, the Edison Papers had published 162 reels of microfilm, each reel averaging slightly more than 1,000 images. The market for microfilm scanning was largely driven by institutions such as banks and insurance companies, which had huge collections on microfilm and which wanted greater access to those images. They were not interested in subtlety, but were content with black-and-white (1-bit) images. Such images can suffice if a typed or printed document is scanned at a sufficiently high spatial resolution (300-600 dpi [dots/pixels per inch]), but most of the documents in the Edison Papers microfilm edition were handwritten, much of it in pencil or light pen, and much of it on paper that had darkened in the century since it was written. There was no doubt that the documents would have to be scanned as eight-bit images (256 shades of gray), a capability the scanner manufacturers were just beginning to explore. Increasing the bit depth of the images allowed us to scan them at a relatively modest resolution of 200 dpi.

After some misadventure, we settled on a vendor. We soon found that the scanning produced better images when we used negative film, as the amount of light that came through positive film tended to overload the sensors and wash out fainter lines. Besides straightforward quality control issues of light or dark images, we found that the scanner occasionally trapped dust particles between the sensors and the film, creating a streak across dozens or even hundreds of images. Those problems aside, we were pleasantly surprised by the quality of the images. Because the documents were recorded on high-contrast microfilm, we expected little in the way of fine distinction, but in fact the images often revealed details that were nearly or actually impossible to see using a microfilm reader.
The original time estimate for the job was six months. There were some kinks to be worked out in the technology, and it took about two years. When it was done, we had nearly 1,500 CDs holding a terabyte of data. The images, captured as uncompressed TIFF files, averaged around 6 MB each, and before we could deliver them over the Internet we would have to create smaller versions. Even more important, we had to somehow link the images to the appropriate document information, so that when a user called up a particular document the correct images would appear.

The creation of derivative images, like much editorial work, is repetitive but not routine and requires a finicky intelligence. We did not have sufficient storage space to put the full-size images online as a viewing option, so the user was going to get one derivative image and it had to be legible. We aimed at reducing the spatial resolution to 80 dpi, which is about life-size on most computer screens, and using JPEG compression to wind up with an image that was about one percent the size of the original (an average of 60 Kb). Because the microfilm edition groups documents by subject or type, the images could often be batch-processed and then reviewed for quality. We did not hesitate to lighten or darken an image if the alteration made it easier to read. (Some editors initially winced at this, reflecting a general uneasiness about manipulating digital images, but it is philosophically no different from changing the lighting when microfilming in order to enhance the contrast of a document; an illegible document is of little use.) Although most of the film was shot at a constant 14:1 reduction, some unusually large or small documents were filmed at other ratios. We tried to scan all the images at the equivalent of 200 dpi on the original. However, in the case of agate-type newspaper clippings and other documents with fine detail, we increased the spatial resolution of the online images to make them legible for the user.

Linking images to their document information was a painstaking process. There is no simple one-to-one correspondence between images and their information. A document might contain one or more attachments or enclosures, for example, in which case a user who retrieves the covering document will want to see the enclosures as part of the document, while at the same time the enclosures might be recorded as separate documents themselves. That is, the same image may be linked to more than one document. With the help of an outside programmer (and a 21-inch screen), we created an interface that displayed successive digitized images on one side and database information on the other. Using the microfilm frame numbers in the database, the program would calculate the number of images in a document. Most of the time the calculation was right, but when it was wrong the operator—working with the digital images, the database, and the microfilm—could easily override it.

The result is an online image edition that allows the user to sample or assemble the documents in a number of ways—name, date, document type, editorial organization—and to view as a group documents scattered across many reels of microfilm. At the same time, there are certain characteristics of microfilm that are useful to preserve in an electronic edition. A user landing on a page in the middle of an experimental notebook, account book, or scrapbook is likely to
want to browse forward and backward through the entire item. This problem was solved by the creation of a new data table, but the solution was only possible because the structural information identifying the collection of individual documents as a unit was already present in the database.

**Text**

As might be expected, the other side of the edition—creating live, linked text from the transcribed documents and their editorial apparatus—presented its own set of issues. Again, the electronic text and apparatus, if not identical to those in the published volumes, reflected the same principles of selection, transcription, and annotation. There were both theoretical and practical considerations in the creation of the digital text for the Edison Papers, as is always the case, but the practical issues predominated.

At the start of the project we had chosen ‘a conservative expanded approach [to transcription] that does not try to ‘clean up the text’ ... to strike a balance between the needs of the scholar for details of editorial emendation, the requirements of all users for readability, and the desire of the editors that all readers obtain a feel and flavor for Edison, his associates, and their era’ (Jenkins lv-vi). Because of the nature of the documents, establishment of an authoritative text—in the sense of choosing between alternative readings—was rarely an issue. Only a tiny percentage of the documents (such as contracts, letters to the editor, or patent specifications) were written or printed more than once.

Traditional text editing conventions generally proved a comfortable fit. Where it was not, we tried to keep our improvisations as close as we could to the spirit of traditional models. For example, we used traditional abbreviations to describe the documents: ‘A’ for autograph, meaning the document was in the hand of the author; ’L’ for letter; ’D’ for document; ‘S’ for signed; and so on. However, there was no existing symbol for an artifact, nor was there one for technical notes or notebook entries, both of which we had in abundance. So we created new ones: ’M’ (Model) for physical objects, and ’X’ for technical materials. Almost immediately, we realized that notebook entries and similar materials presented a significant new entanglement. From quite early in his career Edison had co-workers beside him at the bench, people who helped him carry out his research plans. Sometimes one of them would work and another would take notes; sometimes the experimenter would make his own notes; sometimes a group of them would work together with one keeping notes; sometimes the researcher was more or less autonomous; sometimes they were pursuing a line of thought that Edison had assigned. What stumped us was the question of authorship raised by such documents. Was the author the person who did the work? Recorded the work? Had the idea for the work that was carried out? Even if those puzzles had answers, most of the time we couldn’t assign those roles with much surety. Finally we simply cut the Gordian knot and declared that documents of type ’X’ had indeterminate authors, even when Edison wrote them. This turned out to reflect the way work was pursued as well as the way many of Edison’s co-workers felt about the work. They realized that
they were active participants, but they also recognized that when Edison was not in the laboratory work slowed after a couple of days and that in fact the work would not have existed without Edison to drive it.

Although the quantity of Edison's drawings and their importance is very unusual for documentary editions, with only a few published scientific documents as distant prototypes, there is one documentary category in which the Edison Papers pioneered and which so far remains unique to the project—technological artifacts (Rosenberg). Edison was, after all, an inventor, and the things he created were the core of his work. In the mid-1980s, when we confronted the problem posed by physical objects, we considered several options for presentation, even exploring videodisc (and, in a lighter moment, paper pop-up constructions). Finally we decided to give each artifact an annotated introductory headnote, presenting the object as a photograph if we had one—preferably our own, if we had access to the object, or a historical image if we didn’t—or, failing that, a historical sketch, patent drawing, or other representation. Dating was a challenge—we settled on design as comparable to textual composition—as was the slippery analogy of transcription for both photographs and drawings. We have found the system satisfactory for print, and the edition's users have seemed to agree. The electronic edition will offer similar presentation of photographs and drawings. For those instances where an artifact still exists, though—stock tickers, electric lighting, and motion pictures, for example—the internet's potential for displaying sound and motion opens fascinating possibilities for annotation, affording the user detail, depth, and understanding simply not possible with static images and text. As Edison's designs stretch the notion of 'artifact' to include his electrical central stations and the Ogdensburg ore-milling plant of the 1890s, questions of presentation will doubtless adapt with them.

It was planned from the beginning of the electronic edition that the text of the print volumes would be included, marked up with SGML (later XML) to take full advantage of the capabilities of live digital text. But before doing any markup, the text itself had to be established, letter-for-letter as accurate as in print. This was not a problem for volumes three and four, because all the corrections, alterations, and additions to the documents and editorial material in those volumes made after their submission to the press—through galleys and page proofs—had been entered into the electronic files of the documents. For the first two volumes, however, edited and published before the electronic edition was on the horizon, the electronic files were uncorrected. Neither scanning with optical character recognition (OCR) nor simply retyping the published text was accurate enough to avoid another full proofreading of the text. However, those two processes combined offered an extremely accurate text, since the computer and the typist did not make the same kinds of mistakes. The scanned text was divided into individual documents, and a second copy of the documents was typed in individually from the volumes. The project secretary then used the 'compare' function of WordPerfect to find differences between the two versions and create a corrected one, which she stored. Editorial material, such as front matter, headnotes, and back matter, was treated the same way.
What was not clear at the outset was what DTD would be used (even less clear was what software would be used to deliver the text on screen \(^3\)). The work done by David Chesnutt, Michael Sperberg-McQueen, and Susan Hockey for the Model Editions Partnership was immensely helpful, as was the DTD they developed. The MEP tackled a diverse collection of electronic documentary presentations, and the resultant *Markup Guidelines for Documentary Editions* is now the only reasonable starting place for anyone preparing text for an electronic documentary edition. \(^4\)

We found, as everyone who attempts markup does, that tagging the text is a painstaking process. \(^5\) The situation is admirably described in the *Markup Guidelines*:

> Some of the tagging we recommend can be automated in very simple ways, e.g. by simple macros in a word processor, or in some cases even by simple global changes. Other tagging we recommend can be automated successfully only by a skilled programmer. Some things fall between the two extremes, and can be performed by an astute editor or a journeyman programmer. Some kinds of tagging cannot be fully automated, even by expert programmers, but an automatic process can propose tagging for a human editor to accept or reject, in much the same way that a selective global change in a word processor allows the user to decide whether or not to make the change, on a case by case basis. And, finally, some tagging is done most simply by hand.

Automated and semi-automated tagging can substantially reduce the cost of tagging an edition, but failed attempts to automate what cannot be automated can consume alarming amounts of time, patience, and money. The art and challenge of managing the creation of an electronic edition using limited resources lies, in no small part, in automating what can be automated, doing manually what must be done manually, and deciding (perhaps with a sigh) to leave untagged what cannot be tagged automatically and is not essential to the edition. It will not always be easy to decide where to class a particular kind of tagging: some kinds of information require manual tagging in some collections of documents, but can be tagged automatically or semi-automatically in others. Right judgment will depend on the body of materials being edited, on the time and resources at hand, and on the skills of the available programming assistance. (*Markup Guidelines*, Section 2.4)

Those paragraphs constitute guiding principles for editors. We had the advantage of long familiarity with our word processor, which allowed us to write fairly complex macros that greatly simplified much of the tagging. Nevertheless, the actual work required as much intelligence and care as any other in the project.

The structural tagging of the documents tended to be straightforward (which did not always mean easy). Markup targeted the structure and physical
presentation of the text—date lines, paragraphs, closings, signatures, damage—rather than such details as names of people or places. Perhaps the most important structural decision concerned images in the documents—what in the volumes we call ‘art’ (as opposed to ‘illustrations,’ which are images we supply). These images are usually integral to the meaning of the documents, especially in notebook entries, patent documents, and other technical material. Occasionally a selected document has no words on the page at all, and we have to insert bracketed letters next to drawings as hooks from which to hang annotation. Often enough the placement of text relative to drawings is meaningful, and there are times when no transcription can fully capture that relationship. On paper we tackle these issues with careful explication, thoughtful design, and sometimes inclusion of text in reproduced art; still, there is sometimes meaning lost in a patchwork of text and excised drawings. But in the electronic edition, with the original document instantly available, the transcribed representation of such documents no longer carries the full burden of interpreting the document to the user. It becomes an aid to reading, and the design of individual documents—still a nightmare of fluidity on screen—becomes a secondary consideration. With the text transcribed, <figure> elements that contain the ID of their respective page image are inserted where images exist on the page. Those elements will appear on screen as icons that, when clicked, will call up the entire page image.

Embedding editorial material proved challenging. The first significant decision concerned the index. The back-of-the-book index, although created for use in a codex, is a sophisticated intellectual tool that maintains and arguably increases its strength when applied to electronic texts. Full-text searching can help find specific text known to exist, but it is at best a marginally effective way to explore a body of information. A good index not only provides direction to implicit meaning in the text, but it reveals to the user what may be found in the work. In print volumes it is often used to that end as a browsing aid; online, where the scope and depth of a work is harder to judge, such an aid is that much more valuable. The full index—or that for any selected group of documents—can be assembled on the fly for browsing and access. Moreover, index entries can appear on the screen with documents or editorial text, and so serve to link to related material.

We had final electronic versions of the volumes' indexes, and we used a macro to convert the alphabetically ordered entries

```
Ammonium nitrate, 264n.2
Anders, George, 104, 122, 128,
132n.1, 664; agreement with TAE and Welch, 118-19; and Boston instrument,
109; and Financial and Commercial instrument, 133n.1; and magnetograph, 82n.2,
104; and polarized relay, 110n.9
Anderson, Frank, 253
```

into a tagged, page-ordered list
Having the index prepared this way made it relatively straightforward for taggers to place entries in the text as they serially processed the documents. Afterwards a macro inserted an ID attribute in the element for linking.

The other complex editorial decision involved references and is still very much in process. In an edition where only a small percentage of the documents are transcribed, annotations often direct the reader to the images of other documents that are not transcribed. In the books, these take the form of lists of documents on the microfilm, identifications accompanied by reel and frame numbers. For correspondence and many other types of documents, the references can be translated into direct links. However, references to material in notebooks are rarely coordinate with the way notebooks are divided into documents. Moreover, Edison and his crew recorded their work in whatever notebook lay to hand, and so the background information for a technical document that represents a week's work is as often as not collected from more than one source. What in the book are haphazard strings of frame numbers must become a new type of online document, an artifice that allows the user to see the relevant notebook pages as the assemblage we intend. Just as we made notebooks, account books, and scrapbooks browsable by creating a new data table, these compound references will be a creation of the database. Like the tagging, this will be half-automated and half-handwork; with the tagging—and the scanning, the data entry, the myriad editorial decisions, and the work that follows on them—it is part of a foundation for an edition we couldn't imagine when the project started twenty-five years ago, one that combines the known strengths of microfilm and books with the remarkable power and access of the digital world.

Notes

1. The aspiring editor can do no better than to acquire copies of Kline and Stevens and Burg, and to join the Association for Documentary Editing (http://etext.lib.virginia.edu/ade/), which makes cordially available the aggregate expertise of the profession (and discounts both books to members). The two volumes lay out in careful detail the strengths and weaknesses of the many methods used by
the best practitioners. I must mention that, although I use ‘we’ for convenience throughout this chapter, the Edison Papers was organized in 1978-9 and I did not join until 1983. I left the project in 2002.

2. There are services that will recapture printed text by double-keying and comparing, and the accuracy is at least as high as can be obtained any other way. Because the work is done overseas, we were concerned about the labor practices involved; our situation allowed us to recapture the text this way. However, it seems that the principal firms doing such work are mindful of their employees' well-being, and it would certainly be well worth considering their services if an edition consisted of or included a significant quantity of previously published materials (Farrell and Olsen A35).

3. This question remains unresolved, awaiting completion of a substantial portion of the markup before a commitment is made. Choices are very limited; without a programmer, they are nonexistent. The Anastasia suite, created by Peter Robinson and Scholarly Digital Editions, requires programming expertise we do not have but offers a promising mix of power and flexibility.


5. When we started, in 1998, we did preliminary tagging on each document in Corel's WordPerfect and then switched to SoftQuad's Author/Editor (which no longer is supported) for most of the work. Author/Editor was superior at tagging the text, but we were frustrated by its text-handling and proprietary file format. When later versions of WordPerfect proved capable of handling markup with sufficient sophistication and stability, we used it exclusively. The significance of Corel's 2002 acquisition of SoftQuad and its XML editor XMetaL remains to be seen.

6. In April 2000, the University of Vermont and the National Historical Records and Publications Commission funded a meeting to discuss ways to improve and standardize intellectual access to electronically published historical documents; that is, online indexing and searching. The result of those deliberations, known as the Burlington Agenda, can be found at http://chipmunk.uvm.edu:6336/dynaweb/burlingtonagenda/burlingtonagenda.

7. For a related group of editions such as the Founding Fathers, a common index would obviously be a remarkable tool. However, merging existing indexes, which are often created with different intellectual perspectives, would require significant groundwork.

8. This is another idiosyncratic treatment of the documentary record. In order to make the microfilmed notebooks most useful for readers, the database treated the appearance of a different date in a notebook as the beginning of a new document. This allowed us to sort the notebook materials by date in the printed index. The electronic edition uses the same document definitions.
Choosing Content Management Technologies

For those working in libraries, archives, museums, galleries and other cultural institutions there is an increasing amount of software and hardware available to manage, catalogue, edit, store, retrieve, display and publish their content. How can organisations choose a management system that’s right for them?
Open Source Software (OSS) is often free, or subject to inexpensive licence fees. Help or technical support can often be found through an online community of volunteers (although an increasing number of OSS projects offer paid support). Because the source code is openly available, it is easier to adapt the software to suit your organisation’s needs, as long as appropriate technical support is available to your organisation. The user can potentially fix problems or bugs immediately without having to wait for the next release of commercial software.

Proprietary software requires the payment of licensing fees which are generally higher than those associated with Open Source solutions. Proprietary software generally comes with a single, dedicated source of technical support, making it easier to contact someone if anything goes wrong. Proprietary software may be appropriate for organisations where staff lack technical skills and experience, especially in regard to implementation and maintenance.
Draw up a list of requirements. Every organisation has different needs and it is important to understand these fully before committing to any one system over another. When your requirements gathering and analysis is complete, you should have the following:

### Existing Legacy Data and Systems
- An assessment of any legacy systems you may need to integrate with – for example an existing CMS or legacy cataloguing done in Excel spreadsheets or an Access database.
- A decision on whether your new system will need to integrate with any existing software in your organisation or whether the new system can replace it entirely. The fewer systems deployed on site, the more effective a support team can be.

### Organisational Support
- A budget for the project.
- An awareness of the level of IT support in your organisation, and whether staff time will be available to help set up and administer the new system.
- An awareness of the IT and cataloguing skills of the potential users of the new CMS, and whether they will require training.
- An understanding of the duration for this software choice (is it a short-term or long-term commitment by your organisation?). In the case of a short-term commitment, an exit plan will also be necessary.

### Functionality of the new system
- A list of the types of collection you will be managing with the new system, and the potential file formats you will need it to support (images, documents, audio-visual formats, etc.).
- A decision on whether you need your software to be internal, public facing, or both, and whether it requires a web interface.
- An overview of which metadata standards your new system should support.
- An awareness of the potential number of users of the new system (this may affect the number of licences required for proprietary software).
- A decision on whether you will need to export your data and metadata in standard formats, for example to allow interoperability with DRI or Europeana.
- An assessment of the security and back-up solutions required for your data.
- A decision on preferred hosting options for the data, whether local, third party or cloud, and how control and IP of the data are influenced by your choice.
Undertake thorough research on the systems that are available. A great deal of information is freely available through software websites and wikis, community forums, Twitter, online journals, blogs and through other people working in your domain. When researching, some questions to consider are:

- What kind of support exists for the system – websites, forums, mailing lists? This is especially relevant for open source.
- How long has the system been available and is it well established in your domain?
- What do other organisations in your domain (libraries, archives, museums, etc.) use for similar content?
- Is the software still under development? Is it still supported by either the vendor or the community?
- Is a free trial version of the software available for you to assess?
- Will the system support different user roles – for example, different access permissions for different members of staff?
- How will the new system integrate with existing systems in your organisation?
- Will the system support your chosen metadata standards? Can the fields and data model be customised?
- Can your existing legacy data or metadata be ingested into the new system?
- How will the new system integrate with existing systems in your organisation?
- Can the user interface be customised easily to suit your organisation’s needs?
- Does your organisation require multiple licences/log-ins to use the new software?
- For proprietary systems, if support is offered, where is the vendor based and will live support be available in your time zone?
- Will the system support different user roles – for example, different access permissions for different members of staff?

And finally, consider feeding your own experience and findings back into the community to help inform other organisations’ research.
When Not to Use Drupal

Submitted by Quinn on Tue, 07/03/2012 - 20:41

WordPress is better sometimes

Of all the CMSes available to humanists, WordPress is both the most convenient to use and the most clearly aligned with traditional scholarship. After a relatively painless installation, WordPress provides the capacity to display long narrative text with embedded images and videos. Without any modification, a WordPress site can be built to contain both a running blog and a set of static sub-pages, open for comment or not, and available to the outside world.

Many universities offer WordPress installations through their Academic Computing or Information Technology departments, making this process even easier. With a wide variety of available themes, the look and feel of a WordPress installation can be dramatically changed to emphasize text or image content.

The scholarly community has shown much love for WordPress. For many academics, a blog is the first foray into digital content creation and WordPress is often the first blog that’s utilized as an application rather than subscribed to as a service from sites like Blogger and WordPress.com. The blog has been dealt with as a form, theoretically and productively, for quite some time. Even back in 2006, projects like the Institute for the Future of the Book was trying to develop methods to make the blog form more nuanced through the creation of WordPress extensions like CommentPress.

WordPress is particularly well-suited as a mechanism to support individual scholarship and commentary using traditional linear narratives. If an individual or project is only interacting with the outside world via reporting their findings in the format of traditional articles or short posts, then WordPress is a perfectly acceptable solution.

That said, once you try to move beyond the blog form to integrate the kind of collaborative workspaces and community-oriented functionality typically needed for a moderately collaborative digital project, WordPress is lacking. And if you want to deal with data within that site, and taxonomies and data models and visual representation of such data, then you’re better off in Drupal, which can provide you with all the functionality necessary to present long narratives with embedded images but also provide you the capability to integrate datasets and sophisticated collaborative tools.
Omeka is better sometimes

Omeka ([http://omeka.org/](http://omeka.org/)) is, like Drupal, a CMS built on the MySQL-PHP technology stack. Omeka is designed for exhibiting narratives and collections to audiences, and with the recent release of Neatline ([http://neatline.org](http://neatline.org)), which extends Omeka to include sophisticated geospatial and temporal functionality, these narratives can include rich geographic and chronological data. Omeka, especially enabled with Neatline, is also well-supported by two key institutions in digital humanities research: CHNM ([http://chnm.gmu.edu/](http://chnm.gmu.edu/)) and the Scholars' Lab ([http://www2.lib.virginia.edu/scholarslab/](http://www2.lib.virginia.edu/scholarslab/)) at UVa. Another benefit of Omeka is that it's focused on metadata standards for public humanities work like that found in museums and libraries, and handles much of the taxonomic effort that you may find unwieldy in Drupal. As such, if your project is focused on the presentation of a narrative with visual, textual and spatial data, then you should seriously consider using Omeka.

For any digital humanities project where the aim is to present time and space-enabled data, you should try out Omeka with the Neatline extensions. If you're just getting started in creating a collection of objects, places, and so on, you should also first take a look at Omeka. The value of Drupal over Omeka comes when your use case becomes too specific or complex and you need to create specific data models, views into the data, and services to connect to your work.

TEI is much better for certain things

When a humanist ventures into the world of digital textual scholarship and data archiving, they quickly encounter TEI ([http://tei-c.org](http://tei-c.org)). Many scholars are first exposed to the idea of separating style from content—indicating what something means (e.g. book title, ship name, emphatically spoken word), rather than how it should be displayed (italics)—through a workshop on TEI. The TEI Consortium offers a set of extremely detailed and extensible guidelines for how to add metadata to text, with a major use case being the development of digital critical editions. Projects involving the preparation of textual materials are all but expected to use TEI, particularly if they are applying for grants, and using TEI is widely seen as a way to increase the likelihood that digital textual scholarship will be accessible in the long term.

Scholars who are only used to working with word processing software face a learning curve with TEI. There is something initially counterintuitive about taking a human-readable text and rendering it nearly unreadable with angled brackets full of additional information. The fact that rendering a TEI encoded document readable again involves writing XSLT exacerbates this issue. There have been various efforts aimed at addressing these challenges, with the TEI-ann and TEILiteEditor plugins providing an alternative to manually typing angled brackets, and TEICHI ([http://www.teichi.org](http://www.teichi.org)) enabling basic search and display of TEI texts in Drupal, to name a few.

The impulse to use a durable format for the sake of preservation can push a scholar towards TEI, even when it doesn't align well with his overall goals for the textual material. For a project on Thomas Jefferson, constructing a TEI bibliography of his library might be the best choice, particularly if the project also involves marking up his personal papers in TEI. But a continually evolving bibliography of secondary literature about Vladimir Nabokov—even if its compilation is a research project that spans a scholar's entire career—is less of a good fit for TEI, and even less so if the scholar wants to solicit additional entries from Nabokov scholars worldwide. A TEI-based
approach to such a project might involve the following steps:

1. Taking the Word document where the data has been collected (probably without any markup whatsoever) and copying the text into an XML document
2. One-by-one, marking up the data using the `<biblStruct>` element, and its numerous sub-elements (`<analytic>`, `<monogr>`, `<imprint>`, `<pubPlace>`, and `<biblScope>`, to name a few typical examples)
3. Writing XSLT to output various kinds of listings (all entries sorted by author, all entries sorted by date, only monographs, etc.)
4. Realizing that some of the shortcuts and errors in your TEI “break” your listings (e.g. listing all authors of a book together makes it difficult to group together works by author)
5. Cleaning up the TEI XML to address the issues in #4
6. Generating static HTML from your XML and XSLT, and posting them on your personal university web space
7. Emailing a list of Nabokov scholars and soliciting additional bibliography entries, which you manually add to your TEI document when they email them to you
8. Realizing that it’s a pain to have to re-generate all your HTML files every time you add a new entry to your TEI document, and resigning yourself to the drudgery, or...
9. Trying to get your local IT support to configure and support a server with the rather specialized Apache Cocoon web framework, which can run the XSLT “on the fly”, so you only need to upload the updated XML and the server does the rest of the work

This workflow requires knowledge of how to use an XML editor, how to mark up texts using TEI, and how to write XSLT, as well as an IT department with the resources to support niche projects and the knowledge to set up and maintain a Java-based framework they’re probably not already using.

Achieving the same goals with Drupal might involve:

1. Having your local IT support install Drupal; its requirements are much simpler than Apache Cocoon, and your institution might already be running it
2. Install the **Bibliography module** ([http://drupal.org/project/biblio](http://drupal.org/project/biblio))
3. Enter the bibliography entries from the Word document into web forms that are easy enough to understand that you might feel comfortable delegating some or all of the data entry to student assistants
4. Go to yoursite.yourinstitution.edu/biblio at any point during or after the data entry process to browse your bibliography and sort and filter by a variety of metadata fields
5. Install the Views module and use the interface to generate any listings not provided out-of-the-box by the Bibliography module
6. Email a list of Nabokov scholars and invite them to create an account on your site, and provide them with simple instructions for how to add new listings
7. Sit back and watch your bibliography grow

All of these steps can be done by a humanist without any specialized training or a background in technology, following the instructions laid out in this book.

A Drupal-based bibliography facilitates collaboration and display of the bibliography, but what about preservation?
It’s unrealistic to expect your Drupal site will run on a server forever, or that your library will archive your data with the full Drupal environment in tow. At some point(s) in the project’s lifecycle, the data will need to migrate to a more archivable form, and there are Drupal modules that can help you do that; see [section about import/export].

When deciding between TEI and Drupal, consider the following factors:

- Is there a module that aligns well with the kind of markup I’d be doing in TEI (e.g. the Bibliography module)?
- If not, is my markup simple enough that I can achieve the same kind of granularity in my data storage by creating a custom content type (e.g. if you’d only be marking up your text for poem, stanza, and line)?
- What are potential funders’ expectations for the use of TEI in projects similar to mine?
- How important is it that my data interoperate with other TEI texts as part of a larger initiative? This may not preclude the use of Drupal for data entry, if you have a strategy worked out for exporting and converting the data into TEI.
- How much collaboration is required for this project?

If your project uses detailed markup that covers more than just structural elements of the text, chances are that TEI is the right choice. For simpler text encoding projects, it’s important to weigh the pros and cons of TEI and Drupal as they relate to your goals for your data, amount of technical support available, and workflow. Drupal might be a better choice, but it may also require that you explain to colleagues why you did not use TEI; be thorough in your exploration of both options and clear about the rationale behind your decision if you choose the Drupal path.

Hand-crafted .js for maps/visualizations is better sometimes

Most humanities research using digital tools will produce some form of representation of relationships, demographics or other phenomena related to research. This often takes the form of sophisticated plots and diagrams as is produced in software like Tableau, or cartographic format as results from spatial analysis using ArcGIS, or network representation like that produced by network analysis toolkits like Gephi. Each of these software packages provides the ability to not only represent the rich content of a dataset but also to interrogate the individual components and interact with the data on its own. When a scholar or project needs to share the results of these explorations between members or with the community at large, these tools provide the capacity to output static images with legends for just such a purpose. Posting these images in a web-accessible manner is trivial, and can be accomplished in any CMS.

However, there is often a need to provide researchers or community members with the capacity to explore the dataset with some of the same interactive functionality provided by the desktop tools but in web form. This is typically accomplished by bringing to bear one of the rich Internet scripting platforms, like Flash, Silverlight or JavaScript in the browser. These platforms have various libraries to provide data visualization and interactivity that can work off the same data the scholars are examining with their desktop applications.

Typically, you would not think of using a CMS to deal with rich data visualization and interactivity, but that’s beginning to change, especially with geographic data in Drupal, as will be explained in the Spatial Drupal chapter.
Integration of data as content types into Drupal comes with significant overhead and typically provides less functionality and slower performance than hand-crafted scripts. If all one wants to do is show a flat dataset in a well-known visualization pattern, then the costs of building it in Drupal far outweigh any benefit.

But, if a project is utilizing some form of visual representation directly in their research, such as geographic representation of the research matter, then Drupal allows that scholar or project to represent a timeline or map-based view of the ongoing, collaborative research.

Other CMS (Django/Flask/ZoobieX, etc.)

To be written...

Tags:  WordPress,Django,TEI,Omeka,Javascript,maps

http://drupal.forhumanists.org  (http://creativecommons.org/licenses/by-sa/3.0/)
Choosing a platform for your project website

Learning HTML is no longer a requirement for building a website for your project. There are many platforms-- general-purpose platforms and ones tailored to specific kinds of projects--that allow you to build much more sophisticated project sites than would be possible if you were building from scratch. When choosing a platform for your project website, the major factors to consider include functionality, familiarity, community, support, and cost.

FUNCTIONALITY

What do you want your project site to do? Are you developing an exhibit or collection of material, which needs to be displayed in a sequential order? Are you developing a directory, that you want to be browsable and searchable based on metadata you’ve entered (like "author", "publication date", "media used", etc.)? Do you want to use your site to transcribe content, or add annotations? Will users be able to create their own accounts, and will having an account provide them with additional access or unlock new tools on the site? What format(s) does your content take (text, audio, video, still images, downloadable files, etc)? Will your content be stored on the site itself, or is it coming from another hosting provider, like YouTube, a library website, or an institutional repository? How do you want to display your content-- in an image gallery, a timeline, a map, a list, or some other way?

These are just a handful of the considerations that should influence your decision about what platform to choose for your project. Functionality is the most important factor to address. While the other factors-- familiarity, community, support, and cost-- can help you choose...
between multiple options that provide more or less the same level of functionality, choosing a platform that does not do what you need, because it's free or support is available for it is only a good idea as a stop-gap measure (e.g. to establish a URL and web presence in time for a conference or grant proposal) while you explore better options.

Ideally, you'd find a platform that does everything you want your project site to do, with minimal extra configuration, allowing you to focus on preparing and entering your data. This rarely happens, in part because what you want your site to do is often tied to the unique traits of your data itself, which a pre-made system isn't designed to accommodate without a little (or a lot) of work.

Some platforms, like Drupal, are extremely generic, and almost certainly won't do what you want out-of-the-box. To build a scholarly project site with Drupal, you have to add numerous "modules"-- or pieces of packaged-up functionality that someone has written code for. Drupal has a large international community of developers who write modules, and most modules can be installed and configured without you or your assistants ever having to look at the underlying code. Choosing a generic platform requires more time investment upfront, but leaves you with more flexibility later. For instance, if your project starts off with only text, but you later decided to incorporate video, it may be considerably easier to make that change if you've chosen a generic platform like Drupal (which some people use for textual content, others for video content, others for spreadsheet-like data) than a platform designed specifically for managing texts.

Other platforms make it very easy to build certain types of sites. Omeka is an example of a platform for sharing collections and exhibits. There are "add-ons" available for Omeka that extend its functionality, much like Drupal "modules", but they are intended to improve its collections and exhibits, not to fundamentally transform it into a different kind of platform, as add-ons for more generic platforms sometimes can (e.g. the BuddyPress "plugin", which turns WordPress from a blogging and generic content management platform into a social networking platform). Similarly, MediaWiki is a platform for building a wiki, and its "extensions" generally provide additional wiki functionality.

If you're not sure how your project may evolve, and all things are equal with regard to the other factors (like community, cost and support), you may be better off choosing a generic
platform, to keep your options open. If you have a clear sense of the limits of the project's scope, it may be better to choose a more specialized platform, if an appropriate platform exists.

FAMILIARITY
 Platforms that started out having very different user interfaces are increasingly converging around certain design approaches and choices. All commonly-used platforms have (or can have, with the help of a module or plugin) a text authoring interface with WYSIWYG capabilities ("what you see is what you get" -- e.g. buttons you can click on to do things like make the font bold, or add a link, rather than making the user write HTML). Designs, or themes, that you can download and use for sites running any platform are increasingly adopting adaptive or responsive design techniques, which render the site differently depending on whether it's being viewed on a high-resolution laptop, a tablet, or a phone. This convergence makes it easier to make choices about the platform for your site without overly concerning yourself with what platforms are already being used by other popular sites in your field: chances are, you can make your site behave like other sites, even if it's running on a different platform. That said, the more you expect users to interact with your site-- be it through adding new content, or providing transcriptions or annotations, or engaging with other users-- the more important it is to minimize the learning curve required of your users. One of the easiest ways to do that is using platforms and plugins that they're already familiar with. For example, if you're building a scholarly network, and you know your users are MLA members, you may want to use the Commons In A Box package, which powers MLA Commons.

COMMUNITY
"Community" here refers to the group of people who are using the platform. Do other scholars in your field, or in related fields, use the platform that you are considering? A platform that's widely used by scholars may be a better choice than a platform whose major user base is small business owners, but if all the example sites you can find come from the sciences, where their data is considerably different than yours, you may want to make sure the platform meets your needs. Choosing a platform that's already being used by a community of humanities scholars may make it easier for you to ask questions, and get tips and advice on
how to deal with problems that arise, without having to translate your questions into language more easily understandable for technologists without a humanities background or scholars in another discipline.

SUPPORT
Who can help you develop a site using this platform, and what skills are required to do so? Most universities provide faculty with free access to some sort of web publishing platform, and offer training workshops and/or one-on-one consultation. Many universities also have a web development group with professional staff who may be available to consult or directly help you build your project site, at lower-than-market rates, but they may place restrictions on what platforms you can choose.

If support from a central or departmental IT group isn't an option (either because it's unavailable, or because the platforms they support are truly a bad fit for your project), there are ways of finding support on your own. There may be formal or informal meet-up groups around certain platforms, which provide an opportunity for people who are using the platform to exchange tips and suggestions, and you may be able to find someone in one of those groups who could do some freelance work. (The Berkeley Drupal Users' Group is one example.) You may be able to find a graduate or undergraduate student who can provide you with hands-on assistance, but it's important to know what you're looking for. Do you need someone to help you configure the platform by selecting, installing, and configuring a set of "modules" that are already available? This work requires a significantly less technical skill set than if you need someone to write new code to provide functionality that doesn't currently exist. If you need the person to write code, be sure you know what language(s) they will have to use-- it can vary depending on the platform, and the nature of what you need done. Depending on how elaborate your site design is, you might need to look for someone who has experience developing themes for your platform, which can be a very different skill set than writing modules. Your hosting choices may also be relevant here: are you hosting the site with a service that takes care of setting up the database and installing the platform, or do you need support from someone who can do that kind of work? Step-by-step installation guides can be found for all major platforms, but someone who's mostly comfortable configuring modules for the platform may not be comfortable working on the server level.
With the exception of writing code for modules, which does require specialized knowledge, building a site using most platforms is not beyond the capabilities of a curious humanities graduate student, given some time and opportunity to experiment and take advantage of the numerous how-to guides, books and video tutorials available online. However, some platforms may be more appealing to learn than others, especially for students considering alternative academic careers.

COST

The cost of a project website takes many forms-- hosting, configuration, ongoing maintenance, and the cost of developing new modules, if needed. Sometimes these costs are bundled together, for instance, if you're using proprietary software that's developed, hosted and maintained by a company. In most cases, though, you'll have to estimate these costs, which can vary wildly: inexpensive commercial hosting can cost around $100/year whereas deluxe packages where you have dedicated server resources can cost $100/month; undergraduates available through a research apprenticeship program may work for course credit, while professional web developers can charge $100+/hour. If you're using a freely-available open source platform, and you don't need to have new modules developed, your major costs will be site configuration and data entry. Finding the time to learn the platform well enough to do the configuration work yourself, and/or having a research assistant do that work, can cut costs considerably.

Because of the interplay of these various factors, it's hard to provide a general recommendation about what platform to use. Here are some commonly used platforms (all free and open source):

- **Drupal** - a general-purpose platform with a large international community of developers, including in higher ed and digital humanities. The [Berkeley Drupal Users Group](mailto:quinnd@berkeley.edu) meets monthly on campus. A new working group of people who use Drupal for research will start meeting in spring semester 2014; contact Quinn (quinnd@berkeley.edu) for more information.
- **MediaWiki** - wiki platform used by Wikipedia and the [Brueghel Family Research Website](http://www.brueghel.org).
- **Scalar** is often used for multi-modal projects, and excels in multimedia annotation.
The Scalar site includes [a few examples of projects that use it.]

- [WordPress](http://wordpress.org) - great for simple web publishing and blogging out of the box; lots of plugins are available to turn WordPress into a platform for text annotation ([CommentPress](http://commentpress.org)), social networking ([BuddyPress](http://buddyress.org)), etc.

If you'd like to discuss what might be a good fit for your project, please email [digitalhumanities@berkeley.edu](mailto:digitalhumanities@berkeley.edu) or use our contact form.
Information Visualization for Humanities Scholars

Stéfan Sinclair, Stan Ruecker, and Milena Radzikowska

1 Information visualization for humanities scholars needs to accommodate a mix of evidence and argumentation. The humanities approach consists not of converging toward a single interpretation that cannot be challenged but rather of examining the objects of study from as many reasonable and original perspectives as possible to develop convincing interpretations (for a fuller argumentation of this approach in a digital context, see Drucker). In this sense, we can evaluate a visualization system by determining how well it supports this interpretive activity: a visualization that produces a single output for a given body of material is of limited usefulness; a visualization that provides many ways to interact with the data, viewed from different perspectives, is better; a visualization that contributes to new and emergent ways of understanding the material is best.

In this context, there is an important difference between static and interactive visualizations. A static visualization aims to produce a single perspective on available information. Conventional pie charts, bar charts, and graphs are good examples—these eighteenth-century inventions of William Playfair provide the reader with useful ways to understand information (Tufte), but they are fundamentally tools for display (nevertheless subject to a variety of interpretations of the visual information). Interactive visualizations, on the other hand, aim to explore available information, often as part of a process that is both sequential and iterative. That is, some steps come before others, but the researcher may revisit previous steps at a later stage and make different choices, informed by the outcomes produced in the interim. In a pie chart, by contrast, a static, synchronic object, the visual subdivision of the whole into parts can be useful, but the format does not readily lend itself to experimentation.

The Web site Many Eyes, designed “for shared visualization and discovery,” includes a wide range of information-visualization interfaces that work with user-provided data (see Danis, Viegas, Wattenberg, and Kriss). On the single-output side are graphics like the word cloud and the more compact and visually striking Wordle, originally designed and developed by Jonathan Feinberg, a cluster of words in which the size of each word corresponds to its relative frequency in a document (fig. 1). Although Wordles are sometimes more decorative than functional—even according to the Wordle site (see also Harris’s criticism of word clouds)—some people use them to produce eye-catching summary images for documents or collections of documents (such as conference proceedings), and others use them to consider possible keywords for documents. Although each Wordle is a static visualization, the tool allows users to modify the layout and styling parameters in ways that can draw attention to different features or produce different aesthetic effects. For fiction, they often show the prevalence of proper nouns, which typically need to be removed along with function words (prepositions, articles, and conjunctions) to increase the value of the image (depending on one’s purposes). Wordles (and Wordle-inspired visualizations) are noteworthy for how widely they are used, especially in popular contexts such as advertising (see Smith on how word clouds can quickly convey the differences in vocabulary in toy advertising for girls and for boys).
The Wordle tool provides a very convenient and fast way of generating a view of a document; it can answer a question like, “What words appear most frequently in this text?” Although it is possible to compare two or more Wordles to get a sense of how similar or different documents are in terms of their most commonly ocuring words, the interface is not conducive to document comparison. The Bubblelines interface in Voyant Tools, developed by Stéfan Sinclair and Geoffrey Rockwell, is an example of a visualization tool that is better suited to document comparison. Each document is represented as a separate line, and the interface allows the user to experiment with different search terms to compare across each document (fig. 2).

The New York Times has published a collection of specialized interactive visualization interfaces that are designed to allow users to compare terms in several documents (they are specialized because they do not allow users to provide texts). Like the Wordles, these are noteworthy exemplars of widely seen and used visualization interfaces. “Inaugural Words: 1789 to the Present” provides word clouds for inaugural speeches, and “The State of the Union Address” compares the words of George W. Bush across seven years of speeches. One of the more interactive visualizations in Many Eyes is the word tree, which allows people to study text sequences (phrases) by navigating a tree with phrase-frequency information indicated by type size (fig. 3).
In proliferating the perspectives on data, visualizations can be useful to humanities scholars by providing additional insight into small amounts of text or data, thus supporting what John Unsworth calls “scholarly primitives,” especially for showing patterns that result from filtering, sorting, grouping, and otherwise visually rearranging the material. Visualizations can provide similar insights across large amounts of information that would otherwise be too abundant to grasp or process (e.g., the Google Books Ngram Viewer, based on over 5 million books as of December 2010; see also Michel et al.). Visualizations can also produce comparisons between a pair of documents or of one target document against many other documents.

In the sciences, information is often numeric, and visualizations typically involve converting observed numeric data to visual form and then manipulating the form. In the humanities, information is often textual or graphic from the outset, and the visualizations can support directly manipulating these visual forms in useful ways. Numeric visualization has a useful role to play, but some of the most striking and innovative visualizations in the humanities deal directly with texts and images. These visualizations provide an additional lens that can help us tease out the kinds of information that are not otherwise readily apparent or manageable.

In the context of the humanities, the availability of data to be visualized has been steadily increasing with the proliferation of digital collections (see Greenstein). In addition, relatively sophisticated tools are being developed for working with the various digital materials (see, e.g., the DiRT list). Some of the tools perform tasks behind the scenes, such as cleaning, searching, or sorting. Others consist of visualization components in the interface, which make the results from the analytic operations accessible in meaningful ways.

Taken individually, many of the available tools for digital humanities scholarship seem relatively simple (e.g., TAPoRware), intended to perform a single function well (as per the Unix model of piping data through a variety of tools), and the visualization component may either supplement that function or make it possible (the TAPoR recipe “Compile Textual Data and Visualize the Results Using Excel” describes such a process).

Many of the simpler tools feature an open-ended search mechanism (similar to a search engine like Google). When browsing a collection of items like a library database, for instance, the most common strategy for finding out what is in a collection is to do a series of keyword searches and look at the search results. This approach is most effective
when the goal is a single document for which the searcher knows the title, author, or some other identifying piece of information (see Seal, Girdlestone, and Warwick).

In the humanities, however, an equally important task is to locate (or discover) new material, with no prior knowledge of the kinds of details used for retrieval. Using a search-based system, the task becomes a guessing game, where the searcher begins with some likely candidates for search terms and then reviews the results and tries again until some relatively sufficient set of items is assembled for further study. Adding visualizations to the browsing task can supplement this function in several ways, giving the searcher the option to assemble groups of documents visually instead of working sequentially through a list. These kinds of interfaces are often referred to as exploratory interfaces (see Shen et al.).

**Browsing by Grouping**

The literature on human-computer interaction provides a wide range of experimental systems that enable grouping by browsing. The scatter-gather browser developed by Peter Pirolli and his colleagues is a good example; it uses dots to represent documents for users to collect and disperse (Pirolli, Schank, Hearst, and Diehl). Ben B. Bederson’s *PhotoMesa* similarly allows the user to interact with a page display, as seen in the thumbnails of digital images in figure 4, visually grouping them by keywords. Kerry Rodden, Wojciech Basalaj, David Sinclair, and Kenneth Wood address the question of how similarity among the grouped images supports browsing tasks. In 2007, Jonathan Harris and Sepandar Kamvar produced a suite of dynamic displays involving posts about emotions scraped from blogs and visualized as colored dots arranged, in one of the displays, by attraction to the current cursor location.

We have developed two experimental interfaces that deal with browsing collections or documents through visual grouping. The first is the *Mandala* browser (fig. 5), which allows researchers to open a document or multiple
documents at the same time and iteratively construct visual Boolean queries that draw on the underlying data (such as the XML-encoded version of *Romeo and Juliet* in fig. 5). The user may perform conventional searches, but there is also a variety of automatic and semiautomatic mechanisms for exploring the data.

Figure 5. The Mandala browser, originally developed by Oksana Cheypesh, Constanza Pacher, Sandra Gabriele, Stéfan Sinclair, and Stan Ruecker, is a universal space for researchers investigating structured text collections or individual text documents encoded with XML (although additional formats are supported, such as PDF, plain text, and comma-separated values). This screenshot shows the browser loaded with *Romeo and Juliet*, where each dot represents a speech in the play. Multiple search terms can be defined that attract the items (speeches) toward the label (dots on the periphery represent speeches that do not match any queries). In this example, the user has defined several magnets to identify different speakers (Romeo, Juliet, Nurse, Mercutio), as well as several magnets to match terms that indicate family relations in the text (mother, father, sister, brother, cousin). Dots that are displayed close to the magnet represent speeches that only match that query (e.g., speeches by Romeo or speeches that only contain *mother*), whereas dots between magnets match more than one query (e.g., speeches by Romeo that contain *mother*).

For example, someone might be interested in studying a selection of plays by Shakespeare and looking at invocations of the natural elements to see how they are used to help define supernatural figures. The hypothesis might be that Lear on the heath is transformed not just into a madman but also into a supernatural creature. The researcher could iteratively construct a query space that first asks for all the speeches by selected characters, such as Puck, Ariel, the three weird sisters, and Lear. Then key vocabulary could be added to the query, with terms such as *wind*, *water*, and *air*. The point would be to collect the set of speeches from a specific set of characters that could serve as a baseline for the comparison. It might turn out that there is very little actual invocation of the elements, even by the acknowledged supernatural characters, which would invalidate one of the felicity conditions of this line of thought, and the researcher could stop there. Alternatively, the subsets that appear might prove fruitful, and the study could extend to other figures with a high degree of agency, such as Prospero in *The Tempest* or the Duke in *Measure for Measure*.

The significance of the Mandala browser is that the queries are built up in single steps, with aggregated visual feedback at each step, and the sets and subsets of results are usually easy to see and understand. It is possible to carry out this kind of inquiry using digital copies of all the plays, a word processor, the search function, and a fairly significant amount of cutting and pasting (Brown et al.). In Mandala, however, the query space consists of one small dot for each speech and one colored magnet for each part of the query, making the entire investigation possible in a
single screen that changes dynamically as the user progresses through the different stages. The *Mandala* browser provides the crucial mechanism for the user to read text represented by the items. Our various user studies with humanists have repeatedly indicated that researchers want to be able to return to the text. *Mandala* is designed to be open-ended (fed by user-provided data), but we have experimented in particular with Shakespeare’s plays, the *Orlando* project archive, the early modern English witchcraft trial documents from Chadwyck-Healey, the course descriptions from the 2008 calendar at the University of Alberta, and a variety of interview transcripts.

The second of our experimental interfaces for collection browsing through visual grouping is a family of transferred interfaces that we call the showcase browsers (fig. 6; see Chow and Ruecker). Like *Mandala*, each showcase browser is based on rich-prospect browsing principles; that is, it shows on the default screen one meaningful representation of every item in the collection, combined with tools for visually grouping the items.

![Figure 6. A screenshot of a showcase browser of biodiversity projects in Edmonton. The projects can be grouped and subgrouped by any combination of the criteria shown on the buttons on the left.](image)

Our experiments with the showcase family of browsers have included browsers for the following items: pill identification, conference delegates, historical buildings, crayon drawings, research faculty, and wasp wing features. There are some particular requirements for the information that is useful for this kind of browser. There should be a single image that can be used to represent every item in the collection. The metadata should consist of faceted items, where each collection item has a single value for each facet, although it is possible to construct sets of items that meet multiple matches. Eye color was used for the conference delegate browser, for example, since most people have a single eye color. In the rare cases of people who have one blue eye and one green one, we have to address the design question, Should these people have their images multiplied when necessary, so that they go into both groups, or should they form a new, third group? We typically opt for the creation of a new group, which happens automatically, since the system constructs the groups based on the available metadata. If there was an error in the metadata, for instance, and someone had an eye color of “square,” then an eye color group with the group heading “square” would appear with one item in it (we have repeatedly found that visualization interfaces can also serve as useful checks for the integrity of data).

Although we originally conceived of the showcase browser as a browser for images and their associated metadata, we have subsequently developed a version where the photos of pills or people or buildings have been replaced with tiles containing text—something like the statistics on the reverse side of sports trading cards. In this case, the tiles show the bibliographic data for the items in a text collection. Our current design, as seen in figure 7, is dynamic and allows the researcher to add or subtract metadata from the tiles, so that one researcher might decide to use tiles that show...
author name, title of work, and date of publication, whereas another might omit authors and dates and show instead the genre and word count (persistence of data and user-rights management can be challenging aspects of some visualization interfaces, especially Web-based ones). The number of kinds of metadata is determined by what is available in the archive.

Figure 7. This image shows a screenshot of a digital image browser for text collections (designer Ian Craig; programmer Alejandro Giacometti).

25 Using this kind of faceted browser, the display of the complete set of items is successively reduced as the user chooses how to group the items and then selects the group of interest. Typically, instead of removing unselected items entirely, we create a subset of them at a much smaller scale, located at the bottom of the screen (the trail of remaining smaller items is an important representation of intellectual work that has been accomplished, as argued by Vannevar Bush in 1945 in his description of the links followed in the Memex machine, a precursor to the Web). The experience is one of visually sifting from a large body of data to a smaller one until the remaining items represent the subcollection of interest.

26 The showcase design would not facilitate the kind of task described for Mandala, where the goal was to find the speeches in several plays that might support the hypothesis that Lear is transformed on the heath into a supernatural figure. An appropriate research task for the showcase design would be the creation of a work set consisting of a smaller number of items from a large collection, in which the bibliographic characteristics of the items are the distinguishing features. Someone interested in sifting through a collection for plays by women authors in the eighteenth century, for example, could locate this subset by subdividing the collection by genre, then selecting the group containing plays. This group could be subdivided by the sex of the authors; then, the user could choose the subset by women. Finally, the user could divide the plays by women according to date of first publication or performance, depending on what kinds of dates the archive made available. Choosing those plays from the eighteenth century would complete the process.

27 This example described steps that proceeded from genre to sex to date, but since the data is faceted, the user is free to carry out the steps in any order. At each intermediate stage, the resulting subsets will be significantly different (and potentially generative of different insights), depending on the sequence the user follows. From the perspective of the interface designer, providing this variety of choices means that it is not necessary to predict what the most likely sequence will be. The user is not navigating a hierarchy of information in any conventional sense but is instead working through a series of faceted subdivisions that are dynamically grouped at each step. The interface is enabling iterative steps that can fit into a fluid interpretive process.
The experience of navigating a collection using a showcase browser is typically somewhat different from the experience of navigating a collection using a menu hierarchy. Both systems work by successively subdividing the collection, but in a showcase browser the order of the subdivision is chosen by the user. It is therefore optimal to use a showcase browser for collections and their associated metadata where there is a one-to-one relationship between the subcategories in the metadata and the individual collection items. Most literary works, for instance, can be categorized under a single genre: *Middlemarch* is a novel. However, if a work has more than one genre, then it is necessary to duplicate that work so that it will appear under both genre categories (or create new, hybrid categories): *Oklahoma!* is both a musical and a western. We see how decisions by designers of a tool can have a critical impact on ontological and epistemological aspects of research.

**Revealing Features**

Browsing a collection can be a fascinating and useful scholarly activity, but it is typically only one early step in a longer scholarly process of reading and interpretation. Once the user has settled on relevant results after having done a retrieval task or browsed a collection, what remains is a list of works. A typical approach to this list is to begin reading at the top and through to the end, perhaps making notes along the way and beginning the process of interpretation that may lead to a research result. Designers of humanities visualizations, however, hope to produce systems that can also assist in these subsequent steps. We already have some evidence that visualizations of this kind have a role to play (Rockwell and Bradley).

Grouping tools like the *Mandala* browser can also be used in this context, since one of the useful mental processes in interpretation is to group similar things together and see what kinds of patterns, if any, emerge from the groupings. The activity is usually iterative, leading to many dead ends before something worth pursuing further begins to emerge.

The *Mandala* browser, however, is suitable only for some of these explorations, since it proceeds by small increments defined by the researcher. Someone looking for material enclosed in a particular XML tag and combining it with a search term can benefit from the visualization of *Mandala*. But *Mandala* is not currently designed to go beyond Boolean combinations of specific tags and specific search strings. The system cannot produce a set of magnets, for example, that represent a semantic class rather than a single word. This situation is common to many tool sets: they enable some processes but not others, and often it is best to think of how to use several tool sets in combination.

The researcher who is interested in seeing all the speeches in Shakespeare that mention the natural elements either has to be lucky enough to find them grouped by the collection developers into one tag or has to list them individually. If the terms are used to define individual magnets (it is possible to define magnets that correspond to one of several terms), the display quickly becomes overwhelmingly complex, and there is always the chance that some key elements might be missed.

Other systems circumvent this problem. Two strategies in particular hold significant promise: clustering and classification. We experimented extensively with both in the *MONK* project (*Metadate Offer New Knowledge*). Clustering requires some fairly sophisticated logic in the system, allowing the reader to identify a set of texts and then to have the interface show how the texts—in variable units, such as paragraphs—group together into automatically identified clusters. There are dozens of available algorithms for performing this kind of clustering, which emphasize different features of the works being clustered. *Mallet*, developed by Andrew Kachites McCallum, is a widely used topic-modeling tool, although it is less often used in the digital humanities, partly because it is neither especially user-friendly nor designed for humanistic inquiry. As Martin Mueller, Jean-Frédéric de Pasquale and Jean-Guy Meunier, and others have pointed out, for a system to be useful to literary scholars, the clustering must involve more than just an automatic identification of topics, because, in literary studies, how something is said is often as significant as what is said.

In the classification tool set ([fig. 8](#)), *MONK* provides an interface that uses a variety of supervised classification algorithms to find similarities between works or parts of works that have been suggested by the user (supervised classification is a process in which the user identifies characteristics of a subset of items and the system uses this
training set to classify the remaining items). The result is a tool set (“Search by Example”) with which the user can collect a set of documents that have some common feature of interest, then see how the system views the similar features of those works or parts of works and also what other works in the collection share those features. In one case study, for example, Catherine Plaisant and her colleagues identified a set of the poems by Emily Dickinson that had erotic content (i.e., “hot” versus “not hot”), then used a classification system to see what those poems had in common with one another (e.g., an unusual prevalence of possessive pronouns) and also what other poems the system thought should belong in the set. Other projects are looking at subjects such as the sentimental in novels (Steger), language in literary works versus newspapers (Horton), and reports of curses and spells in reports from early modern witchcraft trials (Uszkalo).

Figure 8. The MONK workbench, designed by Milena Radzikowska and Stan Ruecker and programmed by Amit Kumar, Andrew Macdonald, and Stéfan Sinclair, combines tools into tool sets that allow the user to carry out a variety of multistep research tasks with literary texts. In the classification tool set shown here, the user creates a set of passages representing some phenomenon of interest, and the system suggests similar documents while also making the set of distinguishing features available for examination.

It is worth saying a few words about the tool-set design of MONK. Although individual tools can be helpful, often a scholarly process can take advantage of more than one tool in sequence. Several digital humanities projects have therefore made use of the idea of recipes or tool sets, which guide the researcher through a process where some steps are facilitated by tools and others are manual, and different kinds of data are accessed or produced with each step. Examples of these kinds of projects include TAPoR (Text Analysis Portal for Research), MONK, and JiTR (Just-in-Time Research). Similarly, some interfaces are designed such that multiple modular tools can interact, even if they do not appear in sequential steps; the Voyant Tools interface is an example, as with this corpus from Shakespeare. Both tool sets and modular tools are conducive to iterative, interactive processes that depart from a simple model of data input and output.

Time and Space

Another general class of phenomena deals with patterns involving spatial organization, chronological sequence, or both. Some standard approaches to visualizing this kind of information exist. For spatial data where a map is involved, geographic information systems provide considerable flexibility in plotting data of various kinds (see, e.g.,
F. Black, MacDonald, and J. Black), and there is a growing interest in areas such as literary cartography. For chronological data, the timeline is a venerable visual format, whether manifested statically or interactively. Examples include the Simile project from MIT, ThemeRiver from Pacific Northwest Laboratories, and the somewhat more unusual Spiraclock from L’Ecole des Mines de Nantes. The Scholars’ Lab has recently released Neatline, an Omeka-based tool that “allows scholars, students, and curators to tell stories with maps and timelines.” The nondeterministic and interpretive aspects of Neatline are essential and make it an excellent example of how tools developed by digital humanists can differ from those developed in more purely scientific disciplines.

Yet not all spatial data can or should be related to geography, and in some cases the sequential data and the spatial arrangement are related to each other. In Watching the Script (fig. 9), for example, a reader can read the speeches of a play and also think about the blocking of a particular theatrical performance (Sinclair, Ruecker, Gabriele, and Sapp). This interface represents characters as colored circles that are positioned on a stylized stage and speeches as scrolling text. Student directors can arrange the blocking, adding annotations about their reasons for putting each character in a given location. Readers can choose which portions of the play or which characters to watch and also control the playback speed.

Figure 9. Watching the Script, designed by Sandra Gabriele and programmed by Stéfan Sinclair and others, draws on XML-encoded plays to allow students, directors, and actors to see a stylized reproduction of a play or screenplay that includes the blocking information.

Recently, we have been working on reimplementing this interface in a prototype called the Simulated Environment for Theatre (SET), which uses a gaming engine to provide a more realistic 3-D experience (fig. 10). Building on the design of previous visualizations for theater, such as Michael Best’s Scenario, SET allows the user to choose the viewpoint for seeing the stage, originating from anywhere in the theater (Roberts-Smith et al., “Visualizing Theatrical Text”). The development of SET was in large part motivated by feedback from domain experts in theater who, while recognizing the value of our initial interface, felt that the potential for the interface was limited from a practitioner’s point of view, especially because of its focus on text rather than on time and space.
Typographic Form as Interactive Visualization

Although many visualizations introduce additional graphic elements, it is also possible to use only the organization of words to convey meaning. A classic example is W. Bradford Paley’s TextArc, which combines words plotted in sequence around the periphery with the words that appear more than once plotted in the interior of the circle according to their average position, sometimes called a weighted centroid (fig. 11). The TextArc interface is less about allowing the user to generate different visualizations based on specified parameters and more about representing the text in a novel way. TextArc is a time-based visualization in which the system reads the original text and produces a curving line that navigates within the textual space depending on where each word is plotted.
Among our projects, we have a variety of interfaces where text is privileged as a visual indicator of data, including a series of repetition graphs, designed by Piotr Michura, that show patterns of repetition in the context of the rest of the document (fig. 12). By plotting the words horizontally in sequence but shifting down one row for every repetition, the graph results in a visual thumbnail of a book where repeated phrases show as sharp, steep curves (Michura, Ruecker, Radzikowska, and Fiorentino). This interface is a radically different view of the text that can help provoke new interpretive insights. The glyphs in figure 12 are actual words, and, although they are illegible at this scale, they convey graphical meaning. The first text is an excerpt from Shakespeare’s *Hamlet*, and the second is from Gertrude Stein’s *Making of Americans*; Stein’s propensity for repetition and Shakespeare’s relatively high vocabulary richness are neatly summarized and confirmed by the graphs.
Interactive Glyphs

To this point, our discussion of visualizations for humanities researchers has focused on displays where image and text are treated as primary visual objects. Another set of our experiments are slightly more conventional scientific visualizations, in that they deal with various visual objects that stand in place of numeric data. These kinds of visualizations allow the user to see the forests that might otherwise appear to be a lot of individual trees.

For the literary scholar interested in studying patterns, there are two related but distinct needs. First, an interesting pattern of some sort must be identified. Stan Ruecker, for example, described how, in reading the letters of Lady Mary Wortley Montagu, he was struck by the repeated introduction of absurd fantasies. Montagu had a wide range of correspondents, and, in her letters to many of the principal ones, she introduced an absurd motif. In more than one instance she proposed leaving society to run off together to a desert island. In writing to her daughter from Italy, she develops at some length the idea that the local people have identified her as a witch, and so maintains that she is
considering taking up witchcraft. The presence of these absurd fantasies suggests several questions: What does it mean that this writer chooses to suggest them? What purpose might they serve? What implications does it have for her relationship with her correspondent, and what does her correspondent's response reveal about the nature of that relationship? Looking further afield, is Montagu the only letter writer of her period who includes absurd fantasies? Do writers from other periods use them too?

The last two questions lead to the second need of the scholar interested in identifying and studying patterns: knowing the scope of a pattern. Is it idiosyncratic to a particular writer, genre, historical period, and so on, or is it a widespread phenomenon? And if it is widespread, is it subject to meaningful local variations? To take another more general example, many writers employ extended metaphors. In the case of metaphor, the use is general, spread across genres and historical periods. But the localized use of particular metaphors by a given writer can be significant enough to be of scholarly interest, especially when understood against a comparative background.

A visualization intended to show patterns of interest should therefore provide these two features: a way to find possible patterns to investigate and a means to locate those patterns against a baseline provided by a set of other relevant works. What constitutes the set of relevant works will vary from project to project and needs to be selectable by the scholar. For some projects, such as tracing absurd fantasy in Montagu's letters, it may be important to compare a pattern in one work with any instances of that pattern in other works of the same genre and century. For other projects, it may be more useful to look at a pattern in one work against a comparative background of other works by the same author. Still another project might be interested in identifying patterns across an entire collection, without reference to a particular work or author.

An example of this kind of complex visualization is the MONK project’s lexical-bibliographic glyph, which we placed as the central element in a scrolling list of significant words, lemmas, or parts of speech (fig. 13). The system identifies the list of words and associates them with a background of bibliographic data. As the scholar scrolls through the list of words, lemmas, or parts of speech, for each item in the list the system produces a circular object—an information glyph. The size of the circle in the middle of the glyph is related to the relative significance of the item and therefore changes from item to item. Plotted in segmented rings around this central circle are the bibliographic data, which break down the collection by work, author, date, genre, and so on, depending on which rings the user thinks might be useful for comparison.

Figure 13. The lexical-bibliographic glyph, designed by Carlos Fiorentino as part of the MONK project, allows the user to simultaneously
view combinations of data from the micro level of the word, lemma, or part of speech with the macro level represented by the bibliographic information associated with the works. In this screenshot, the user can see the relative frequency of the selected term *say* in the works of Stein, Blake, and Dickinson and then further explore the associated metadata.

The conceptual complexity and analytic potential of a visualization interface such as the lexical-bibliographic glyph come at the cost of accessibility; we are a long way from the relatively intuitive Wordle visualization. We believe that through careful visual design and iterative user testing it is possible to create complex visualization tools that are useful to a wide range of scholars, although any one tool is unlikely to satisfy all types of humanities scholars. Rather, we should strive to create a full ecology of visualization tools that are sensitive to the particularities of humanities research and that offer a range of analytic functionality, from the most basic and recognizable to the most sophisticated and defamiliarizing.

A primary index to the quality of visualizations for humanities scholars is the quality and originality of scholarship that the systems support. In each of the projects mentioned here, we have been working with humanities researchers in an effort to produce a useful visual form of the data. Since humanities scholarship is often exploratory, we have also come to believe that interactive formats are in most cases preferable to static ones, since they allow the person using the system to add and subtract elements, experiment with different forms, pursue hunches or insights, and so on. It is therefore important that the expectations of the scholar correspond to the affordances of the visualization. It is important for the scholar to know enough about the visualization tools to understand that the interpretive work is being guided and biased by the data and software. Failing that, we need to have methodologies that are sufficiently well tested and understood for scholars to be able to use the tools with confidence. The question remains whether humanistic inquiry lends itself well to well-trodden methodologies when originality and idiosyncrasy are the norm.

**Works Cited**


Harris, Jacob. “Word Clouds Considered Harmful.” *Nieman Journalism Lab*. Nieman Foundation for Journalism,


DOI: 10.1632/lsda.2013.6

---

**Comments**

1 Comment on the whole page

1. [Arnim Alex Seelig](#) March 27, 2013 at 2:52 pm

   Thank you for the fine work, Stéfan, Stan, and Milena! A great entry for an exciting new publication venue, as is the whole anthology.

5 Pingbacks and trackbacks

1. [CCR 733: Rhetoric, Composition, and Digital Humanities (sp14)](#) | [Collin Gifford Brooke](#) August 10, 2013 at 9:55 pm

3. Ways of Seeing and Reading: The Use of Data Visualization in Literary Studies | The New Humanities: English 5850
December 12, 2013 at 6:13 am

[...] Stéfan Sinclair, Stan Ruecker, and Milena Radzikowska discuss the issues and benefits of using data visualization in humanities studies. One of the biggest problems within humanities studies is that, as with the CASS work on representation in media, researchers often “[begin] with no hypothesis about what [is] happening in the data” (Baker, Gabrielatos, & McEnery). Sinclair, Ruecker, & Radzikowska suggest that “[i]nformation visualization for humanities scholars needs to accommodate a mix of evidence and argumentation” (para. 1). Since humanities research is not generally falsifiable, any tool for research needs to offer the ability to “examin[e] the objects of study from as many reasonable and original perspectives as possible to develop convincing interpretations” (Sinclair, Ruecker, & Radzikowska). This group suggests that data visualization processes, particularly interactive processes, are perfectly poised to aid in such research at least as much as they aid in presentation. Interactive visualizations offer a representation of large data sets and allow the researcher the opportunity to go back and make changes as they data offers new arguments and perspectives. Ultimately, they decide that “visualizations can be useful to humanities scholars by providing additional insight into small amounts of text/data, thus supporting what John Unsworth calls ‘scholarly primitives,’ especially for showing patterns that result from filtering, sorting, grouping, and otherwise visually rearranging the material” (para. 10). [...]
Like other Mapping the Republic of Letters projects, ours has included numerous phases. One of the first issues we confronted as early as 2008 was the need to limit our scope in terms of the volume of letters we dealt with at one time. After a few years of careful and copious data transcription of the over 15,000 letters that Benjamin Franklin either wrote or received over the course of his life, the entirety of Franklin's correspondence proved too large a mapping endeavor for our first go at it. At this point, we stepped back to examine our goals for converting Franklin's correspondence into "data" (the easily computer-digested dates, correspondents, locations of letters sent and received). What questions did we hope to answer by compiling such "data" from Franklin's letters? Did such questions lend themselves to an examination of a smaller set of letters?

After much deliberation, the Franklin team decided to limit our focus—for the time being—to the years between 1756 and 1763 when Franklin traveled to England and Scotland for the first time (Franklin sailed in 1757, but we extended our focus to a few month before and after his six year trip). The years between 1756/7-1762 were an exciting and formative moment in Franklin's life. During these years, Franklin met many leading intellectual luminaries of his day, including David Hume and Lord Kames. In the two years since we began the intensive processing of Franklin's correspondence between 1756 and 1763, we have formulated new and important questions about the nature of Franklin's correspondence network—questions that lend themselves to the deeper and more precise nature of our investigation now structured around Franklin's first trip to London.

These explorations have also lent themselves to thinking broadly about the republic of letters and where the Americas fit into it. For an in-depth survey of these matters, see: Caroline Winterer, "Where is America in the Republic of Letters?" Modern Intellectual History 9, 3 (Nov. 2012): 597-623.
Current Goals and Interests of the Project

Using the collection of Franklin's correspondence found online, we extracted from Franklin's correspondence the following information about each letter: its date, its author and recipient and their genders, communities (aka “professions” in an eighteenth-century sense of the word), and places of birth; its source location and the location of its recipient. Once we had this information from each letter, we could shift our attention from the physical letters (Franklin’s correspondence) to the people behind the letters (Franklin’s correspondents) By engaging both with the letters and the people writing and receiving the letters, our team could formulate and answer two sets of questions: (type 1) those about the physical letters themselves and (type 2) those about the correspondents of whom Franklin’s correspondence network was comprised between 1756-1763. Thinking about Franklin’s correspondents themselves required that we separate out each individual correspondent from the “letters” spreadsheet to create a separate “Individuals” spreadsheet. Armed with these two spreadsheets, we could begin the process creating computer-generated visualizations that enabled us to formulate and begin to answer very specific questions about the nature of Franklin’s correspondence and correspondents network.

A few examples of type 1 questions with some preliminary conclusions reached by examining visualizations from the “letters” spreadsheet:

During what months did Franklin receive letters from particular locations (i.e. England, Scotland, and America) and how did the numbers of letters from one location relate proportionally to letters from other locations?

One conclusion drawn:

- In December 1756, before Franklin set sail for London, 100% of letters he received came from America.
- In January 1762, when Franklin is overseas, 100% of letters he received came from England.
- In December 1763, when Franklin had returned from London, over 75% of letters Franklin received were from America.

![Letters Written to Franklin, 1757-1763 by Month by Source Country](image)

When did Franklin receive letters from Scotland?

Franklin received letters from Scotland in only four months between 1757-1763: in September 1759, October 1759, March 1761, and May 1762.
From what countries did Franklin receive most of his letters between 1757 and 1763?

Franklin received nearly all his letters from two places: British America and England. Very few other places are represented, challenging us to think about the timing of Franklin's ascent into the ranks of the "cosmopolitan" on the world stage.

A few examples of type 2 questions with some preliminary conclusions reached by examining visualizations from the "correspondents" spreadsheet:

Who were Franklin's top correspondents by volume of correspondence?

1. Isaac Norris
2. Mary Stevenson
3. David Hall
4. Deborah Franklin
5. William Strahan
Grouped by gender, who were Franklin's top correspondents?

1. Male: Isaac Norris, David Hall, William Strahan, Peter Collinson
2. Female: Mary Stevenson, Deborah Franklin, Jane Mecom, Elizabeth Graeme

Grouped by country of birth, who were Franklin's top correspondents?

1. David Hall and William Strahan appear to be Franklin's top two "Scottish" correspondents, but neither one actually lived in Scotland at the time of Franklin's trip there.
2. Mary Stevenson, the daughter of Franklin's London landlady, Margaret Stevenson is Franklin's top "English" correspondent during this time.

With what community association or professional group did Franklin correspond the most? (Top five)

1. professionals
2. secular officials
3. professionals and secular officials
4. artisans
5. church officials

Where do women rank in his network?
Franklin wrote most of his letters to men.

With these conclusions from the visualizations of Franklin's correspondence network, our team is generating a second set of questions, which will result in a new interpretation of Franklin's presence on the world stage during this first trip to London.

Initial Question Answers from the visualizations of the data Level 2 Question inspired by the Answers from the Initial Question

Interpretative Conclusions about the Level 2 question that help us understand Franklin in the context of his eighteenth century world.

Defining Terms: "Cosmopolitanism"

Some of the level 2 questions we are engaging with now include the following: To what extent was Benjamin Franklin a "cosmopolitan" and how might we begin to define this term for maximum analytic utility during the eighteenth century? What can Franklin's correspondence network tell us about the relationship between centers and peripheries in the eighteenth century? Are there significant or tangible differences between a process of Anglicization (someone like Franklin in a colonial periphery engaging more with and becoming integrated into centers like London and Edinburgh) and becoming a "cosmopolitan"?

One vivid example of how Franklin's correspondence network reveals the prospects and limits of "cosmopolitanism" is to compare it with Voltaire's. The geomap below show Franklin's network (top) and Voltaire's (bottom). Many of Franklin's letters crossed the Atlantic; only a few of Voltaire's did. One could argue that Franklin was more cosmopolitan than Voltaire because his letter network was more trans-Atlantic; yet any scholar working in the Enlightenment would vigorously dispute that conclusion, arguing that other measures should also help us to define cosmopolitanism. Visualizations of these kinds, in short, challenge us to think about how we might define analytic terms such as "cosmopolitanism" more rigorously, including both impressionistic measures and numerical measures.
Note: Images are sketches, they need refinement and need to be checked for accuracy.

How to cite this page:
Claire Rydell and Caroline Winterer, “Benjamin Franklin’s Correspondence Network, 1757-1763,” Mapping the Republic of Letters Project, Stanford University, October 2012

Related

Voltaire’s Correspondence
View an instance of Ink loaded with Voltaire’s correspondence. In combines mapped letters with a timeline stacked bar chart and a relationship viewer.

Voltaire’s Publications
Mapping Voltaire’s publications with data from the BNF.

D’Alembert’s Correspondence
See a demo of D’Alembert’s correspondence network using Ink.

Humanities+Design
Visit the laboratory web site to learn more about the visualizations.
Text Analysis of Martha Ballard’s Diary (Part 1)

“mr Ballard left home bound for Oxford. I had been Sick with the Collic. mrs Savage went home. mrs foster Came at Evening. it snowd a little.”

This is the first entry in the diary of Martha Ballard. Martha Ballard was a rural Maine midwife who kept an extensive diary between 1785 and 1812 and whose life was immortalized in 1990 by the historian Laurel Thatcher Ulrich’s award-winning *A Midwife’s Tale*. Over the course of three decades, Ballard kept a meticulous, near-daily accounting of her life spanning over 10,000 entries.

When reading *A Midwife’s Tale*, I was struck by how readily the text would seem to lend itself to digital analysis. In an interview, Ulrich noted, “The very thing that had attracted me to the diary in the first place was also the thing that made it difficult to work with. I mean there’s just so much.” To ground herself, she began by simply counting things: “And I would go day by day for every other year of the diary, and I would tick off what was in each entry: baking or brewing, spinning or washing, or trading, sewing, mending, deliveries, general medical accounts, going to church, visitors, people coming for meals, etc.” Because of the sprawling scope, she took this quantitative approach only for the even-numbered years in the diary. The fact that she was working in the late eighties without a computer...
makes her work even more impressive.

After poking around online I came across DoHistory.org, a website developed and maintained by the Film Study Center at Harvard University and hosted by (who else, really) George Mason’s CHNM. The website presents the diary to the public in two formats: the viewer can either browse through photographed pages of the diary or read the transcript of the pages (transcribed through a monumental effort by Robert R. McCausland and Cynthia MacAlman McCausland):

When I realized the entire diary was online, it got me thinking about possibilities for text mining. As an aspiring digital humanist with little “hard” skills beyond basic GIS, I had been meaning to learn how to program for quite some time. In Martha Ballard’s diary, I had an intriguing source of data with which to learn how to do so. Now I just had to learn how to program. With the patient help of several programming-savvy family members, I gradually learned the basics of Python and how to apply it to Martha Ballard’s diary. What follows are the first steps we took to process the diary’s raw data into an accessible digital format.

**Process**

At first, I briefly considered learning how to scrape the text of the diary off the
website. After some investigation, I decided that was a little beyond my abilities, so I copped out to the much easier route of sending an email to Kelly Schrum at CHNM, who kindly forwarded my request to Ammon Shepherd, who emailed me a zip file containing 1,431 html documents, one for each page of the diary. The html files of the transcribed diary are a basic, 3-column table that look this. My first step was to find a way to strip out the html tags and organize the text into a systematic database of individual entries. Fortunately, Ballard’s meticulousness and consistency lent itself well to such an approach.

The diary’s format translates quite nicely into creating a list of lists – the “main” diary being a list of all the entries, and each entry being a list in and of itself. The first program we wrote was to open each html file and begin extracting the different sections of text (which were conveniently marked by html tags). Iterating through each entry allowed us to separate the different columns in her diary into different items in the list. Here is the breakdown of our “list of lists”:

1. Diary
   1. Entry
      1. Date
         1. Month
      2. Day
      3. Year
   2. Day of the Week
   3. Main Text of Entry
   4. Day Summaries (Column 3 of actual diary entry)
   5. Birth(s) (Recorded in Column 1 of actual diary entry)

In creating the list, we had to separate out the raw data from the html tags that
formatted it. Fortunately, the folks who built the html files originally used an extremely systematic formatting process that actually made the job of distilling one from the other quite straightforward. A Python module called Pickle allowed us to export the list of entries as a manageable single file that we could then easily import into future programs to manipulate.

For example, the third entry in the diary would translate a bit into something like this:

1. Diary
   1. Entry (3)

   1. Date
      1. 1 (January)
      2. 3
      3. 1785

   2. 3 (Tuesday – Ballard numbered the weekdays, beginning with Sunday as 1)

   3. “Tuesday. mrs. Foster went home. I had threats of thee Collic; by takein peper found releif.”

   4. Empty

   5. Empty

The list allows us to access pieces of information by “calling” their position. It helped me to think of the entire diary list as a warehouse containing almost 10,000 boxes (entries) inside it, with each box containing five compartments, with the first of those compartments divided into three sub-compartments. If you were to open any of the boxes (entries) and look inside the first compartment, then inside sub-compartment number two, you would always find a number that
represented the month of that particular entry. If you were to look inside the third compartment of the entry/box, you would always find the main text for that day’s entry.

The advantages of setting up the data in a list structure is the ability to access these specific pieces of information easily and to compare them across entries. In many ways, processing the text to make it readable and programmable is one of the biggest challenges to text mining. Deciding on the most logical way to organize and break down over 1,400 files will lay the groundwork for the fun part: writing programs to actually analyze the diary of Martha Ballard.

***Special-edition sneak preview of future posts in this series***

A simple counting program reveals that the main text of Martha Ballard’s diary alone contains 377,315 words, spanning I-couldn’t-make-this-number-up 9,999 entries. That is a lot of data to play with.

---

**#A Midwife's Tale #Laurel Ulrich #Martha Ballard #Programming #Python #Text analysis**

---

**Comments**

**Ben Brumfield - August 31, 2009 @ 9:48 am**

I’m looking forward to see what you come up with. I’ve made a couple of stabs at mining diary data in a naive way on my own project, but suspect that my approach is too strongly influenced by the text I’m working with.

* Cameron Blevins - August 31, 2009 @ 11:28 pm

  Ben,

  Thanks for the comment – I think it’s pretty normal for methodological approaches to get shaded by the text, which I admit can be both a good and bad thing. From what I understand you have a far
more sophisticated grasp of programming than I do, so I welcome any and all advice from your own experience.

-Cameron

Ben Brumfield - August 31, 2009 @ 11:59 pm
That’s a bit comforting. I’d love to chat with you about this, but suggest that we wait until you’ve posted your observations. I’m very interested to see what sort of data you think is extractable.

Kelly in Kansas - September 3, 2009 @ 7:26 am
Thanks for letting us follow your interesting journey virtually. We will all learn from your experience.

- McCausland - July 6, 2010 @ 10:46 am
You boys have taken on one hell of a job
and I for one would consider it a personal favor if
you would alert me when you have this project well under way.
I’m the guy (with Cyn) that transcribed the 9,999 words that Martha wrote.
Robert McCausland

- Cameron Blevins - July 7, 2010 @ 8:56 am
Thanks Rob! Good to hear from you again. I’ll definitely be keeping you posted.

-Cameron

Leave a Reply
Your email address will not be published / Required fields are marked *

Name*

Email*
Text Analysis of Martha Ballard’s Diary (Part 2)

Given Martha Ballard’s profession as a midwife, it is no surprise that she carefully recorded the 814 births she attended between 1785 and 1812. These events were given precedence over more mundane occurrences by noting them in a separate column from the main entry. Doing so allowed her to keep track not only of the births, but also record payments and restitution for her work. These hundreds of births constituted one of the bedrocks of Ballard’s experience as a skilled and prolific midwife, and this is reflected in her diary.

As births were such a consistent and methodically recorded theme in Ballard’s life, I decided to begin my programming with a basic examination of the deliveries she attended. This examination would take the form of counting the number of deliveries throughout the course of the diary and grouping them by various time-related characteristics, namely: year, month, and day of the week.

Process and Results

The first basic step for performing a more detailed text analysis of Martha Ballard’s diary was to begin cleaning up the data. One step was to take all the words and (temporarily) turn every uppercase letter into a lowercase letter. This
kept Python from seeing “Birth” and “birth” as two separate words. For the purposes of this particular program, it was more important to distill words into a basic unit rather than maintain the complexity of capitalized characters.

Once the data was scrubbed, we could turn to writing a program that would count the number of deliveries recorded in the diary. The program we wrote does the following:

1. Checks to see if Ballard wrote anything in the “birth” column (the first column of the entries that she also used to keep track of deliveries)

2. If she did write anything in that column, check to see if it contains any of the words: “birth”, “brt”, or “born”.

3. I then printed the remainder of the entries that contained text in the “birth” column but did not contain one of the above words. From this short list I manually added an additional seven entries into the program, in which she appeared to have attended a delivery but did not record it using the above words.

Using these parameters, the program could iterate through the text and recognize the occurrence of a delivery. Now we could begin to organize these births.

First, we returned the birth counts for each year of the diary, which were then inserted into a table and charted in Excel:

At the risk of turning my analysis into a John Henry-esque woman vs. machine, I compared my figures to the chart that Laurel Ulrich created in *A Midwife’s Tale* that tallied the births Ballard attended (on page 232 of the soft-cover edition). The two charts follow the same broad pattern:
Note: I reverse-built her chart by creating a table from the printed chart, then making my own bar graph. Somewhere in the translation I seem to have misplaced one of the deliveries (Ulrich lists 814 total, whereas I keep counting 813 on her graph). Sorry!

However, a closer look reveals small discrepancies in the numbers for each individual year. I calculated each year’s discrepancy as follows, using Ulrich’s numbers as the “true” figures (she is the acting President of the AHA, after all) from which my own figures deviated, and found that the average deviation for a given year was 4.86%. Apologies for the poor formatting, I had trouble inserting tables into WordPress:

<table>
<thead>
<tr>
<th>Year</th>
<th>Deliveries Count</th>
<th>Difference</th>
<th>Deviation (from Ulrich)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Manual (Ulrich)</td>
<td>Computer Program</td>
<td></td>
</tr>
<tr>
<td>1785</td>
<td>28</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>1786</td>
<td>33</td>
<td>35</td>
<td>2</td>
</tr>
<tr>
<td>1787</td>
<td>33</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>1788</td>
<td>27</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>1789</td>
<td>40</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td>1790</td>
<td>34</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>1791</td>
<td>39</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>1792</td>
<td>41</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>1793</td>
<td>53</td>
<td>50</td>
<td>3</td>
</tr>
<tr>
<td>Year</td>
<td>Deliveries</td>
<td>Total</td>
<td>Difference</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>1794</td>
<td>48</td>
<td>48</td>
<td>0</td>
</tr>
<tr>
<td>1795</td>
<td>50</td>
<td>55</td>
<td>5</td>
</tr>
<tr>
<td>1796</td>
<td>59</td>
<td>56</td>
<td>3</td>
</tr>
<tr>
<td>1797</td>
<td>54</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>1798</td>
<td>38</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>1799</td>
<td>50</td>
<td>51</td>
<td>1</td>
</tr>
<tr>
<td>1800</td>
<td>27</td>
<td>23</td>
<td>4</td>
</tr>
<tr>
<td>1801</td>
<td>18</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>1802</td>
<td>11</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>1803</td>
<td>19</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>1804</td>
<td>11</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>1805</td>
<td>8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>1806</td>
<td>10</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>1807</td>
<td>13</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>1808</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>1809</td>
<td>21</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>1810</td>
<td>17</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>1811</td>
<td>14</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>1812</td>
<td>14</td>
<td>14</td>
<td>0</td>
</tr>
</tbody>
</table>

Keeping the knowledge in the back of my mind that my birth analysis differed slightly from Ulrich’s, I went on to compare my figures with other factors, including the frequency of deliveries by month over the course of the diary.
If we extend the results of this chart and assume a standard nine-month pregnancy, we can also determine roughly which months that Ballard’s neighbors were most likely to be having sex. Unsurprisingly, the warmer period between May and August appears to be a particularly fertile time:

Finally, I looked at how often births occurred on different days of the week. There wasn’t a strong pattern, beyond the fact that Sunday and Thursday seemed to be abnormally common days for deliveries. I’m not sure why that was the case, but would love to hear speculation from any readers.

Analysis

The discrepancies between the program’s tally of deliveries and Ulrich’s delivery count speak to broader issues in “digital” text mining versus “manual” text mining:

Data Quality

Ulrich’s analysis is a result of countless hours spent eye-to-page with the original text. And as every history teacher drills into their students when conducting research, looking directly at the primary documents minimizes the degrees of interpretation that can alter the original documents. In comparison, my analysis is the result of the original text going through several levels of transformation, like a game of telephone:
Each level increases the chance of a mistake. For instance, a quick manual examination using the online version of the diary for 1785 finds an instance of a delivery (marked by ‘Birth’) showing up in the online HTML, but which does not appear in the “raw” HTML files our program is processing and analyzing.

On the other hand, a machine doesn’t get tired and miscount a word tally or accidently skip an entry.

Context

Ulrich brings to bear on the her textual analysis years of historical training and experience along with a deeply intimate understanding of Ballard’s diary. This allows her to take into account one of the most important aspects of reading a document: context. Meanwhile, our program’s ability to understand context is limited quite specifically to the criteria we use to build it. If Ballard attended a delivery but did not mark it in the standard “birth” column like the others, she might mention it more subtly in the main body of the entry. Whereas Ulrich could recognize this and count it as a delivery, our program cannot (at least with the current criteria).

Where the “traditional” skills of a historian come into play with data mining is in the arena of defining these criteria. Using her understanding of the text on a traditional level, Ulrich could create far, far superior criteria than I could for counting the number of deliveries Martha Ballard attends. The trick comes in translating a historian’s instinctual eye into a carefully spelled-out list of criteria for the program.
One area that is advantageous for digital text mining is that of revising the program. Hypothetically, if I realized at a later point that Ballard was also tallying births using another method (maybe a different abbreviated word), it’s fairly simple to add this to the program’s criteria, hit the “Run” button, and immediately see the updated figures for the number of deliveries. In contrast, it would be much, much more difficult to do so manually, especially if the realization came at, say, entry number 7,819. The prospect of re-skimming thousands of entries to update your totals would be fairly daunting.

#A Midwife's Tale  #Digital History  #Laurel Ulrich  #Martha Ballard  #Programming  #Python  
#Text analysis

Comments

Ben Brumfield - September 13, 2009 @ 2:05 pm
What a great idea to perform the same analysis (birth count by year) that Ulrich did, and compare your results to hers! I’m curious if you tried the manual method for the years with the widest divergence to see exactly why your program disagreed with Ulrich 5 times in 1795. That might offer an opportunity to improve the algorithm, but would more likely illustrate the limitations of data mining via text searches, with some concrete examples of why some analysis is non-computable.

Given the content of the diary, I wonder if you could look for correlations between births and other events that Ballard mentioned in the text of her entries. For example, she mentions the weather in the early pages I’ve examined, so you might be able parse out her descriptions (looking for strings like “fine” or “snow”), assign weather values to dates, then look for correlations between the weather and the deliveries she attended. Other events may be harder to identify: Ballard mentions her own health, but also comments on other people who are unwell. This might make it impossible to correlate deliveries to Ballard’s health.

Thanks for posting such a detailed description of your work.

Cameron Blevins - September 14, 2009 @ 9:02 am
Ben,

Interesting ideas – comparing my program to Ulrich’s analysis certainly reinforced some of the limitations of data mining, but gave me hope in that it’s not so difficult to tweak the program and make it more effective.

I really like the idea of looking for correlations between births and other events. I think the next step for an in-depth and systematic analysis of the text would be to create first a dictionary of unique words, then start grouping the words together under different categories (Religion, Death, Marriage, etc.). From there it would be really cool to then look for patterns using those groupings. Unfortunately Ballard’s unique spelling system presents a challenge – she spells each word about 3-4 different ways, and has an incredible use of shorthand that contributes to around 37-38,000 “unique” words that would need to be cataloged. But if that gets done, the possibilities really become endless.

Thanks again for the support!

-Cameron

- Ben Brumfield - September 14, 2009 @ 8:53 pm

I’ve encountered similar challenges editing the Julia Brumfield diaries, where proper names are spelled inconsistently — sometimes within the same page. Because I’m identifying terms for indexing/analysis as I transcribe the text from scanned images, I can resolve the spelling irregularities during transcription/editing. However, I’ve still found that full-text searches will identify terms I missed during the mark-up phase, so I can’t say that my technique for data extraction is substantially better.

I like your idea of extracting words from the text to identify variant spellings. I presume you’d do a frequency count over the entire corpus and sort the word/count pairs alphabetically to look for variants.

Another possibility is to manually pull all the variant spellings of something you’re interested in (say, weather) from one year’s worth of entries. You could then execute the search against a different year and then manually identify missed variants there to see how representative your orginal sample was. Someone with more statistics than I command could probably come up with reasonable figures for your extraction algorithm’s accuracy. At any rate, you’d then be able to extract the data about that subject for your analysis.

The downside of my approach is that you can block yourself from discovering subjects to investigate. Because you set out with one topic of analysis in mind (weather, say), you might miss the sort of things that a high-frequency word list could suggest. In my own project, I did not identify clothes washing as a domestic activity worthy of analysis until I was around 500
Thank you for an interesting and helpful post. And yet it seems to confirm my fairly uninformed and perhaps knee-jerk reaction to a lot of these text mining projects—the conclusions are quite modest compared to the effort that went into the process. You have shown that Mainers had sex more often in the summer. No, you have shown that Mainers who hired Martha Ballard to midwife their babies were more likely to have had sex in the summer.

I hope I don’t come off as snarky, I truly am impressed by your technical abilities. I think that is exactly what makes me read your conclusions and say “Is that it?”

Ben Brumfield - September 14, 2009 @ 7:17 pm

Larry, I’m afraid that you’re confusing the technique (parsing and extracting data from the text) with the analysis (what you do with the data you’ve extracted, and whether you attempt to extract more data from the text). In this case, Cameron’s extracted births and dates from the text—a single fact (Ballard’s attendance at births) with a single dimension (the date the birth occurred). There’s not very much analysis he can perform on this data, since there are only a limited set of questions to be asked from it. So of course the conclusions are modest.

However, I’d wager that those conclusions—the graph of births per year—were determined through far less effort by Cameron than the effort spent by Ulrich to manually tabulate births by year. The fact that he’s able to compare his low-cost effort to Ulrich’s and such minor deviation lets us know the quality-to-cost trade-offs of his methodology. That in itself is worth knowing.

The real question is—having exhausted the interesting questions he can ask of the data he’s extracted—is there other data to be extracted from this text that might lend itself to more interesting analysis? How often was Ballard paid, and how many clients stiffed her? What were the geographic limits of her practice? If you had the misfortune to enter labor during a snowstorm, did that reduce the likelihood that you’d be attended by a midwife? If so, does weather explain the trough in births Ballard records between November and January, thus canceling out the summer-conception effect Cameron’s initial analysis finds?

Some of these may simply not be extractable from the kind of full-text search Cameron’s performing here—geography in particular requires contextual information about where her clients lived that is not internal to the text, and which we’re unlikely to have elsewhere. But it’s a useful exercise to figure out which of these questions are answerable, which are impossible, and why.

Cameron Blevins - September 14, 2009 @ 10:45 pm

Larry,
Thanks for the feedback, you raise some good points. One fundamental issue with text mining in the humanities has been a gulf between promise and delivery – there seems to be so many things that could potentially be done, but that in the end prove to be either impossible or involve even more work than doing it by hand. There’s also the issue of what I believe you termed “parlour tricks” on your blog, of analysis that may be superficially interesting or catchy, but adds little substantive value to the investigation. Both of these are fair criticisms.

In response to “Is that it?”, I’d say its a valid question to ask since the analysis I’ve done so far isn’t particularly deep, but that it’s a bit like watching a ten year old learning how to play basketball and saying “Okay, but can they dunk?” Much like a ten year old struggling to learn how to play a new sport, the process for me (admittedly somewhat selfishly) has been more about the learning experience than about producing earth-shattering results.

Having said that, even my limited experience so far has affirmed for me the potential and ability of text mining to study history. I’m fascinated by ways it can be applied that would be either impractical or impossible to accomplish manually. What I’ve done here can be done (and has, obviously) without the magic of computers. But in the hands of a more skilled programmer than I, text mining offers up the real ability for both deep analysis and a degree of flexibility that goes beyond the typical scope of traditional methodology. When paired up with the massive digitization projects going on already that lowers the barrier to processing digital data (and, I fully admit, presents its own issues and problems), I think the tradeoff between the quality of results vs. time/effort is going to continue to shift in favor of text mining.

Larry H Cebula - September 14, 2009 @ 11:07 pm

Cameron: Thanks for taking my comments in the friendly spirit in which they are intended.

As a profession, we have been here before. In the 1960s the term Cliometrics was coined. Historians created punch cards based on census data and city directories and so on. It was going to REVOLUTIONIZE EVERYTHING. But nothing much ever came from it so far as I know. The one book title that pops up in my mind is Fogel and Engerman’s Time on the Cross–a controversial book.

And yet–I am pretty sure that the application of digital technology is actually going to revolutionize everything–eventually. I want it to work.

Can you point me towards some historical text mining scholarship that has produced unique and compelling insights?

Ben Brumfield - September 15, 2009 @ 5:57 am

Larry, I think you make a fair point here. Text searching (not necessarily text mining) some parts of some scholarship, according to Patrick Leary’s Googling the Victorians, but in most cases it’s probably harder to figure out the questions to ask than to do the programming.
In my on project, for example, writing an analysis tool to look for correlation among subjects I'd already extracted was the matter of a single evening’s hacking. But is it really that insightful to see that stripping tobacco occurs alongside clouds and rain? Or that mentions of the tenant farmer are common next to plowing? So far the most use I’ve gotten out of the tool has been in identifying unfamiliar names during the annotation process by looking for the context in which they’re mentioned. Which is nice, but that’s only happened twice in a few hundred pages. Web searches for unfamiliar names have worked just as often.

Despite those modest—even disappointing—results, I’m not sorry I built the tool, not least because it required such a modest effort. I think that perhaps we’re moving beyond the model of large scale, resource-intensive text mining projects with unrealistic expectations to a model in which text mining is just another tool in the humanist’s chest. Like a set of Allen wrenches: you may not need them very often, but they only cost a couple of bucks so you don’t mind the expense.

Leave a Reply

Your email address will not be published / Required fields are marked *

Name*

Email*

Website

Post Comment

- Notify me of follow-up comments by email.
- Notify me of new posts by email.
One of the most basic applications of text mining is simply counting words. I began by stripping out punctuation (in order to avoid differentiating mend and mend. as two separate words), put every word into lowercase, and then ignored a list of stop words (the, and, for, etc.). By writing a program to count occurrences of the 500 most common words, I could get a general (and more quantitative) sense for what general topics Martha Ballard wrote about in her diary. Unsurprisingly, her vocabulary usage followed a standard path of exponential decay: like most people, she utilized a relatively small number of words with extreme frequency. For example, the most common word (mr) occurred 10,050 times, while her 500th most common word (relief) occurred 67 times:

Because each word has information attached to it – specifically what date it was written – we can look at long-term patterns for a particular word’s usage. However, looking at only raw word frequencies can be problematic. For example, if Ballard wrote the word yarn twice as often in 1801 as 1791, it could mean that she was doing a lot more knitting in her old age. But it could also mean that she was writing a lot more words in her diary overall. In order to address this issue,
for any word I was examining I made sure to normalize its frequency – first by dividing it by the total word count for that year, then by dividing it by the average usage of the word over the entire diary. This allowed me to visualize how a word’s relative frequency changed from year to year.

In order to visualize the information, I settled on trying out sparklines: “small, intense, simple datawords” advocated by infographics guru Edward Tufte and meant to give a quick, somewhat qualitative snapshot of information. To test my method, I used a theme that Laurel Ulrich describes in *A Midwife’s Tale*: land surveying. In particular, during the late 1790s Martha’s husband Ephraim became heavily involved in surveying property. In the raw word count list, both *survey* and *surveying* appear in the top 500 words, so I combined the two and looked at how Martha’s use of them in her diary changed over the years (1785-1812):

```
      survey(ing)
```

Looking at the sparkline, we get a visual sense for when surveying played a larger role in Martha’s diary – around the middle third, or roughly 1795-1805, which corresponds relatively well to Ulrich’s description of Ephraim’s surveying adventures. As a basis for comparison, the word *clear* appeared with numbing regularity (almost always in reference to the weather):

```
      clear
```

Using word frequencies and sparklines, I could investigate and visualize other themes in the diary as well.

**Religion**

Out of the 500 most frequent words in the diary, only three of them relate directly to religion: *meeting* (#28), *worship* (#143), and *god* (#220).
Meeting, which was used largely in a religious context (going to a church meeting), but also in a socio-political context (attending town meetings), had a relatively consistent rate of use, although it trended slightly upwards over time. Worship (which Martha largely used in the sense of “went to publick worship”), meanwhile, was more erratic and trended slightly downwards. Finally, and perhaps most interestingly, was Martha’s use of the word god. Almost non-existent in the first third of her diary, it then occurred much more frequently, but also more erratically over the final two-thirds of the diary. Not only was it a relatively infrequent word overall (flax, horse, and apples occur more often), but its usage pattern suggests that Martha Ballard did not directly invoke a higher power on a personal level with any kind of regularity (at least in her diary). Instead, she was much more comfortable referring to the more socially and community-based activity of attending a religious service. While a qualitative close reading of the text would give a richer impression of Martha’s spirituality, a quantitative approach demonstrates how little “real estate” she dedicates to religious themes in her diary.

Death

death

dead

funeral
Most of the words related to death show an erratic pattern. There are peaks and valleys across the years without much correlation between the different words, and the only word that appears with any kind of consistency is interd (interred). In this case, word frequency and sparklines are relatively weak as an analytical tool. They don’t speak to any kind of coherent pattern, and at most they vaguely point towards additional questions for study – what causes the various extreme peaks in usage? Is there a common context with which Martha uses each of the words? Why was interd so much flatter than the others?

**Family**

In this final section, I’ll offer up a small taste of how analyzing word frequency can reveal interpersonal relationships. I used the particular example of Dolly (Martha’s youngest daughter):

The sparkline does a phenomenal job of driving home a drastic change in how Martha refers to her daughter. In a matter of a year or two in the mid 1790s, she goes from writing about Dolly frequently to almost never mentioning her. Why? Some quick detective work (or reading page 145 in *A Midwife’s Tale*) shows that the plummet coincides almost perfectly with Dolly’s marriage to a man named Barnabas Lambart in 1795. But why on earth would Martha go from mentioning Dolly all the time in her diary to going entire years without writing her name? Did Martha disapprove of her daughter’s marriage? Was it a shotgun wedding?

The answer, while not so scandalous, is an interesting one nonetheless that text
analysis and visualization helps to elucidate. In short, Martha still writes about her daughter after 1795, but instead of referring to her as Dolly, she begins to refer to her as Dagt Lambd (Daughter Lambert). This is a fascinating shift, and one whose full significance might get lost by a traditional reading. A human poring over these detailed entries might get a vague impression that Martha has started calling her daughter something different, but the sparkline above drives home just how abrupt and dramatic that transformation really was. Martha, by and large, stopped calling her youngest daughter by her first name and instead adopted the new husband’s proper name. Such a vivid symbolic shift opens up a window into an array of broader issues, including marriage patterns, familial relationships, and gender dynamics.

**Conclusions**

Counting word frequency is a somewhat blunt instrument that, if used carefully, can certainly yield meaningful results. In particular, utilizing sparklines to visualize individual word frequencies offers up two advantages for historical inquiry:

1. Coherently display general trends
2. Reveal outliers and anomalies

First, sparklines are a great way to get a quick impression of how a word’s use changes over time. For example, we can see above that the frequency of the word expired steadily increases throughout the diary. While this can often simply reiterate suspected trends, it can ground these hunches in refreshingly hard data. By the end of the diary, a reader might have a general sense for how certain themes appear, but a text analysis can visualize meaningful patterns and augment a close reading of the text.
Second, sparklines can vividly reveal outliers. In the course of reading hundreds of thousands of words over the course of nearly 10,000 entries, it’s quite easy to lose sight of the forest for the trees (to use a tired metaphor). Visualizing word frequencies allows historians to gain a broader perspective on a piece of the text, and they also act as signposts pointing the viewer towards a specific area for further investigation (such the red-flag-raising rupture in how frequently Dolly appears). Relatively basic word frequency by itself (such as what I’ve done here) does not necessarily explain anomalies, but it can do an impressive job of highlighting important ones.

#A Midwife's Tale  #Laurel Ulrich  #Martha Ballard  #Programming  #Python  #Text analysis

Comments

Agnieszka Kielkiewicz-Janowiak - December 7, 2009 @ 5:19 am
Cameron,
I am really impressed by your work and dedication to getting an in-depth understanding Martha’s story. I came across this wonderful text resource when I was researching New England women’s private writings over 10 years ago. Working mostly from Poland, I found it invaluable to be able to access this manuscript online (while I had to retrieve others from microfilm in a UMass library when I was there a short time). My special interest, as a sociolinguist, was language patterns and the occurrence of structures such as do-less negation, periphrastic do in declaratives, conjunctions (e.g. the almost obsolete “ere”), be and have with mutative intransitives, modal verbs, pronouns, etc. I was not just the word count I was after, but most importantly the context (broad and narrow). How I wish I had access to your expertise then! I published a book on the language of a few New England women in 2002. Thank you for renewing my interest in the text of Martha’s diary. And congratulations on your results!

Cameron Blevins - December 7, 2009 @ 9:27 am
Agnieszka,
Thanks for the kind note. Your research sounds fascinating! I have close to zero background in sociolinguistics, but I can imagine the applications are pretty wide-ranging. I think one of the
fundamental challenges to my approach here is one of context – figuring out ambiguities or references without a human reader looking at each instance is tough. I’m hoping to devote some more time to exploring the diary in the next month or so, and will post whatever else I find here.

Thanks again!

-Cameron

Leave a Reply

Your email address will not be published / Required fields are marked *

Post Comment

Notify me of follow-up comments by email.

Notify me of new posts by email.
In *A Midwife’s Tale*, Laurel Ulrich describes the challenge of analyzing Martha Ballard’s exhaustive diary, which records daily entries over the course of 27 years: “The problem is not that the diary is trivial but that it introduces more stories than can be easily recovered and absorbed.” (25) This fundamental challenge is the one I’ve tried to tackle by analyzing Ballard’s diary using text mining. There are advantages and disadvantages to such an approach – computers are very good at counting the instances of the word “God,” for instance, but less effective at recognizing that “the Author of all my Mercies” should be counted as well. The question remains, how does a reader (computer or human) recognize and conceptualize the recurrent themes that run through nearly 10,000 entries?

One answer lies in topic modeling, a method of computational linguistics that attempts to find words that frequently appear together within a text and then group them into clusters. I was introduced to topic modeling through a separate collaborative project that I’ve been working on under the direction of Matthew Jockers (who also recently topic-modeled posts from *Day in the Life of Digital Humanities 2010*). Matt, ever-generous and enthusiastic, helped me to install MALLET (Machine Learning for Language Toolkit), developed by Andrew McCallum at UMass as “a Java-based package for statistical natural language
processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text.” MALLET allows you to feed in a series of text files, which the machine will then process and generate a user-specified number of word clusters it thinks are related topics. I don’t pretend to have a firm grasp on the inner statistical/computational plumbing of how MALLET produces these topics, but in the case of Martha Ballard’s diary, it worked. Beautifully.

With some tinkering, MALLET generated a list of thirty topics comprised of twenty words each, which I then labeled with a descriptive title. Below is a quick sample of what the program “thinks” are some of the topics in the diary:

- **MIDWIFERY**: birth deld safe morn receivd calld left cleverly pm labour fine reward arivd infant expected recd shee born patient
- **CHURCH**: meeting attended afternoon reverend worship foren mr famely performd vers attend public supper st service lecture discoarst administred supt
- **DEATH**: day yesterday informd morn years death ye hear expired expird weak dead las past heard days drowned departed evinn
- **GARDENING**: gardin sett worked clear beens corn warm planted matters cucumbers gatherd potatoes plants ou sowd door squash wed seeds
- **SHOPPING**: lb made brot bot tea butter sugar carried oz chees pork candles wheat store pr beef spirit churnd flower
- **ILLNESS**: unwell mr sick gave dr rainy easier care head neighbor feet relief made throat poorly takeing medisin ts stomach

When I first ran the topic modeler, I was floored. A human being would intuitively lump words like *attended*, *reverend*, and *worship* together based on their meanings. But MALLET is completely unconcerned with the meaning of a
word (which is fortunate, given the difficulty of teaching a computer that, in this
text, discoarst actually means discorsed). Instead, the program is only
concerned with how the words are used in the text, and specifically what words
tend to be used similarly.

Besides a remarkably impressive ability to recognize cohesive topics, MALLET
also allows us to track those topics across the text. With help from Matt and using
the statistical package R, I generated a matrix with each row as a separate diary
entry, each column as a separate topic, and each cell as a “score” signaling the
relative presence of that topic. For instance, on November 28, 1795, Ballard
attended the delivery of Timothy Page’s wife. Consequently, MALLET’s score for
the MIDWIFERY topic jumps up significantly on that day. In essence, topic
modeling accurately recognized, in a mere 55 words (many abbreviated into a
jumbled shorthand), the dominant theme of that entry:

“Clear and pleasant. I am a\textsuperscript{t} mr Pages, had another fitt of y\textsuperscript{e} Cramp,
not So Severe as that y\textsuperscript{e} night past\textsuperscript{t}. mrss Pages illness Came on a\textsuperscript{t}
Evng and Shee was Deliverd a\textsuperscript{t} 11\textsuperscript{h} of a Son which waid 12 lb. I
tarried all night She was Some faint a little while after Delivery.”

The power of topic modeling really emerges when we examine thematic trends
across the entire diary. As a simple barometer of its effectiveness, I used one of
the generated topics that I labeled COLD WEATHER, which included words such
as cold, windy, chilly, snowy, and air. When its entry scores are aggregated into
months of the year, it shows exactly what one would expect over the course of a
typical year:
As a barometer, this made me a lot more confident in MALLET’s accuracy. From there, I looked at other topics. Two topics seemed to deal largely with **HOUSEWORK**:

1. house work clear knit wk home wool removd washing kinds pickt helping banking chips taxes picking cleaning pikt pails

2. home clear washt baked cloaths helped washing wash girls pies cleand things room bak kitchen ironed apple seller scolt

When charted over the course of the diary, these two topics trace how frequently Ballard mentions these kinds of daily tasks:

Both topics moved in tandem, with a high correlation coefficient of 0.83, and both steadily increased as she grew older (excepting a curious divergence in the last several years of the diary). This is somewhat counter-intuitive, as one would think the household responsibilities for an aging grandmother with a large family would *decrease* over time. Yet this pattern bolsters the argument made by Ulrich in *A Midwife’s Tale*, in which she points out that the first half of the diary was “written when her family’s productive power was at its height.” (285) As her children married and moved into different households, and her own husband experienced mounting legal and financial troubles, her daily burdens around the house increased. Topic modeling allows us to quantify and visualize this pattern, a pattern not immediately visible to a human reader.

Even more significantly, topic modeling allows us a glimpse not only into Martha’s tangible world (such as weather or housework topics), but also into her
abstract world. One topic in particular leaped out at me:

feel husband unwel warm feeble felt god great fatagud fatagued thro life time
year dear rose famely bu good

The most descriptive label I could assign this topic would be EMOTION – a tricky and elusive concept for humans to analyze, much less computers. Yet MALLET did a largely impressive job in identifying when Ballard was discussing her emotional state. How does this topic appear over the course of the diary?

Emotion

Like the housework topic, there is a broad increase over time. In this chart, the sharp changes are quite revealing. In particular, we see Martha more than double her use of EMOTION words between 1803 and 1804. What exactly was going on in her life at this time? Quite a bit. Her husband was imprisoned for debt and her son was indicted by a grand jury for fraud, causing a cascade effect on Martha’s own life – all of which Ulrich describes as “the family tumults of 1804-1805.” (285) Little wonder that Ballard increasingly invoked “God” or felt “fatagued” during this period.

I am absolutely intrigued by the potential for topic modeling in historic source material. In many ways, it seems that Martha Ballard’s diary is ideally suited for this kind of analysis. Short, content-driven entries that usually touch upon a limited number of topics appear to produce remarkably cohesive and accurate topics. In some cases (especially in the case of the EMOTION topic), MALLET did a better job of grouping words than a human reader. But the biggest advantage lies in its ability to extract unseen patterns in word usage. For instance, I would not have thought that the words “informed” or “hear” would cluster so strongly
into the DEATH topic. But they do, and not only that, they do so more strongly within that topic than the words dead, expired, or departed. This speaks volumes about the spread of information – in Martha Ballard’s diary, death is largely written about in the context of news being disseminated through face-to-face interactions. When used in conjunction with traditional close reading of the diary and other forms of text mining (for instance, charting Ballard’s social network), topic modeling offers a new and valuable way of interpreting the source material.

I’ll end my post with a topic near and dear to Martha Ballard’s heart: her garden. To a greater degree than any other topic, GARDENING words boast incredible thematic cohesion (gardin sett worked clear beens corn warm planted matters cucumbers gatherd potatoes plants ou sowd door squash wed seeds) and over the course of the diary’s average year they also beautifully depict the fingerprint of Maine’s seasonal cycles:

Note: this post is part of an ongoing series detailing my work on text mining Martha Ballard’s diary.

Comments

Jason Boyd - April 1, 2010 @ 8:30 am
Fascinating. I work for Records of Early Drama (REED) — we publish collections of pre-1642 documents,
and I was very interested to see how effective MALLET was in dealing with a linguistically complex text like Martha Ballard's diary. Was the diary text you used marked up at all? Or was it a plain text file?

Another question: although MALLET is unconcerned with word meanings, instead focusing on patterns of word usage, how does it overcome the problem of text that predates standardized spelling, punctuation, and grammar? Could it handle texts that were authored by numerous people over time, each of whom had their particular idiosyncrasies?

- Cameron Blevins - April 1, 2010 @ 8:56 am
  
  Jason,

  All good questions.

  1. The diary was not marked up at all. It was processed using Python into a basic list/array (with date, day of the week, text from the entry, etc.). From there I just exported the main text from each entry into ~10,000 separate .txt files, which MALLET could then treat as separate documents. Tracking them over time was a matter of naming the txt files by their date, such as 18070225.txt (2/25/1807).

  2. I was pleasantly shocked at how well MALLET handled the messiness of Ballard’s shorthand style of writing. I think there were a few factors that contributed to this:

     – Stretched over 10,000 entries and 27 entries, the vagaries of different spellings tend to smooth out. Big data can overcome a lot of problems.

     – In a way, MALLET has an advantage in overcoming spelling variances. Provided the variances are somewhat consistent, it doesn’t care whether the word is “delivd” or “delivered,” all it knows is that particular string of characters tends to appear alongside “birth” words.

  3. MALLET can handle many different texts/authors – in fact, that’s precisely what Matt Jockers has been doing. This has particular potential for clustering different authors together. The downside is that you tend to get “topics” that form based on unique words in an author’s vocabulary. If you were to feed it contemporary British fiction, for instance, you’d probably get a topic of words like “Potter” “Hogwarts” and “Quidditch” – not particularly useful for analyzing trends your entire corpus. It all probably depends on just how variant the particular idiosyncrasies are from author to author.

  Hope this helps.

  -Cameron

- erik steiner - April 2, 2010 @ 5:11 pm

  Cameron,
This is awesome. I’m very intrigued by the possibility that this approach can be used to accurately model geographically varying patterns – such as climate. It would be very cool to track down actual weather data and correlate it with her references – or at least overlay it on your graphs. In theory, you could also reverse-geocode diaries (or newspapers) to determine based on their content where they were from. Since you know the locations of newspapers, it might be an interesting way to test this idea.

Also, I’m wondering about MALLET and the topics it defines – does it tell you how related two topics are to one another, and can you see this change over time? It would be interesting, for example, to see if Martha becomes has less EMOTION around DEATH as she gets older.

Great work. I look forward to more cool stuff from this.

-erik

- Cameron Blevins - April 2, 2010 @ 11:32 pm
Erik,

Thanks for the feedback. I really like the idea of reverse-geocoding, especially if you had a known-location training corpus for the program to work with.

MALLET doesn’t necessarily tell you how related two topics are to one another (at least I think, like I said I’m pretty shaky on how it works from a technical standpoint). But since I have all the temporal data associated with their “scores” for each entry, it’s easy to do. I’ve actually played around a bit and set up a correlation matrix to see which topics move in tandem or apart. Mixed results so far, but it was interesting to see one topic that I was having trouble identifying move almost exactly opposite (coefficient of -0.9) with the COLD WEATHER topic over the course of a typical year. I still don’t really know what the topic is (weakly associated with rainy weather?), but whatever it is seems to appear in the warmer months:

cloudy afternoon rain home foren fore flax shower tn showers thunder af aft combd heavy turns misty dress pulld

-Cameron

david blei - April 6, 2010 @ 10:01 am
this is fascinating.

re: geocoding. i work a lot on developing topic modeling tools. we recently developed a topic model that might account for location, by associating each document with a location and encoding which locations are adjacent to each other. (it’s not exactly geocoding, but it kind of gets you there…)

we wrote about it in this paper, which is forthcoming from the annals of applied statistics:
the code is implemented in the “lda” R package. (in fact, this package lets you fit a number of types of topic models.)

best
dave

○ Cameron Blevins - April 6, 2010 @ 9:11 pm
  Dave,

  Thanks for the comment! Although most of your paper was a bit over my non-quanty humanities head, it was interesting to see the intersection of topic modeling and geographic analysis. I’ll also be sure to check out the LDA package, thanks for the suggestion.

  -Cameron

Lisa - April 17, 2010 @ 12:50 pm
Hi Cameron:

Thanks to you and Matt for introducing MALLET — I found your analysis of the product very interesting. I’m curious to know whether MALLET would also work for languages/scripts other than English? Say, Chinese?

By the way, the Archivist of the United States’ most recent blog entry on the Library of Congress’ acquisition of Twitter. He references Martha Ballard’s Diary.
http://blogs.archives.gov/aotus/?p=172

Thanks again for a fascinating read.

○ Cameron Blevins - April 19, 2010 @ 9:31 am
  Lisa,

  I’d be interested to see if it works on other languages, could have some fascinating potential there.

  Thanks for the link to the Archivist post, that was an interesting analogy between Ballard’s diary entries as tweets.

  -Cameron

David Mimno - May 18, 2010 @ 7:20 am
Hi Cameron,

Thanks for using our MALLET topic modeling tools! This is exactly the type of research that got me
interested in statistical text mining.

Regarding irregular spellings: I've run this code on large early English collections, and it tends to find "clusters" of spelling variations, rather than smoothing over all variation and all time. For example you usually don't get 17th century spellings mixed with fully modern orthography. For a single-author corpus like this diary, it should work very well even with substantial variation.

On multiple languages: MALLET will support any language, although you may need to do some extra work creating "stoplists" of very common words and tokenizing the text (for example using the Stanford Chinese word segmenter). If you have documents aligned across multiple languages (such as wikipedia articles), MALLET also supports "polylingual" topic modeling: use the option –language-inputs instead of –input to learn topics in many languages simultaneously.

-David

- Cameron Blevins - May 18, 2010 @ 8:03 pm

David,

And thanks to you all at UMASS for building and maintaining such a great tool.

I'm interested to hear about your experience with different corpora, especially ones that encompass several centuries. Do you think it's finding clusters of spelling variations because of the actual spelling patterns themselves, or their placement in the text? I think one reason MALLET seems to work so well on this is the fact that it's a single author, but I haven't had much experience with larger (and broader, or polylingual) corpora.

Please send along my appreciation to the rest of the MALLET team.

-Cameron

- Steven - September 8, 2010 @ 11:18 pm

Hey,

One question anyone used MALLET on Social Media data specifically on blogs?

FM

- Ron - October 4, 2011 @ 5:08 am

Any particular reason to use the one “I” spelling of “Topic Modeling”?

- Datafiend - April 14, 2014 @ 4:31 am

Reblogged this on Austen, Morgan and Me and commented:

Detailed blog post exploring the use of MALLET to topic model a diary.
Leave a Reply

Your email address will not be published / Required fields are marked *

Name*
Email*
Website

Post Comment

☐ Notify me of follow-up comments by email.
☐ Notify me of new posts by email.