DHSI

DIGITAL HUMANITIES SUMMER INSTITUTE

Designing Digital Publications

Mary Borgo Ton and Dan Tracy

ETCL
ELECTRONIC TEXTUAL CULTURES LAB

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada

Welcome to DHSI 2023!

Thank you for joining the DHSI community!

In this coursepack, you will find essential workshop materials prefaced by some useful general information about DHSI 2023.

Given our community's focus on things computational, it will be a surprise to no one that we might expect additional information and materials online for some of the workshops—which will be made available to you where applicable—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 any information that's not in this coursepack.

Please also note that materials in DHSI's online workshop folders could be updated at any point. We recommend checking back on any DHSI online workshop folder(s) that have been shared with you in case additional materials are added as DHSI approaches and takes place.

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.

We hope you enjoy your time with us!

DHSI
DIGITAL HUMANITIES SUMMER INSTITUTE

ETCL
ELECTRONIC TEXTUAL CULTURES LAB

Statement of Ethics & Inclusion

Please review the DHSI Statement of Ethics & Inclusion available here: **https://dhsi.org/statement-of-ethics-inclusion/**

DHSI is dedicated to offering a safe, respectful, friendly, and collegial environment for the benefit of everyone who attends and for the advancement of the interests that bring us together. There is no place at DHSI for harassment or intimidation of any kind.

By registering for DHSI, you have agreed to comply with these commitments.

Virtual Sessions

Your registration in DHSI 2023 also includes access to the virtual **institute lecture** sessions. Access details for these talks will be shared as DHSI approaches.

Due to the high volume of attendees, please ensure your DHSI registration name or DHSI preferred name and your Zoom name match so that we know to let you into the virtual sessions.

DHSI Materials

DHSI materials (ex. videos, documents, etc.) are intended for registrant use only. By registering, you have agreed that you will not circulate any DHSI content. If someone asks you for the materials, please invite them to complete the registration form to request access or contact us at **institut@uvic.ca**.

Auditor and participant registration

If you registered to **audit** any workshops, note that auditor involvement is intended to be fully self-directed without active participation in the workshop. The auditor option offers more flexibility regarding pace and time with the workshop content. Your registration as an auditor will include access to some asynchronous workshop materials only and does not include access to live workshop sessions and/or individual/group instruction or consultation. Please direct any questions about DHSI workshop auditing to **institut@uvic.ca**.

If you registered as a **participant** in any workshops, your registration includes access to asynchronous content + active participation in live workshop session(s). The workshop instructor(s) will contact you about the date(s), time(s), and platform(s) of the live workshop session(s).

If you are unsure whether you registered as an auditor or participant, please check your registration confirmation email. Further questions can be directed to **institut@uvic.ca**.

Schedule

The at-a-glance schedule of DHSI 2023 courses, workshops, institute lectures and aligned conferences & events can be found here:
https://dhsi.org/timetable/

All times are listed in North American **Pacific Time Zone**.

For those who registered as participants in any workshops, live sessions for online workshops are not currently listed on the above-referenced schedule. **Instructors will be in touch with registered participants directly about the exact date(s) and time(s) of their live workshop session(s).**

Acknowledgements

We would like to thank our partners and sponsors (including the Social Sciences and Humanities Research Council), workshop instructors, aligned conference & event organizers, institute lecturers, local facilitators, and beyond for making this possible.

Further information

General DHSI 2023 information: **https://dhsi.org/program/**

Full course listings (in-person): **https://dhsi.org/on-campus-courses/**

Full workshop listings (online): **https://dhsi.org/online-workshops/**

Aligned conferences & events (in-person): **https://dhsi.org/on-campus-aligned-conferences-events/**

Aligned conferences & events (online): **https://dhsi.org/online-aligned-conferences-events/**

Institute lectures: **https://dhsi.org/institute-lectures/**

Frequently asked questions: **https://dhsi.org/faq/**

Any questions not addressed in the above pages? Please email us at **institut@uvic.ca**!

Designing Digital Publications

DHSI 2023
5th-9th of June at the University of Victoria

Mary Borgo Ton and Dan Tracy

This workshop will focus on strategies for designing, building, and publishing long-form scholarship in fully digital formats. The synchronous portion of this workshop will focus on discussion and exploration of design elements that will be important to your project. As we consider commonly-used platforms like Pressbooks, Omeka, and Scalar, we will cover strategies for working with multimedia, structuring prose content, and designing digital reading experiences. Participants will have ample opportunities to reflect on their own research, professional goals, and audiences as they make choices about the content and design of their own projects. Following the synchronous session, participants will have access to sandbox sites and resources to guide hands-on experimentation with Pressbooks, Omeka, and Scalar. This workshop is ideal for graduate students who are contemplating a born-digital dissertation, scholars who are working heavily with multimedia, and those who are curious to explore alternatives to print-based scholarship.

Mary Borgo Ton is the Digital Humanities Librarian and Assistant Professor at the University of Illinois Urbana-Champaign. She received her Ph.D. in British Literature with a concentration in Victorian literature and a Graduate Certificate in Digital Arts and Humanities from Indiana University and has contributed to digital collections of materials from the global south, including *Livingstone Online*, *One More Voice* , and *Archivo Mesoamericano*. As the former Digital Publishing Specialist, she supported authors and editors in all stages of the publication process as they created long-form digital scholarly works in Pressbooks, Omeka, and Scalar the Illinois Open Publishing Network (IOPN).

Dan Tracy is Associate Professor and Head, Scholarly Communication and Publishing, at the University of Illinois Library. His responsibilities include directing the Illinois Open Publishing Network (IOPN), a library-based scholarly publisher of journals and long-form works, including DH publications. He holds a PhD in English with a concentration in 20th Century American Literature, as well as an MS-LIS, from Illinois. His research in recent years has focused on user experience of digital publications and publishing platforms. Currently he has a library-based grant to develop approaches to digital editions using IOPN infrastructure, and in the course of developing these approaches is the editor of *Gentlemen Prefer Blondes: A Critical Edition*, which is available in beta. Along with other IOPN colleagues, he is a member of the grant team for the Mellon-based AFRO Publishing Without Walls 2, which seeks to expand capacity for digital publishing in Black Studies, in partnership with the Illinois Department of African American Studies and North Carolina Central University.

Welcome

Thank you for joining us! Through this workshop, you will learn how to:

- Create a project charter for a digital publishing project
- Develop writing workflows for digital-born publications
- Identify and evaluate digital publishing tools
- Record metadata for media items
- Design accessible web-based content
- Assess hosting options
- Consider preservation and storage options
- Engage in ethical collaboration practices

Before the Workshop

Please fill out this form to help us get to know you, your research interests, and your publishing project ideas: https://forms.illinois.edu/sec/404951639

Schedule

Monday	Morning 10:30-12	1 Introductions / Overview *Exercise: Evaluate model projects*
	Afternoon 1-2:30	2 Scoping your project *Exercise: Draft a basic project charter*
Tuesday	Morning 9-12	3 Exploring writing workflows for digital publishing *Exercise: Map your ideas* *Exercise: Setting up your websites*
	Afternoon 1-4	4 Working with media: metadata, copyright, fair-use, and accessibility *Exercise: Add media to your project*
Wednesday	Morning 9-12	5 Structuring the reading experience *Exercise: Add prose to your project* *Exercise: Map multiple paths through your content*
	Afternoon 1-4	6 Exploring advanced layout *Exercise: Universal design principles* *Exercise: Modify HTML/CSS*
Thursday	Morning 9-12	7 Planning for the long term • Selecting a publisher/hosting service • Developing a preservation plan • Evaluating third-party content (Tableu, ArcGIS) Considerations for special contexts. Choose your own adventure! • Incorporating multimodal writing into your pedagogy • Proposing and advising digital dissertations and theses • Digital publishing and the tenure dossier
	Afternoon 1-4	Revising your project charter + developing your prototype • Ad hoc consultations
Friday	Morning 9-10:15	Conclusions + next steps

Activity: Comparing Publications

Choose one site from each of the following categories.

Pressbooks	Omeka	Scalar
• *Claude Monet: The Water-Lilies and other writings on art* • *A Person-Centered Guide to Demystifying Technology* • *Instruction in Libraries and Information Centers* • *Elementary Arabic I*	• *Sugar Production Stories for Children and the History of Slavery* • *Starkiller to Skywalker: How* Star Wars *Evloved from Script to Screen* • *History Harvest at the University of Illinois* • *Wearing Gay History*	• *Lost in the City: An Exploration of Edward P. Jones's Short Fiction* • *Love and Suspense in Paris Noir: Navigating the Seamy World of Jake Lamar's Rendezvous Eighteenth* • *Et Al.: New Voices in Arts Management* • *Why Busing Failed: Race, Media, and the National Resistance to School Desegregation* • *Bodies and Structures* • *Claude McKay's Early Poetry (1911-1922): A Digital Collection*

Pressbooks

Which site are you reviewing?

What is the overall goal or argument of the site?

Who is the primary audience for this site?

What is the relationship between prose and media (including proportion of each, page layout, relative significance)? How are images, video, and audio materials incorporated?

How are readers expected to engage with the material—read like a book, explore through searching? And what kinds of navigational features does the site include to support that mode of reading? (drop down outline, navigation bar, search)

How are citations and scholarly sources displayed?

What do you find effective about the presentation of prose and media on this site?

What would you change?

Omeka

Which site are you reviewing?

What is the overall goal or argument of the site?

Who is the primary audience for this site?

What is the relationship between prose and media (including proportion of each, page layout, relative significance)? How are images, video, and audio materials incorporated?

How are readers expected to engage with the material—read like a book, explore through searching? And what kinds of navigational features does the site include to support that mode of reading? (drop down outline, navigation bar, search)

How are citations and scholarly sources displayed?

What do you find effective about the presentation of prose and media on this site?

What would you change?

Scalar

Which site are you reviewing?

What is the overall goal or argument of the site?

Who is the primary audience for this site?

What is the relationship between prose and media (including proportion of each, page layout, relative significance)? How are images, video, and audio materials incorporated?

How are readers expected to engage with the material—read like a book, explore through searching? And what kinds of navigational features does the site include to support that mode of reading? (drop down outline, navigation bar, search)

How are citations and scholarly sources displayed?

What do you find effective about the presentation of prose and media on this site?

What would you change?

All Sites

What did these three sites have in common?

What were features that were unique?

Which aspect of the website seems the most technically challenging or complex?

Which one do you think is the best model for your project?

Platform Overview

Pressbooks

Pressbooks is ideal for monograph-length works that want to mimic the reading experience of print but with multimedia and interactive content. It is often the go-to choice for Open Educational Resources and textbooks. Key features:

- Editorial interface is built on WordPress
- Table of contents and search function to support browsing
- Prose with embedded images, video, and interactive H5P content.
- Import content from .dox file
- Export content to PDF

Platform Guide: https://guide.pressbooks.com/

Omeka

Omeka Classic and Omeka S create curated collections of images, video, and audio. It is often used by libraries to create exhibits of special collections material, by History Harvest for local history, and by authors looking to create digital companions to print publications. Key features:

- Advanced search functionality
- Expanded capacity for detailed metadata
- Encourages reader to explore material
- Supporting prose to contextualize objects

Omeka Classic Platform Guide: https://omeka.org/classic/docs/
Omeka S Platform Guide: https://omeka.org/s/docs/user-manual/

Scalar

Scalar is ideal for publications that want to offer a “choose-your-own-adventure" reading experience. It enables authors to create multiple sequences of pages and alternate outlines of the content. Scalar tends to be used for collections of essays with multiple thematic ties and for media-centric projects. Key features:

- Media annotation features
- Multiple paths through the material
- Built-in visualization tools that expose connections between pages, media, and other content

Platform Guide: https://scalar.usc.edu/works/guide2/index

Metadata Scoping Exercise

In the context of digital libraries, metadata supports core search functionality as well as providing essential descriptive information about resources. In a digital project implemented with a specific scholarly perspective for a specific audience, these issues may all still apply, but the description carries additional weight because you may select metadata that is tailored to a narrower audience than the typical digital library. You may also have metadata that supports technical functionality beyond search.

As an example, if I have digital copies of films or film stills in my media items, the original aspect ratio of the film might be considered essential information for one audience, but completely unnecessary for another.

Consider

1. What types of media are included as part of your project? You may want to distinguish not just video vs image, but types of image, such as film stills vs portraits.

2. For each type of media above, what are essential metadata fields your digital project's audience would expect to have for every item as basic scholarly information or to adequately understand your argument or the relevance of the items?

3. Looking at all of the types of media listed under #1, what do you think is essential metadata that your audience might not immediately care about but you have a legal or ethical responsibility to provide regardless of content type?

4. Of the types of information above, what do you have reliably? What would be hard to come up with?

5. What are some metadata fields that someone might want but are out of scope for your project?

Accessibility Resources

Below are resources for creating accessible content.

Accessibility Checker

https://www.accessibilitychecker.org/

This free, web-based tool will review the structure and visual layout of your website. The report will include information on which parts of the site do not follow best practices and offers suggestions on how to make them more accessible.

AInspector for Firefox

https://addons.mozilla.org/en-US/firefox/addon/ainspector-wcag/

This Firefox add-on inspects web-pages for issues related to WCAG 2.0 level A and AA requirements.

Reasonable Colors

https://reasonable.work/colors/

This website is a free, open-soutce color system for building accessible color palettes.

Coblis—Color Blindness Simulator

https://www.color-blindness.com/coblis-color-blindness-simulator/

Upload images to this page simulate how various forms of colorblindness can affect someone's perception of an image.

Amy Cesal's Data Visualization Guidelines for the CFPB

https://www.amycesal.com/portfolio/#/cfpb-design-manual-data-visualization/

This guide provides a short summary of best practices when developing grafts and charts.

Hosting Options

A hosting service stores your website and makes it available to readers (or "clients") via the internet. Because it takes server space and electricity to provide access to your website, hosting services often charge the person who makes the site a fee. Sometimes the cost is offset by university funding or through the addition of advertisements.

Questions to consider:

- What are my short-term, long-term goals for this site?
- What is my budget?
- Will I be able to export my site to a different host if I need to?

Hosting	Advantages	Disadvantages
Sandbox hosted by a university • USC Scalar • Omeka.net	• Free • No set up required • Ideal for proof-of-concept or experimental sites	• Limited storage capacity • Limited features • No guarantee that the site will remain available
University hosting services	• Usually free to graduate students and faculty • May support set up of WordPress or Omeka sites	• Limited choice of platforms • Limited ability to customize site • Site may disappear if no longer affiliated with the university
Reclaim Hosting or other private hosting service	• One-click installation of Wordpress, Omeka, Scalar and other content management systems • Automatically updates platforms	• Annual fee • Limited tech support for customized websites

Preservation

Creating an archived version of your site will preserve some, if not all, of the website in the event that it is no longer live.

Questions to consider:

- Who is my perceived audience for the archived copy?
- Which features of the site do I want to represent in the archived copy? Is the exact look and feel part of the argument, or am I primarily preserving text and images that can be represented in other ebook formats without losing scholarly content?
- How do I want to be able to access the archived copy?

Saving Ukrainian Cultural Heritage Online (SUCHO) provides a brief overview of commonly used web archiving tools, including Browsertrix, Internet Archive's Wayback Machine, and Manual WebRecorder, as well as links to tutorials for each platform.

Glossary

Here are commonly used terms in the world of digital publishing.

Accessibility: a set of design principles that shapes the structure and layout of webpages in order to make content accessible to all people, regardless of disability type or the severity of impairment

Alt-Text: a short description of images and other visual resources that can be read by a screen reader

Application: a program designed to carry out a specific task

Catalog: an official list of publications

Cascading Style Sheets (CSS): a computer language that determines the format, layout, and style of a webpage

Content Management System (CMS): software used to create and modify digital content

Creative Commons Licenses: a means for the creator of a publication to describe the conditions under which others may use, remix, and share their work

Dashboard: an area of the website that is not seen by normal readers of publications, where editors and authors can manage settings and access specific options to upload, create, and arrange content

Direct Object Identifier (DOI): a unique string of numbers, letters, and symbols that identifies serves as a permanent link to articles, documents, or websites

Domain name: a string of letters, numbers, and symbols, that identifies a website

Embedding: a method of adding content hosted on a different site (i.e. third-party content) to a web page using a link and html

Hosting service: an ecosystem of resources that stores your website and makes it available to readers (or "clients") via the internet

IFrame: an inline frame is an HTML element that enables you to embed content from one website into another

Institutional repository: an archive for collecting, preserving, and disseminating digital copies of the intellectual output of an institution

International Standard Book Number (ISBN): a unique number used by publishers to identify each separate edition and variation of a publication

JavaScript: a programming language that makes web pages interactive

Metadata: information about images, media items, and other resources that often includes the creator's name, year the work was created, and copyright

Metadata profile: establishes a set of predefined attributes (metadata fields and values) determined to be key for a set of similar digital content. In Omeka, these are called **resource templates**. (Adapted from NDLRN.)

Hypertext Markup Language (HTML): a computer language that describes the structure of a web page (i.e. headings, paragraph breaks, ordered lists)

Open Access: a publication with no financial, legal, or technical barriers to accessing it

Open Source: software which has been made freely available so that its original source code can be redistributed and modified

Page (of a website): a document that contains a hypertext file and related files for images, scripts, and graphics. Often created with HTML, CSS, and JavaScript

Plug-in: software that adds new functions to a host program without altering the host program itself (source: Britannica)

Publishing Platform: an ecosystem of software and related resources used to create and publish web-based content

Server: a computer or another device that makes information available to users (or "clients")

Shortcodes: code that appears within square brackets that are recognized by Pressbooks to perform special functions.

Third-party content: resources, images, text hosted on another website

Universal design: the practice of creating web content and interfaces that are accessible to all individuals, regardless of their age, ability, or technology used to access the internet

Website: a collection of web pages that share a common domain name

WYSIWYG (wizz-ee-wig) editor: abbreviation for "What you see is what you get." A display generated by a content management system that represents how content will appear when published.

Recommended readings

The following articles have been added to the course packet under terms of fair use or under the terms of their Creative Commons licenses.

Brainstorm and Design. (2021). From *The Colored Conventions Exhibit Guide*. Penn State University. https://docs.google.com/document/d/13zU6HRW26iW7yp8_WCeLTbSo8JlfVS5L99ahGnBFvp8/edit

> The Colored Conventions Exhibit Guide contains three preparatory exercises that help content creators to conceptualize and design digital exhibits. The brainstorming process can be adapted for thinking about content in Pressbooks, Omeka, and Scalar.

Clement, T. E., Reside, D., Croxall, B. Flanders, Fraistat, N., Jobes, S., Kirschenbaum, M., Lodato, S., Mandell, L. Marty, P. Miller, D., Nowviskie, B., Olsen, S., Scheinfeldt, T., Seaman, D., Tebeau, M., Unsworth, J., Walter, K. (2011). Collaborators' Bill of Rights. From *Digital Pedagogy in the Humanities* Modern Language Association. https://hcommons.org/deposits/item/hc:31187

> This brief, but influential Bill of Rights lays out best practices for attributing labor in collaborative projects.

Di Pressi, H., Gorman, S., Posner, M., Sasayama, R., Schmitt, T. (2015). A Student Collaborators' Bill of Rights. *HumTech*. UCLA. https://humtech.ucla.edu/news/a-student-collaborators-bill-of-rights

> This statement describes ethical labor practices when building digital humanities projects with students.

Gallon, K. (2016). Making a Case for the Black Digital Humanities. In M. K. Gold & L. F. Klein (Eds.), *Debates in the Digital Humanities* 2016, 42–49. University of Minnesota Press. https://doi.org/10.5749/j.ctt1cn6thb.7

> While digital publishing presents opportunities for recovering voices that have been systematically marginalized, this article exposes the limitations of digital humanities tools as vehicles for critically engaged scholarship. Gallon offers Black Digital Humanities as a means to both recognize the ways that digital tools are imprecated in systems of oppression while also imaging digital forms of scholarship as sites of resistance.

Jessop, M. (2006). Metadata Creation for Digital Humanities Projects. *Society for Imaging Science and Technology* 2006, 84-87. https://image.dig4e.com/IS&T%20Articles/arch2006-s20LDC-Jessop.pdf

> Jessop describes the process of compiling metadata for digital humanities projects, particularly when working across institutional collections and disciplines. Through three case studies, Jessop highlights how metadata choices relate to specialized and generalist audiences for digital projects.

Miller, A. (2019). Digital Project Preservation Plan: a Guide for Preserving Digital Humanities/Scholarship Projects. Version 2.0. https://jewlscholar.mtsu.edu/handle/mtsu/5761

This guide provides a step-by-step place for assessing digital projects for preservation, creating an inventory of project assets, and selecting appropriate repositories and tools to archive the project's content.

Sunderland, S. (2022). Assessing User Experience using Digital Humanities Projects in a Design Classroom. *Visualizing Objects, Places, and Spaces: A Digital Project Handbook.* https://doi.org/10.21428/51bee781.957d585e

Sunderland describes the learning outcomes and structure of a multi-part, multimodal writing assignmentt in an undergraduate classroom. This series of activities increases students' media literacy, develop strategies for accessible site design, and situate digital humanities projects within digital forms of scholarly communication more generally.

Warwick, C. (2012). Studying users in digital humanities. In C. Warwick, M. Terras, & J. Nyhan (Eds.), Digital Humanities in Practice, 1-22. Cambridge University Press. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.305.6694&rep=rep1&type=pdf

Though the digital tools described in this chapter are dated, Warwick provides strategies for studying user needs and how audiences shape digital humanities project design.

Ibid. (2020). Interfaces, ephemera, and identity: a study of the historical presentation of digital humanities resources. *Digital Scholarship in the Humanities, 35*(4), 944-971. https://doi.org/10.1093/llc/fqz081

In this study of interface design, Warwick describes the evolution of site design in order to accommodate evolving technical and accessibility standards, display screens, and access to material.

Web Accessibility: Basics and Guidelines. (2023). Eastern Kentucky University. https://www.eku.edu/brand/wp-content/uploads/sites/30/2023/03/cbm_web_accessibility_r1.pdf

This web accessibility guide includes practical tips on how to make websites more accessible through website structure and navigation, and how to create accessible audio-visual content.

The CCP Exhibit Guide

Design, Develop, and Release a Digital Exhibit

BRAINSTORM AND DESIGN

This section contains a list of three exercises that will prepare you to create, develop, and publish your exhibit on CCP.org. Please follow the exercises in order and read each section carefully. Doing so will prepare you to design and develop a strong, cohesive exhibit for CCP.org.

Exercise 1: Exhibits Review

Browse an example exhibit on the CCP website. Write down answers to the following questions.

1. How is the overall exhibit organized?
   _____
2. How is the first page of the exhibit organized, and how does that compare to the different kinds of pages in the section?
   _____
3. For each kind of page, how are the text and images arranged?
   _____
4. What purpose does the text on the homepage of the exhibit serve?
   _____
5. Who are the intended audiences and how does that shape the voice of the text in the exhibit?
   _____
6. Does the exhibit contain any image galleries, maps, data visualizations, or other interactive elements? If so, can you tell what tools were used to create those elements?
   _____

Exercise 2: Creating Exhibit Pages

Duration: approx 30-45 minutes

Materials:

- Post-It notes or scraps of paper (at least 10/person)
- Sheets of 8.5” x 11” paper (at least 4 sheets/group)
- Color pencils, crayons, or markers (optional, but helpful)

Sit down as a group and complete each step in this exercise. The point is to start on paper so that we can see the whole exhibit come together before we begin to implement the technical aspects. For materials you will need a stack of post-it notes or scraps of paper and a large space to work together.

1. Brainstorm

Begin by coming up with a number of ideas for elements that you might want to include in the pages of your exhibit. At the start, you do not need to figure what all of those elements might be, or how they will need to be arranged on different pages. Instead, each person should begin by working individually to sketch at least ten ideas for elements of the exhibit (better yet, twenty!). Set a timer for ten minutes to allow each person to put ten ideas on ten scraps of paper (or more!). You can note these ideas in short phrases, diagrams, or in rough sketches. Anything goes!

An element can be any mini-ideas, possible concepts, or parts of a page in the exhibit. Here are some possible examples of an element:

- a short narrative for background about a person, place, organization
- a timeline or data visualization graphs
- a map
- a picture/image
- overview of a section with key questions/themes
- possible titles for the exhibit, sections, or pages

2. Share and compare

Once everyone has put at least ten ideas on ten scraps of paper, gather as a group to compare ideas. Please follow these steps:

- Put the sticky notes on the wall or scraps on the table.
- Organize the sticky notes by stacking similar ideas together
- Categorize the ideas by moving them into groups that seem to belong together

3. Explore groups of elements

How might the groups of elements translate into a page in the exhibit? Do the groups seem to fall into any particular ordering? Try moving the sticky notes/scraps around and take a step back. Do the elements seem to fit into any kind of larger organization?

Here, you might draw on your familiarity with some of the cousins of digital exhibits: longform nonfiction, graphic novels, museum physical exhibits, documentaries, art catalogues, or even the kinds of narratives that come from fiction. Does the exhibit seem to have a chronological history with a start, middle, and end? Would it be a novel or a short story collection? Let's borrow the moves of those cousins to encourage readers to explore more than just the first page of an exhibit.

3. Create mock designs for exhibit pages

Once you have started to have a sense of the overall organization of the exhibit, next you should choose three groups of elements to mock up as three different exhibit pages. You might think of the mock pages as the equivalent of an outline. Things change, but it helps to have a blueprint at the start.

As a group, create three mock page designs using a sheet of 8" x 11.5" paper. Fold the paper in half. Use one side to sketch a layout. Draw in or label on the paper where each element might appear on the exhibit page.

Then, on the other half of the paper, create two lists:

- A list of the purposes/functions of the layout
- A list of positives/negatives about the layout

Spend no more than five minutes on each page. (Set a timer!) These pages are supposed to be imperfect, rough sketches. If you are struggling to put things into a layout, consider making one of the pages as the worst possible design. Once completed, discuss them as a group. How do they achieve or fall short of the goals of the exhibit?

4. **Focus on a exhibit page**

Then, choose one page to mock up in more detail. The mock up should also include a section that explains the details of the website layout. What tools, visualizations, maps, or content would be needed? What do we need to learn or research?

5. Ask for feedback

If time, present the detailed page to the rest of your class for feedback. What did each group find in common or in different approaches? Do the mock designs seem to meet the goals of reaching your audiences? What needs to change or be added still?

Exercise 3: Creating Exhibit Storyboards

With the results of Exercise 2, it is now possible to organize the narrative of the exhibits. Before moving to the website, it helps to create a storyboard that will show the general layout and sequence of the exhibit pages.

[INSERT THE EXAMPLE STORYBOARD PAGE HERE]

The exhibits are the primary way that the CCP shares this history with our wider communities. How can the exhibits provide historical background while also creating opportunities for dialogue about the longer history of Black activism? How, we reiterate, can we represent this mass movement while contending with the unequal representation of Black women in the archives of the Colored Conventions? By extension, how might we use the power of narratives to provide context for more abstract representations of Black life in forms of data, visualizations, and so on?

When you are ready to design the storyboard for the exhibit, lay out a sheet of paper for each exhibit page. Add the following details on each page:

1. The title of the exhibit page
2. The purpose or role that this page will serve within the exhibit
3. A sketch of the layout of elements on the page
4. Labels on each elements for who will be responsable for creating the content.

Complete these steps for each intended page of the exhibit. Don't forget to add a sheet for the landing page of the exhibit. Try to rearrange the order of the pages to see what different structures of the exhibits pages might convey. Do some pages seem to fit naturally before or after others? What kinds of transitions or leaps from one page to another can you imagine? Each page should be able to stand on its own, but we also want to motivate people to keep reading.

Here are a few tips to keep in mind as you work:

- Break up text into sections with headings.
- Avoid walls of text by adding images or a multimedia tool.
- Quotes in large fonts (aka pull quotes) can also help to break up walls of text.
- A good guideline is 700-2000 words per page.

Additional Resources on Exhibit Design & Creation

Editorial and Language Guides

1. The CCP Editorial Manual
2. Short Guide to Writing or Teaching About Slavery

Readings about the Craft of Exhibits

1. Rabinowitz, Richard. "Eavesdropping at the Well: Interpretive Media in the Slavery in New York Exhibition." *The Public Historian*, Vol. 35 No. 3, August 2013; (pp. 8-45) DOI: 10.1525/tph.2013.35.3.8
2. Owens, Trevor. "A Draft Style Guide for Digital Collections Hypertext." (2014).

Captions are important!

Captions should describe the source and the content. Captions assist viewers in seeing and interpreting historical objects. Here are two examples of captions. These captions are important because they help ensure that CCP materials meet ADA compliance.

> Image of the exterior of Prudence Crandall School for Negro Girls (1833) formerly known as the Elisha Payne House. Courtesy of the Library of Congress.
>
> Image created by James U. Stead. "Henry Highland Garnet." 1881. National Portrait Gallery, Smithsonian Institution, London. National Portrait Gallery Digital Collection.

CCP's Approved List of Data Viz Tools & Resources

How can our exhibits use maps, graphs, and other kinds of data visualization to provide visual access to our research on the Colored Conventions? How can we use data viz tools to provide exploratory or snapshot representations of the information we have gathered? And, importantly, how should we handle the data visualizations to avoid dehumanizing the people and communities that get represented as data?

Below is a list of the data viz tools approved by the CCP. We have selected these tools according to certain criteria:

1. We know they work.
2. They are relatively friendly to learn, and tutorials are available online.
3. They can be embedded onto the CCP website.
4. We believe the output of the tools will be maintainable for the longer-term future.

If you would like to use a tool that is not on this list, please check with the CCP team first. We are always open to adding new resources to the list of approved tools, but want to ensure beforehand that the criteria have been met.

To create a timeline:

> TimelineJS - timelines
> Timeline Storyteller - multi-faceted timelines

To create a story map (interactive maps that move to a sequence of locations)

- StoryMaps
- OdysseyJS

- ESRI StoryMaps

To create maps with points (note: we find static maps often reach more people)

Palladio
Google My Maps (see tutorial here: Intro to Google Maps)

To analyze text and create graphs:

- Voyant (one of the best explained tools out there: http://docs.voyant-tools.org/)
- Juxta Commons (if you have 2+ versions of a text to compare)

To create graphs and charts:

- RAW by Density - nicely designed graphs
- Datawrapper - flexible graphs with labels
- Flourish - flexible graphs, easy to embed as html

To visualize networks:

- Gephi - network graphs (note: desktop software with a learning curve)

To process data in spreadsheets:

- OpenRefine is the standard tool to use if your spreadsheet is more than 30-40 rows. See the handy tutorial by Thomas Padilla to get up and running quickly: Getting Started with OpenRefine.

If you have a list of place names and want to convert them into latitude/longitude:

- Geocodio

Off the Tracks

Collaborators' Bill of Rights

¶ 1 1) All kinds of work on a project are equally deserving of credit (though the amount of work and expression of credit may differ). And all collaborators should be empowered to take credit for their work.

¶ 2 2) The DH community should default to the most comprehensive model of attribution of credit: credit should take the form of a legible trail that articulates the nature, extent, and dates of the contribution. (Models in the sciences and the arts may be useful.)

¶ 3 a) Descriptive Papers & Project reports: Anyone who collaborated on the project should be listed as author in a fair ordering based on emerging community conventions.

¶ 4 b) Websites: There should be a prominent "credits" link on the main page with PIs or project leads listed first. This should include current staff as well as past staff with their dates of employment.

¶ 5 c) CVs: Your CV is **your** place for articulating your contribution to a collaboration. All collaborators should feel empowered to express their contributions honestly and comprehensively.

¶ 6 3) Universities, museums, libraries, and archives are locations of creativity and innovation. Intellectual property policies should be equally applied to all employees regardless of employment status. Credit for collaborative work should be portable and legible. Collaborators should retain access to the work of the collaboration.

¶ 7 4) Funders should take an aggressive stance on unfair institutional policies that undermine the principles of this bill of rights. Such policies may include inequities in intellectual property rights or the inability of certain classes of employees to serve as PIs.

Comments

19 Pingbacks and trackbacks

1. **Getting Started in the Digital Humanities | Digital Scholarship in the Humanities** October 14, 2011 at 4:22 pm

   [...] in "Care of the Soul," and the Off the Tracks Workshop devised a useful "Collaborators' Bill of Rights.") If you can bring seed funding or administrative backing to a project, that might make it easier to [...]

2. **Who Owns This Stuff? | THATCamp Southeast 2012** March 8, 2012 at 11:01 am

   [...] and build upon the resulting code and artifacts? In this session, I propose we use the "Collaborators' Bill of Rights" as a starting point for discussion. How might we instantiate these recommendations in our [...]

3. **We are RRCHNM | Lot 49** March 8, 2012 at 3:58 pm

   [...] revealed and highlighted the names of everyone who had ever worked on this project before the Collaborator Bill of Rights existed. I asked on Twitter, how many of you look at the About page of a digital humanities [...]

4. **POST-DOCTORAL FELLOWSHIP (CLOSES 2012-07-17) « Occasional Drama** June 28, 2012 at 7:50 pm

   [...] of all persons and and affirms the dignity of all persons. MoEML is committed to honouring the Collaborators' Bill of Rights. Enquiries and applications may be sent to MoEML via Janelle Jenstad at jenstad@uvic.ca. [...]

5. **How Collaboration Works and How It Can Fail | archaeoinaction.info** June 3, 2013 at 3:00 pm

   [...] and the growth of collaborative projects involving humanities scholars, including the excellent Collaborator's Bill of Rights as well as rumination on what dangers collaboration may pose, such as my own article in JDH1-1. My [...]

6. **COLTT 2013: "The Digital Dossier" | Erin M. Kingsley** August 3, 2013 at 7:47 pm

   [...] DH Bill of Rights: including all authors/collaborators must be listed as taking some part of the project (although tasks/credit may vary); individual CVs should list individual, not group, collaboration—what did YOU do on this project – http://mcpress.media-commons.org/offthetracks/part-one-models-for-collaboration-caree... [...]

7. **Credit Transparency and the Collaborator's Bill of Rights | Introduction to Digital History** December 2, 2013 at 12:12 pm

   [...] Projects. In particular, I want to draw your attention to and work through the provisions of the "Collaborator's Bill of Rights," which is part of a larger report entitled "Off the Tracks: Laying New Lines for Digital [...]

8. **Evaluating Non-Traditional Digital Humanities Dissertations | Literature Geek** September 30, 2014 at 7:57 am

   [...] should get credit and thanks for sharing their work with others! (See the awesome "Collaborators' Bill of Rights" that came out of a MITH workshop for more on why correct credit should matter to everyone). [...]

9. **Credit Transparency and the Collaborator's Bill of Rights | Dave DeCamp** October 17, 2014 at 4:59 pm

   [...] Projects. In particular, I want to draw your attention to and work through the provisions of the "Collaborator's Bill of Rights," which is part of a larger report entitled "Off the Tracks: Laying New Lines for Digital [...]

10. **Credit Transparency and the "Collaborator's Bill of Rights" | Boston Public History** November 6, 2014 at 3:08 pm

    [...] humanities project, I want to draw your attention to and work through the provisions of the "Collaborator's Bill of Rights," which is part of a larger report entitled "Off the Tracks: Laying New Lines for Digital [...]

11. **The Pedagogy of Digital Humanities in the Liberal Arts Classroom |** April 7, 2015 at 12:45 pm

    [...] they are encouraged to include DH research projects, experiences, and skills on their resumes. The DH Collaborators Bill of Rights provides some nice initial guidelines for these [...]

12. **Creating a Group Project Charter | Introduction to Digital Humanities** November 13, 2015 at 5:06 pm

    [...] Also you might want to read this Collaborators' bill of rights [...]

13. **Milking the Deficit Internship |** January 5, 2016 at 9:42 pm

    [...] Collaborators' Bill of Rights. Off the Tracks: Laying New Lines for Digital Humanities Scholars. [...]

14. **Disrupting Student Labor in the Digital Humanities Classroom | Research and Destroy** March 18, 2016 at 10:16 am

    [...] for the principles of open access, or the guidelines for professional collaboration outlined in the Collaborators' Bill of Rights. We can develop and share resources for constructively encouraging students to produce durable [...]

15. **CETL Faculty Forum: "Developing Digital Project Assignments" Notes and Resources – Sarah E. Cornish** April 19, 2017 at 12:43 pm

    [...] For a wide selection of readings that may help you think about digital pedagogy and research ideas, browse through Debates in the Digital Humanities edited by Matthew K. Gold of the CUNY Graduate Center. I always incorporate readings on DH into my longer-term projects to get students to engage with the conversation, and I encourage them to read The Collaborators' Bill of Rights. [...]

16. **On Developing a Collaborators' Bill of Responsibilities |** September 4, 2017 at 9:11 am

    [...] guidance on these matters does exist. The Collaborators' Bill of Rights, upon which the UCLA guidelines are based, makes it clear [...]

17. **Digital Book Project – ENG 14-01 Intro. to Literary History and Interpretation** January 13, 2018 at 6:41 pm

    [...] Concerning credit, we will discuss and follow the Collaborators' Bill of Rights. [...]

18. **Collaborators' Bill of Rights – ENG 14-01 Intro. to Literary History and Interpretation** January 26, 2018 at 3:47 pm

    [...] Collaborators' Bill of Rights [...]

19. **Introduction: Issue Fourteen /** January 7, 2019 at 6:54 pm

    Comment awaiting moderation

A Student Collaborators' Bill of Rights

By CDH Guest Author (https://humtech.ucla.edu/author/cdhguest/) on June 8, 2015

By Haley Di Pressi, Stephanie Gorman, Miriam Posner, Raphael Sasayama, and Tori Schmitt, with contributions from Roderic Crooks, Megan Driscoll, Amy Earhart, Spencer Keralis, Tiffany Naiman, and Todd Presner

UCLA's Digital Humanities program emphasizes cross-disciplinary, cross-hierarchy collaboration among students, faculty, and staff. We've created this Student Collaborators' Bill of Rights as a statement of our values and principles in the UCLA DH program.

Collaborations between students and more experienced digital humanities practitioners should benefit everyone. At their best, these partnerships are a way for students to learn new skills and benefit from mentorship, while more seasoned scholars can learn from junior scholars' ideas, skills, subject knowledge, and perspectives.

It's important, though, to recognize that students and more senior scholars don't operate from positions of equal power in the academic hierarchy. In particular, students' DH mentors may be the same people who give them grades, recommend them for jobs, and hold other kinds of power over their futures. Students may not feel entirely comfortable raising objections to certain practices if they feel these objections

could endanger their academic or career prospects. Thus, we think it's important to outline some best practices for collaborations with students on digital humanities projects, so that everyone involved feels they gain from the partnership.

Collaboration can take many forms, from casual brainstorming to full-time employment. As collaborations develop, senior scholars should be mindful that different kinds of relationships entail different responsibilities on the part of each collaborator. A professor who assigns a class project, for example, must primarily consider the student's own intellectual growth, while a senior scholar who employs a student assistant may assign work that primarily benefits the project.[1]

We endorse the principles outlined in the **Collaborators' Bill of Rights (2011)** . As additional safeguards for students, we advise those embarking on collaborations with students to adhere to the following principles:

1. As a general principle, a student must be paid for his or her time if he or she is not empowered to make critical decisions about the intellectual design of a project or a portion of a project (and credited accordingly). Students should not perform mechanical labor, such as data-entry or scanning, without pay.
2. Course credit is generally not sufficient "payment" for students' time, since courses are designed to provide students with learning experiences.
3. We encourage senior scholars to familiarize themselves with the literature on unpaid internships.[2] At a minimum, internships for course credit should be offered as learning experiences, with a high level of mentorship. Those employing interns should be prepared to spend substantial face-to-face time with the student.
4. If students have made substantive (i.e., non-mechanical) contributions to the project, their names should appear on the project as collaborators, and they should be acknowledged in subsequent publications that stem from the project.
5. Students should be empowered to present on projects on which they have collaborated (assuming reasonable limitations about sensitive and embargoed material and on work in progress). Students and senior scholars should discuss the protocol for such presentations at the outset of the project.
6. Students should be empowered to list their collaboration on a project on a CV or résumé, with an appropriate degree of credit. Senior scholars should explicitly encourage this and help students to formulate meaningful statements about their contributions.
7. Senior scholars should recognize that projects on which students have collaborated represent important components of students' scholarly portfolios. Senior scholars should thus make every reasonable effort to either sustain a "live" project or, failing this, either transfer its ownership to student collaborators or distribute to students an archived version or snapshot of the project.
8. When digital humanities projects are required for course credit, instructors should recognize that students may have good reasons not to engage in public-facing scholarship, or may not want their names made public, and should offer students the option of alternative assignments.
9. In meetings and project communication, student collaborators should be treated as full members of the project team, to the extent that this is reasonable, and their contributions should be valued and respected. Students should have a clear sense of how their work fits into the larger project.
10. Digital projects can sometimes branch into multiple projects, or head in multiple directions. Many digital projects are experimental. Mentors and students should set guidelines for re-use of digital scholarly material, as well as for maintaining meaningful artifacts of students' contributions.

1. For more on these roles and the various responsibilities they entail, see Spencer Keralis, "Disrupting Labor in the Digital Humanities; or, The Classroom is Not Your Crowd," in *Disrupting Digital Humanities*, ed. Jesse Stommel and Dorothy Kim (Punctum Books, 2015 [forthcoming]).

2. See, for example, **ProPublica's reporting on unpaid internships** .

Humanities Technology is part of the **Humanities Division (http://humanities.ucla.edu)** within **UCLA College (http://www.college.ucla.edu/)**.**(http://www.ucla.edu)**

1020 Public Affairs Bldg. | Los Angeles, CA 90095 | **P:** 310-206-1414

Chapter Title: Making a Case for the Black Digital Humanities

Chapter Author(s): KIM GALLON

Book Title: Debates in the Digital Humanities 2016

Book Editor(s): Matthew K. Gold and Lauren F. Klein

Published by: University of Minnesota Press

Stable URL: https://www.jstor.org/stable/10.5749/j.ctt1cn6thb.7

Making a Case for the Black Digital Humanities

KIM GALLON

The dust has yet to settle around the debates over what the digital humanities is or is not. Boundaries and demarcations continue to shift within a complex and ongoing conversation about the intersection of technology with humanistic fields. This context, I would argue, has generated the ideal conditions in which to engage the question of how humanity is framed in the digital humanities. To this end, I seek to articulate a relationship between the digital humanities and Africana/African American/Black studies (from here on I will call the field Black studies) so as to highlight how technology, employed in this underexamined context, can further expose humanity as a racialized social construction.

Questions may arise around the use of the term "black." Would not "Africana" or "African American" be more appropriate, some may ask. In other contexts, I am quite sure that my addition of a racial signifier to "digital humanities" would appear at the most racist and at a minimum divisive, leading to questions about who could or could not engage in black digital humanities. Questions of this magnitude are to be expected and are in fact necessary when new areas of inquiries are proposed. At the same time, these sorts of questions obfuscate crucial complexity, making it difficult to chart the paths needed to address much deeper and systemic issues. To get caught up in exact definitions or questions of "who is in or who is out" in black digital humanities is to ignore how the very nomenclature of blackness has a complex and rich history that moves in the same conceptual orbit as the term "digital humanities" (Parham, "Without Innovation").

Although work on racial, ethnic, and national difference is emerging in the digital humanities, discussions about the lineage of Black studies within the digital humanities are almost nonexistent.[1] While a comprehensive history of the intersections between Black studies and the digital is sorely needed, it is outside of the scope of this chapter. Here, I seek to set in motion a discussion of the black digital humanities by drawing attention to the "technology of recovery" that undergirds black digital scholarship, showing how it fills the apertures between Black studies

and digital humanities. Indeed, the black digital humanities help to unmask the racialized systems of power at work in how we understand the digital humanities as a field and utilize its associated techniques. In their work with the #transformDH collective, Alexis Lothian and Amanda Phillips have suggested that putting a name to the unnamed helps to bring a concept into existence (Lothian and Phillips, "Can Digital Humanities Mean Transformative Critique?"). Thus, this piece names the "black digital humanities" as the intersection between Black studies and digital humanities, transforming the concept into corporeal reality while lending language to the work of the black digerati in and outside of the academy.

Like Matthew Kirschenbaum's understanding of the term digital humanities itself, precise definitions of what constitutes the black digital humanities are elusive. The black digital humanities reflects less an actual "thing" and more of a constructed space to consider the intersections between the digital and blackness (Kirschenbaum, 51). Like race, gender, class, and sexuality—all social constructs, if you will—the digital humanities increasingly hold *real* meaning and significance in the academic universe. As Kirschenbaum has observed, there are high stakes for who is and who is not a digital humanist, and for what is or is not digital humanities, when federal grants are hard to come by and academic jobs may hinge on the term (Kirschenbaum, 54–55). Some digital humanities scholars have begun to call attention to the role that race may play in the development of digital humanities programs and centers and in the funding and recognition that particular digital humanities projects might garner (Bailey, "All the Digital Humanists Are White"). A vibrant and critical discourse from #dhpoco, #transformDH, and HASTAC (Humanities, Arts, Sciences, and Technology Advanced Collaboratory), among others, now serves to resist the academic hegemonies that may limit our understanding of what the digital humanities is and will be in the future. My hope is that a critical consideration of the connections between Black studies and the digital humanities will help to advance this work.

The field of Black studies is nearing its fiftieth birthday, having developed out of the civil rights and Black Nationalist movements in the late 1960s. Black studies has long been understood as the comparative study of the black cultural and social experiences under white Eurocentric systems of power in the United States, the larger African diaspora, and the African continent, after all, and these systems of power endure. Contemporary scholars such as Alexander Weheliye therefore describe "black studies as a mode of knowledge production" that "investigates processes of racialization with a particular emphasis on the shifting configurations of black life" (Weheliye, 3). He continues:

> If racialization is understood not as a biological or cultural descriptor but as a conglomerate of sociopolitical relations that discipline humanity into full humans, not-quite-humans, and non-humans, then blackness designates a changing system of unequal power structures that apportion and delimit which humans can lay claim to full human status and which humans cannot. (Weheliye, 3)

Weheliye asks us to consider how Black studies might illuminate the various processes by which nonwhite subjects are systematically shut out from "the category of human as it is performed in the modern west" (Weheliye, 3). His conception of Black studies is powerful in its assertion that modern humanity cannot be dislocated from a racialized hegemony.

What does this mean for digital humanities? Following Weheliye, I would argue that any connection between humanity and the digital therefore requires an investigation into how computational processes might reinforce the notion of a humanity developed out of racializing systems, even as they foster efforts to assemble or otherwise build alternative human modalities. This tension is enacted through what I call a "technology of recovery," characterized by efforts to bring forth the full humanity of marginalized peoples through the use of digital platforms and tools.

Recovery rests at the heart of Black studies, as a scholarly tradition that seeks to restore the humanity of black people lost and stolen through systemic global racialization. It follows, then, that the project of recovering lost historical and literary texts should be foundational to the black digital humanities. It is a deeply political enterprise that seeks not simply to transform literary canons and historiography by incorporating black voices and centering an African American and African diasporic experience, though it certainly does that; black digital humanities troubles the very core of what we have come to know as the *humanities* by recovering alternate constructions of humanity that have been historically excluded from that concept. A discourse on the "politics of recovery" in the digital humanities is beginning to take shape through Amy Earhart's work. She documents a history of what she calls "DIY recovery projects" in the 1990s that sought to disrupt a canon of Eurocentric and male-authored literature. Through the lens of black digital humanities, these efforts at recovery can be understood not only as the recovery of "lost or non-canonical and difficult to locate texts," but also as the recovery of black authors' humanity (Earhart, "Can Information Be Unfettered?").

Applied as a technology in Black studies and in the lives of black people living in the digital era more generally, recovery restores black people's humanity. This technology of recovery operates as the shared basis for black academic and nonacademic digital work, one that dominates the ways by which both Black studies scholars and a black public approach technology. Everyday discursive interactions on social media networks are a case in point. Black people's subsistence in and resistance to the complex oppressive systems of slavery, colonialism, Jim Crow, mass incarceration, and police brutality, across time and space, make black lives ground zero for a technology of recovery using social media. Movements that protest the ongoing police brutality of black women and men, which began on "Black Twitter" and Facebook with hashtags such as #SayHerName, #BlackLivesMatter, and #ICantBreathe, continue black people's centuries-old endeavor to make their collective humanity apparent to the world. These hashtags reveal that black people's humanity is tethered to a racial system that deems black people's lives as insignificant relative to

their white counterparts. Tweets that highlight disparities in social indicators such as employment, education, housing, and healthcare between white and black people show how black people's humanity has material consequences.

In addition to Twitter, scholars and institutions (along with nonacademic users) have developed literary and historical digital recovery projects that similarly represent a search and mission for the collective recuperation of a lost peoplehood. The *Digital Schomburg,* one of the earliest black digitization projects, demonstrates the power of reclaiming black humanity by recovering nineteenth-century black female writing[2] and late nineteenth and early twentieth century images of people of African descent.[3,4] It may then be of little surprise that scholars of the black literary tradition, as a whole, have yet to embrace text mining and other quantitative digital approaches in the same numbers as other groups of literary scholars. Scholars of African American literature may view text mining as counterposed to recovery (Rambsy, "African American Literature and Digital Humanities"). The relatively small number of text mining projects among scholars of black literature is concerning, however, at a time when digital humanities work has shifted its focus to quantitative and computational approaches. But the black digital humanities can highlight the value of specific computational methods. Kenton Rambsy, Assistant Professor of African American Literature at the University of Texas at Arlington and the Project Digital Initiative Coordinator for the Project on the History of Black Writing, models this approach. Noting that mobility and place are predominant themes in African American literary expression, he uses text mining software to geo-tag the occurrence of city and other geographical landmark names in black literary expression (Rambsy, "African American Literature and Digital Humanities"). For example, text mining allows Rambsy to recover Edward P. Jones's use of cities, streets, neighborhoods, and city landmarks to reenvision forms of black humanity that are not completely circumscribed by racism ("Edward P. Jones and Literary Geo-Tagging").

Rambsy's work stresses another key point: digital recovery projects that are either led by or heavily involve black scholars are particularly impactful in how they expand what we understand the digital humanities to be and its potential for critically thinking about power. As a scholar of African descent leading the digital program of the thirty-two-year-old *Project on the History of Black Writing* (HBW)—which was founded by another black literary scholar, Maryemma Graham, with a group of African American literary scholars at an organizing meeting entitled *Computer Assisted Analysis of Black Literature* (CAABL)—Rambsy produces work that disrupts the normative and racialized framework of the digital humanities as led by white scholars.[5] Digital humanities projects exclusively developed by white scholars and information technology staff often reflect the racial hierarchies present in higher education. Mark Anthony Neal views the small number of black scholars in the digital humanities as an administrative issue. He observes, "When all these deans and provosts are looking around for the folks who are going to do cutting edge

work, the last folks they think about are black folks" (*Left of Black*). Neal's comments touch on the unspoken assumption that African Americans are technophobes, even in the midst of the information age. The supposition that black people are averse to technological innovation is tied to the discourse of "black technophobia" that still circulates today, reproducing and reinforcing long-standing "scientific" evidence of black intellectual inferiority (Everett, 19).

From the vantage of black digital humanities, foundational assumptions about humanity, as well as about how we derive meaning about human culture in the academy, remain deeply entrenched in racialization, and the digital humanities are not exempt from this charge. Like many disciplines that study humanity, discussions about digital tools and processes are most often considerations about how majority groups use or might be studied with computational approaches. Thus, the large share of digital humanities projects and related scholarship that pays no attention to race should be defined as the "white digital humanities," for they are, in practice, explorations about human culture based on whiteness as an unmarked category and "standard of the real" (Gordon, 79).[6]

The racialization of black people's humanity therefore poses a fundamental problem to the digital humanities as it is generally defined. Understood as the union of digital technology and the academic disciplines that study human culture, what do we do with forms of humanity excluded from or marginalized in how we study the humanities and practice the digital humanities? What are the implications of using computational approaches to theorize and draw deeper insight into a modern humanity that is *prima facie* arranged and constructed along racial lines? One of the essential features of the black digital humanities, then, is that it conceptualizes a relationship between blackness and the digital where black people's humanity is *not* a given. The black digital humanities probes and disrupts the ontological notions that would have us accept humanity as a fixed category, an assumption that unproblematically emanates in the digital realm. The black digital humanities, then, might be defined as a digital episteme of humanity that is less tool-oriented and more invested in anatomizing the digital as both progenitor of and host to new—albeit related—forms of racialization. These forms at once attempt to abolish and to fortify a taxonomy of humanity predicated on racial hierarchies.

What, then, do the black digital humanities mean for the humanities and its relationship to digital tools? Rather than moving forward with digitizing, text mining, topic modeling, and the like, the black digital humanities would have us seriously consider the political relations and "assemblages" that have racialized the literary, philosophy, and historical texts that we study (Weheliye, 3). Digital tools and platforms should be mobilized to interrogate and disclose how the humanities are developed out of systems of power. The black digital humanities reveals how methodological approaches for studying and thinking about the category of blackness may come to bear on and transform the digital processes and tools used to study humanity. Questions pertaining to digital tool development have much broader

applications, of course. Johanna Drucker, for instance, reminds us that we must use and build digital infrastructure and tools steeped in humanistic theory so that they function in ways that reflect the core values of the humanities (Drucker, 87). However the black digital humanities forces us to move backward before moving forward in thinking about tools, to first consider how the very foundation of the humanities are racialized through the privileging of Western cultural traditions. It then asks us to assess whether those tools would still be used in the same manner had they been developed to explore the texts that were and are marginalized through the racialization of the humanities. It further prompts us to ask how tool building might mirror the material realities of blackness. The black digital humanities therefore foregrounds the digital as a mutual host for racism and resistance and brings to light the "role of race as a metalanguage" that shapes the digital terrain, fostering hegemonic structures that are both new and old and replicate and transcend analog ones.[7]

Ultimately, the task of black digital humanities is to ask, "What aspects of the digital humanities might be made more "humanistic" if we were to look at them from the perspective of blackness?" The black digital humanities raises the question, "How can digital tools and processes such as text mining and distant reading be justified when there is so much to do in reconstructing what it means to be human?"[8] Black digital humanity, with its emphasis on humanity as an evolving category, also changes how we should view the ongoing concerns about sustainability and the future of digital projects. Recognizing that humanity is a construct, a contingent idea, forces digital humanists to come to terms with the contingency of digital projects. How might the sustainability of a digital project be conceptualized from a standpoint that considers humanity as a social construction and subject to change over time and place? Accordingly, the black digital humanities promotes a system of change; it is a mechanism for deregulating the tendency of technological tools, when employed in the digital humanities, to deemphasize questions about humanity itself.

Thus, I make the case for the black digital humanities in order to, as Alan Liu suggests, enlarge the field with "sociocultural meaning" (Liu, 501). Black digital humanities provides a forum for thinking through the ways that black humanity emerges, submerges, and resurfaces in the digital realm through the "racializing assemblages of subjection" (Weheliye, 2). My articulation of this union does not dismiss or marginalize other efforts working at this nexus, such as eblack studies, black code studies, and digital blackness.[9] They all provide compelling methods for describing how the digital comes to bear on blackness and vice versa. But there is a need for these and more theorization on the topic so that they might contribute to a larger black technocultural discourse and Internet activism. Black studies has a unique role to play in dismembering how we think about humanity and the digital humanities by extension. A black epistemology will generate questions about the relationship between the racialization of humanity and the digital as power, ultimately fostering new inquiries and deeper understandings about the human condition.

NOTES

This chapter was developed from a presentation titled "Creating a Digital Culture for Scholarship on the Black Press," which I gave at the African American Expression in Print and Digital Culture at the University of Wisconsin in the fall of 2014. At the time I met Amy Earhart, who encouraged me to continue thinking about the relationship between blackness and the digital. My participation in "Recovering African American and African Diaspora History and Literary in the Digital Humanities: A Roundtable Discussion," with Jessica Johnson, Robby Luckett, and Bryan Carter at the 2015 Annual American Historical Association Meeting, expanded my thoughts about recovery in the digital humanities. Thanks to Roopika Risam, Matthew K. Gold and Lauren Klein for their critical and insightful feedback on this essay. A special thank-you to Lewis R. Gordon and Alexander Weheliye and other scholars of Africana philosophy and the black intellectual tradition, present and past, for providing me with a conceptual language and understanding about blackness.

1. For some of the scholarship on difference in the digital humanities, see http://transformdh.org/about-transformdh/ and https://www.hastac.org/explore/social-political-issues/race-ethnicity.

2. http://digital.nypl.org/schomburg/writers_aa19/toc.html.

3. http://digital.nypl.org/schomburg/images_aa19.

4. http://www.nypl.org/about/locations/schomburg/digital-schomburg.

5. https://hbw.ku.edu.

6. Both Moya Bailey and Tara McPherson implicitly make this argument with their article titles: "All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave" and "Why Are the Digital Humanities So White?"

7. On the "metalanguage of race," see Higginbotham.

8. My question is heavily modeled off the question that Africana philosopher Lewis Gordon poses about the role of philosophy in relationship to Africana philosophy. See *Introduction to Africana Philosophy.*

9. http://eblackstudies.org; http://diasporahypertext.com/2015/02/13/cfp-black-code-studies/; http://www.rutgersdigitalblackness.com.

BIBLIOGRAPHY

Bailey, Moya Z. "All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave." *Journal of Digital Humanities* 1, no. 1 (2011). http://journalofdigitalhumanities.org/1–1/all-the-digital-humanists-are-white-all-the-nerds-are-men-but-some-of-us-are-brave-by-moya-z-bailey/.

Digital Schomburg: African American Women Writers of the 19th Century. New York City Public Library, 1998. http://digital.nypl.org/schomburg/writers_aa19/.

Drucker, Johanna. "Humanistic Theory and Digital Scholarship." In *Debates in the Digital Humanities,* ed. Matthew K. Gold, 85–95. Minneapolis: University of Minnesota Press, 2012.

Earhart, Amy E. "Can Information Be Unfettered?: Race and the New Digital Humanities Canon." In *Debates in the Digital Humanities,* ed. Matthew K. Gold, 309–18. Minneapolis: University of Minnesota Press, 2012.

Everett, Anna. *Digital Diaspora: A Race for Cyberspace.* Albany: SUNY Press, 2009.

Gordon, Lewis R. *An Introduction to Africana Philosophy.* Cambridge: Cambridge University Press, 2008.

Higginbotham, Evelyn Brooks. "African American Women's History and the Metalanguage of Race." *Signs* 17, no. 2 (1992): 251–74.

Images of African Americans from the 19th Century. New York City Public Library, 1999. http://digital.nypl.org/schomburg/images_aa19/.

Kirschenbaum, Matthew, G. "What Is 'Digital Humanities' and Why Are They Saying Such Terrible Things About It?" *Differences: A Journal of Feminist Cultural Studies* 25, no. 1 (2014): 46–53.

Liu, Alan. "Where Is Cultural Criticism in the Digital Humanities?" In *Debates in the Digital Humanities,* ed. Matthew K. Gold, 490–509. Minneapolis: University of Minnesota Press, 2012.

Lothian, Alexis, and Amanda Phillips. "Can Digital Humanities Mean Transformative Critique?" *Journal of e-Media Studies* 3, no. 1 (2013). https://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/xmlpage/4/article/425.

McPherson, Tara. "Why Are the Digital Humanities So White? or Thinking the Histories of Race and Computation." In *Debates in the Digital Humanities,* ed. Matthew K. Gold, 139–60. Minneapolis: University of Minnesota Press, 2012.

Neal, Mark Anthony. "Race and the Digital Humanities." *Left of Black* (webcast), season 3, episode 1, John Hope Franklin Center, September 17, 2012. https://www.youtube.com/watch?v=AQth5_-QNj0.

Nelson, Alondra. "Introduction: Future Texts." *Social Text* 20, no. 2 (Summer 2002): 1–15.

Project on the History of Black Writing. "HBW History: The History of the Project on the History of Black Writing." http://www2.ku.edu/~phbw/about_us_history.html.

Parham, Marisa. "Without Innovation: African American Lifeworlds and the Internet of Things," October 14, 2014. http://mith.umd.edu/podcasts/dd_fall-2014-marisa-parham.

Rambsy, Kenton. "African American Literature and Digital Humanities," January 17, 2014. http://www.culturalfront.org/2014/01/african-american-literature-and-digital.html.

———. "Edward P. Jones and Literary Geo-Tagging," January 17, 2014. http://www.culturalfront.org/2014/01/edward-p-jones-and-literary-geo-tagging.html.

——— ."Text-Mining, Geography, and Canonical African American Short Stories," January 17, 2014. http://www.culturalfront.org/2014/01/text-mining-geography-and-canonical.html.

Weheliye, Alexander G. *Habea Viscus: Racializing Assemblages, Biopolitics, and Black Feminism Theories of the Human.* Durham, N.C.: Duke University Press, 2014.

ARCHIVING CONFERENCE
IS&T imaging.org
https://www.ingentaconnect.com/contentone/ist/ac/2006/00002006/00000001/art00020

Metadata Creation for Digital Humanities Projects

Martyn Jessop; King's College London; London; United Kingdom.

Abstract

The Centre for Computing in the Humanities at King's College London is a research centre that conducts highly collaborative research projects with partners from the academic community and cultural heritage organizations primarily in Europe but also further a field in North America, Asia, Africa and Australia. These projects encompass disciplines such as art history, social history, linguistics, literature studies and music. These projects frequently result in the creation of digital resources utilizing a variety of technologies and methods. They provide an opportunity for CCH to research issues concerning the digitization, design, implementation and delivery of such resources. This paper focuses on issues surrounding the compilation of metadata and the effects that these have by focusing of three case study projects.

Introduction

It is acknowledged that metadata performs many roles within a digital resource but it is easy to overlook the fact that it has to fulfil these roles for different audiences of the same resource. Carefully compiled metadata greatly enhances the value of a resource to specialist users but it can also open up the same material to a much wider audience thus adding substantial extra value and helping to fulfil the wider social role of humanities computing projects. However the compilation of such metadata, even with well defined schemas, is not a simple task.

The Case Studies

This paper discusses the challenges of metadata design and compilation and the effects they have on the audience and usage of the resources produced. The discussion will be conducted through three case studies projects at the Centre for Computing in the Humanities (CCH), King's College London.

The Corpus Vitrearum Medii Aevi

The Corpus Vitrearum Medii Aevi (CVMA) is an international research project dedicated to the publication of all medieval stained glass. The publication resulting from this exploratory pilot project of glass in England and Wales is published on a website [1]. The project digitized 12,500 images of stained glass windows and has made them available via an on-line database.

The main aim of this project was to satisfy the needs of the specialized academic community but also to seek ways in which the resource could be opened up to other academics within the art history community and elsewhere. The objects within the resource are also of interest to members of the public such as local historians and to many casual viewers to whom the high quality images of the windows will appeal. As the metadata for this project was designed primarily for the academic art historian it includes information concerning provenance and detailed metadata describing the type of building and the position of the glass within the building defined by a specialized nomenclature [2]. It also includes information describing the geographical location of the building and the age of the windows.

Many of the images were taken when the glass was removed from the building for restoration purposes. When in-situ these windows are often positioned high up in the walls of the building with restricted views. The resultant digital resource therefore contains many thousands of images that show views of the windows that are far superior to those that can be seen by visiting the original site. This is clearly of great value to art history scholars but it also makes them of interest to members of the public who are interested in local history or casual tourism visits to churches to view the glass.

One of the briefs of the funding body (Arts and Humanities Research Council) was that value should be added to the project by designing the interface in such a way that it could be used by a wider audience within the general public. We had to enable the more casual viewer to discover the resources without having a knowledge of the highly specialized classifications and terminology of medieval stained glass experts. The use of location metadata and a carefully designed resource discovery tool built around it played a key role in this. This interface allows users to locate and view stained glass by county, by place name, or by pointing and clicking on a map.

A further development has been suggested as a result of comments made by members of the wider audience who have seen the pilot project. The windows contain imagery that is of great interest to those who study religious iconography [3], whether from an artistic viewpoint, as a theologian [4], or in anthropological studies. This would involve adding a further set of metadata to describe the imagery used within the windows and increase the value, and audience, of the existing images within the resource. This work is now at an early stage of development.

On-line Slide Library for Classics teaching

The second case study is a much smaller project that was conducted for the Classics department at King's College London. The study of classical civilisations relies heavily on the examination the artefacts left by those civilisations. Researchers will travel to the museum, gallery or archive where these objects reside but this is impractical when teaching involves the presentation of multiple objects from scattered locations to a class of students. The traditional method of showing these objects to a group of students has been via the use of photographic slides and a projector, frequently using two projectors so that images can be projected side by side for comparison. In 2004 a major manufacturer of the projection equipment used for these purposes announced that they would no longer be making projectors and it became clear that the technology for presenting 35mm photographic slides would soon become obsolete. The Classics department at King's approached CCH to see whether it was

feasible to create an on-line slide library that could be used in lectures via the local network and PCs attached to data projectors. It would also make the images available to students for private study on the College network. This project, Humslides, was designed with a much more limited audience in mind than CVMA. It illustrates how the design and creation of even the simple metadata intended for use by a relatively small closed community of Classics department staff and students produced problems that could have limited the use of the resource even among this limited audience.

The metadata schema was relatively easy to devise. The fields that were of importance to the scholars were caption, location (geographical), description, creator, keywords and date. These generated the standard problems of metadata describing historical objects. For example for location do you use the current place name or the name that would have been in use at the time, e.g. for a Roman object found in modern day Colchester should you use the contemporaneous name Camulodunum? Do you give the location where it was originally set up or the location that it is at now? For objects found in areas where a mixture of languages are spoken which do you use for the place name? How do you define dates, by century or more accurately? Historical data is rarely definitive or precise so the previous question raises an important general issue, how do you deal with ambiguity and uncertainty? How do you deal with missing values? Some of these factors can be accommodated by having a carefully prepared schema and guidance notes for the metadata editor but many revolve around scholarly issues and require not only input from experts in the field but also discussion between them to resolve ambiquities and differing opinions (or at least arrive at a context for decisions that can be stated in the metadata).

The project was intended to allow academics to contribute images that they needed for their own teaching. In this case the contributors are the primary end-users of the images, they are also the experts with whom the knowledge required to generate the associated metadata resides. The ideal workflow for this project would have been to allow academics to upload their own images and create the metadata for them. This proved impractical for many reasons. While many people were happy to provide images few were prepared to put in the long tedious hours of thinking about and typing in descriptive metadata. Five thousand slides were prepared for the project and of these we initially used 3600 on the site. Many of these had inadequate metadata, the lecturers who submitted the images knew the reference numbers of their own slides and could therefore find them and use them. This satisfied the basic level of functionality but completely failed the project's aim of opening up a shared resource as it would have been impossible to find images that were submitted by other users other than by browsing the entire collection. Even if this haphazard approach resulted in finding an image that could be of interest there was insufficient information in the metadata to confirm the identity and nature of the object.

Making the submitter responsible for the creation of the metadata also had an unexpected effect. The images were intended for a specialist audience of Classics lecturers and students so one would have expected any member of that group to produce metadata that was of value to the whole group. This was not the case as each submitter often had a very specific reason for selecting an image. The same image could be of interest to other members of the group but for reasons other than those anticipated by the contributor. For example consider an image of a courtyard containing a piece of sculpture standing on an inscribed stone plinth. This would be of interest to someone teaching about the history of the building around the courtyard, others might be interested in the sculpture as a work of art; someone else might be interested in the person depicted by the sculpture. In fact, in this case the image was contributed by a lecturer who wished to use it to study only the inscription and therefore omitted any further information from the metadata. This is understandable but seriously undermines the aim of creating a set of images that can be used by multiple users for a wide range of teaching purposes even within a specialized group. This problem was overcome by introducing an editor role between the submitter and the actual entry of the metadata into the system. This role was taken by someone who had knowledge of the subject material but was also able to see how the content of a particular image could be of value to more than just the submitter and to liaise with them to expand the descriptions within the metadata to cover the possible interests of many potential users.

Digital Image Archive of Medieval Music

The final case study is the Digital Image Archive of Medieval Music (DIAMM) [5] whose aim is to obtain and archive directly-captured digital images of European sources of medieval music. The project has created a new permanent electronic archive of over 14,000 of these images, both to facilitate detailed study of this music and its sources, and to assure their permanent preservation. The sources archived include all the fragmentary sources of polyphony up to 1550 in the UK; all the 'complete' manuscripts in the UK; a small number of important representative manuscripts from continental Europe; a significant portion of fragments 1300-1450 from Belgium, France, Italy, Germany and Spain. Such a collection of images that includes the complete British fragments has never before been possible, and represents an extraordinary resource for study of the repertory as a whole.

The project uses two distinct sets of metadata; one set records information about the capture of the images the other is drawn from existing public catalogues of the source materials. The image capture metadata includes standard photographic information and details of the digital image files and their preparation. It is rather more extensive than one might expect because the creators were trying to build in a degree of 'future proofing' by including information which although limited in value now may be of greater value in the future.

The catalogues on which the content metadata were based began to be compiled in the 1950's and the process took over twenty years. During that time the collections were often re-catalogued so the information contains multiple shelf marks for the same items. Each catalogue was compiled by different people using different criteria and in many different languages. These factors alone posed many problems for the project. However the most vexing problem was that the catalogue entries are written in free prose with no standard layout design and even within a single catalogue there are substantial variations in content that reflect the specific interests of the many individuals who compiled the entries; for example some went into great detail about the bindings and watermarks while others might dwell of the history of ownership of the manuscript at the expense of anything else.

Despite these limitations the staff soon became acutely aware of the richness of the free prose entries as they worked to split them up to form the basis for the metadata extraction. There were inevitable ambiguities, pieces of information that were missing from the original catalogues, and the occasional mistake. In an attempt to overcome these and some of the problems described previously it was decided to include a wiki-style feature of the website based resource that allows scholars to add their own annotations to the images, the intention being that they could supply missing data and correct anything that was wrong or contentious. It was anticipated that although precautions would have to be taken to prevent malicious or unauthorized annotations, and that there may also have been the occasional academic dispute between scholars, this mechanism would provide a useful tool for tackling many weaknesses of the metadata. In practice it was found that there was very little use of this feature; this is probably another facet of the problems surrounding persuading users to supply metadata.

Another possible strategy for coping with the complexity and variability of the metadata was to include a 'fuzzy' search mechanism. This has proved very difficult to implement for a variety of reasons and has not yet been added.

The project would be enhanced considerably by the inclusion of full text transcriptions of the material but this is far beyond the capabilities of the most sophisticated optical character recognition software. The only way of linking text to images in this project is by physical references, for example '4th line down, three inches from the left'.

The creation of the metadata was made harder because it had to be extracted from existing, non-standard, catalogues that were established as important sources themselves. In many respects it would have been easier to have created the metadata from scratch; however it had to be compatible with the standard pre-existing reference works. The project was intended to be primarily a collection of images of music manuscripts and the aim of the metadata was purely to support the image collection. The metadata creation was a difficult and time consuming task but it has proved to be the most popular aspect of the project among the users. The original catalogues were expensive books and therefore available in only a very few institutions. The availability of the standardized metadata derived from them through the website has greatly improved access to the textual content of the original catalogues.

Conclusions

These projects are very different in their missions, content, approach and principal target audiences but by studying them it is possible to draw out a number of common themes.

Each project has four types of audience

- The principal intended audience of scholars with a high level of knowledge about the content
- Students of the subject with a more limited level of specialist knowledge
- Scholars in disciplines other than those the resource was originally intended for who find the material useful in their own fields
- Members of the public with little or no specialized knowledge of the content

The metadata must be designed in such a way that allows each audience to find and identify the object that they are interested in. This can be facilitated by providing browse and search mechanisms that work at different levels of complexity, by using constrained searches that utilize drop down menus, and by graphical navigation aids such as interactive maps of varying scales.

Each project needs a metadata schema that can accommodate the inconsistency, ambiguity and contentiousness that often characterizes historical data. These show the importance, and difficulties of metadata compilation and the need to involve specialists in its creation. The metadata for CVMA project was perhaps the easiest to create because although the data was compiled from older catalogues the areas in which the metadata had to conform with these catalogues were limited. In many respects the CVMA metadata could be compiled according to a design of its own, effectively being created from scratch. The team responsible for this was very small, had full editorial control and on the whole the material was not contentious. The texts were all in English and were structured according to a standard format. The Humslides metadata was simple in design but posed a number of problems, not least of which was the fact that it had to be gleaned from a wide number of academics who were happy to hand over their slides for digitization but understandably daunted when asked to provide descriptive information for several hundred images. In many cases the slides came from two sources; small 'personal' collections and larger departmental collections. The small 'personal' collections of a few hundred slides were usually accompanied by detailed metadata that had been compiled by the submitter. The ownership of the larger departmental collections was often unclear or undefined and it proved to be far harder to obtain metadata for these more sizeable collections. The project was a pilot scheme and as such provided a test-bed for different ways of involving the contributors. Slides were put up with minimal metadata in the hope that this would encourage image submitters to contribute metadata, this worked in some cases but also resulted in complaints from users and, ironically, some of the submitters of the images themselves (who had failed to provide adequate metadata). Where metadata was provided it often reflected only the interests of the submitter and did not allow other users, to whom the images would be useful, to find them. The final project, DIAMM, experienced the greatest difficulties with metadata creation, but also produced a resource in which the metadata itself has proved to be, in the eyes of many of its users, more valuable to scholars than the content itself. The challenges here were integrate the metadata with existing public catalogues that are important reference works themselves but are of differing formats, approaches and languages. This is not an easy task and requires enormous amounts of subject knowledge, technical expertise and hard work over a protracted period of time but DIAMM shows that it can be done and does result in a very worthwhile resource.

Each project needs a workflow that maximises user contributions but ensures the creation of extensive good quality metadata that is suited to a range of potential users even within what are considered to be specialized audiences as well as the general public. It is essential that metadata is compiled by the recognized experts in their field but these are busy people who, while they may be fully committed to the aims of the project, have very little time. These projects have not found a solution to this. Allowing on-line access to metadata records through a Wiki-style

service is an obvious solution but requires careful attention to system security and the issuing of passwords to authorized users. There can still be disputes between different scholars as to the content. The DIAMM project has shown that providing interactive online editing access does not in fact solve the problems of gathering contributions from hard-pressed academics. Humslides does perhaps show a possible way forward, in this case the metadata editor was a graduate student in Classics who could compile an initial entry based on his own knowledge and ask the relevant academic to comment on it. There was a marked improvement in contributions when this scheme was implemented. We were fortunate to have such a person working for us. It has been suggested that documentation could be produced to guide metadata editors. In practice this could only cover a limited range of situations and could not be used to extract metadata from the free-prose style descriptions that constitute the sources of many humanities computing projects.

Projects often go through a pilot stage and several phases of development with each phase being funded separately and not necessarily running contiguously. The aims of the project can easily change in each phase, frequently expanding or changing the focus of the target audience. An example of this is the decision to extend the metadata within the CVMA project to encompass religious iconography, this is an obvious feature of the content but is far beyond the original remit of the project. Very well designed metadata schemas may be able to accommodate this but it is more likely that the metadata will have to be extensible.

The content of metadata can be greatly influenced by the backgrounds and interests of the individuals who compile it. Their specialist knowledge contributes enormously to the success of the final digital resource. However care needs to be taken that the metadata reflects the broader purposes of the project and opens up the resource to as wider audience as possible.

References

[1] Corpus Vitrearum Medii Aevi (CVMA) at www.cvma.ac.uk, last accessed 01/03/2006.

[2] Marks, Richard, *Stained Glass in England during the Middle Ages,* London, 1993

[3] Kemp, Wolfgang, The Narratives of Gothic Stained Glass, Cambridge, 1997

[4] Rushforth, G McN, Medieval Christian Imagery, Oxford, 1936

[5] Digital Image Archive of Medieval Music (DIAMM) at www.diamm.ac.uk, last accessed 01/03/2006.

Author Biography

Martyn Jessop received his BSc from the University of London (1979). He holds the MBCS and CITP awards of the British Computer society and is a fellow of the Royal Photographic Society. He worked on university research projects in the sciences for fifteen years before joining the Centre for Computing in the Humanities at King's College London as a project manager in 2000. He has subsequently worked on a wide variety of digital humanities research projects.

Digital Project Preservation Plan

A Guide for Preserving Digital Humanities/Scholarship Projects

Version 2.0

April 2019

dp^3

Digital Project Preservation Plan: A Guide for Preserving Digital Humanities / Scholarship Projects

Publication Notes [Project Profile]

TITLE

Digital Project Preservation Plan: A Guide to Preserving Digital Humanities/Scholarship Projects

AUTHOR
A. Miller, Assistant Professor, Digital Scholarship Initiatives, Walker Library, Middle Tennessee State University

REVIEWERS
Susan W. Knowles
Digital Humanities Research Fellow, Center for Historic Preservation, Middle Tennessee State University

Molly Taylor-Poleskey
Assistant Professor, Department of History, Middle Tennessee State University

COPYEDITOR
Amy Harris-Aber
Department of English, Middle Tennessee State University

AFFILATION
An OER companion to "Digital Humanities Preservation: A Conversation for Developing Sustainable Digital Projects" by A. Miller and Molly Taylor-Poleskey, In *Transformative Digital Humanities: Challenges and Opportunities*, Eds. M. Balkun & M. Deyrup, Routledge. Forthcoming 2019-2020.

IDENTIFIER
https://jewlscholar.mtsu.edu/handle/mtsu/5761

KEY WORDS
digital humanities, digital scholarship, digital preservation, preservation plan

DATE OF PUBLICATION
Version 1: February 28, 2019 Version 2: April 10, 2019

SUGGESTED CITATION
Miller, A. (2019). Digital Project Preservation Plan: A Guide for Preserving Digital Humanities/Scholarship Projects. Available at https://jewlscholar.mtsu.edu/handle/mtsu/5761

Table of Contents

Summary

A Digital Project Preservation Plan is designed to help with organizing preservation efforts for digital projects. Initially drafted as a companion guide of best methods for preserving digital scholarship or digital humanities projects, it can also be applied to digital projects outside the humanities. This preservation plan will benefit those digital humanities (DH) project creators who need guidance on how to start a digital project with preservation in mind. Although the DH community has shared resources and case studies, the examples available tend to focus on DH development, and less on DH preservation. These resources are also located in disparate locations, making it difficult to synthesize best practices. The Digital Project Preservation Plan is a singular guide, focusing on DH preservation, as a starting point with references to more resources and related DH practices. This is a working document, available to practitioners in whole or part; ideally, it will be used in the early stages of project planning and consulted and revised regularly. A successful project will design and build preservation infrastructure from the beginning as a collaborative effort. As priorities, methods and technologies change, the preservation plan will need to be updated and modified accordingly.

The Digital Project Preservation Plan will be made freely available as an open educational resource (OER) on the Web as follows:

(1) A full plan (guide/instructions and appendices/templates) copy will be available in PDF;
(2) The individual appendices will also be made available as separate documents for ease of access, and will be available as fillable PDFs (if downloaded).

The above contents will be available at https://jewlscholar.mtsu.edu/handle/mtsu/5761

Project Owner:

Preservation Plan

This preservation plan addresses the purpose, objectives, and expected outcomes of the digital project. This is a working document; ideally developed in the early stages of project planning, and consulted regularly, then modified throughout the development lifecycle. A successful project will design and build in the preservation infrastructure from the beginning, which is a collaborative effort. As priorities, methods, and technologies change, the preservation plan will need to be updated and modified.

Objectives

A Digital Project Preservation Plan is designed to help with organizing preservation efforts. The time and level of detail put in this plan depends on preservation plan users. The more the preservation plan is consulted, the more likely project owners are to have successful management of the digital project inventory and preservation. This Preservation Plan is most beneficial to those DH project creators who need more guidance on how to start a digital project with preservation in mind. Although the DH community has shared resources and case studies, the examples available tend to focus on DH development, and less on DH preservation. These resources are also located in disparate locations, making it hard to identify a holistic set of steps. The Preservation Plan instead is a singular guide, which focuses on DH preservation as a minimal starter approach with references to more resources and related DH practices.

The Digital Project Preservation Plan (DP3) is only offered as guidance for digital humanities and digital scholarship projects; it is not intended as the best method or only method. For example, project creators should keep in mind the following:

- *Preservation Plan should be used in consultation with other preservation policies and frameworks standardized at institutional or industry levels.*[1]
- *Preservation Plan will need to be revised as new approaches and technologies are introduced that will help make digital projects sustainable and available for the long-term.*
- *Preservation Plan is merely "a" (singular) optional plan, it is not "the only" plan.*
- *The Digital Project Preservation Plan--Full Plan: Guide and Templates offers guidance and resources, including additional templates in the Appendix. Users of Preservation Plan can pick and choose which to apply to their digital projects.*

Name of Project:

Initial Preservation Plan Date:

Last Modification Date:

[1] Examples are included in the Resources section of this document. Specific categories include general resources, digital education (including methods, tools, sandboxes), project planning, organizations focusing on digital preservation, preservation practices, developing web projects and universal design, and web archiving.

Project Owner:

How to Use this Guide

Overview

This guide consists of tools for creating a Digital Project Preservation Plan including a detailed description of items to consider (a Project Charter, a Digital File Inventory, and Additional Considerations for infrastructure setup), a plan checklist (Preservation Plan-A Summary and Checklist), and usable preservation plan templates.

The appendices serve as the preservation plan templates, and include:

Project Charter
Digital File Inventory
Project Profile
Collaborators Web Publishing Agreement
Universal Design Checklist
Preservation Guidance Checklist

The final section is a glossary of basic terms used throughout this Digital Project Preservation Plan guide.

Instructions

Use this document as a whole or make selections among its individual components. For example, if you already have produced a Project Charter, skip ahead to the Digital File Inventory. Or, if you do not need instructional guides altogether, skip ahead to the Appendices for the fillable PDF templates.

Since preservation standards can change, and storage media formats are vulnerable to risk and obsolescence, a Preservation Plan should be evaluated every 2-5 years. Make note of the date and revisions for each re-evaluation.

[end of How to Use this Guide]

Project Owner:

Project Charter

A project charter is a set of guidelines developed at the beginning of a project. These guidelines are preferably written by the project team, but at the very least, they are drafted by the *project owner* and given to all team members for review at the beginning of the project. This gives the project team the opportunity to look at the entirety of the project. This includes the goals, objectives, limitations, timeline and deliverables, and yet other considerations. It also gives each team member a chance to review their individual responsibilities as described in the project charter. Numerous disciplines use project charters as a project management tool. Some have been specifically developed for use in libraries or cultural heritage institutions.

UCLA Library Special Collections has an extensive array of project management templates, including a Project Charter template. These templates are online as a part of the UCLA Library's Digital Project Toolkit.[2] A template like this one is recommended to use if a project charter would benefit your project. Alternatively, you could draft a text document, then include the sections of a traditional project charter that are most appropriate for your project. The content and design of a project charter is specific to each project. Some suggested sections could include:

Project Description

Provide a brief description of the digital project.

Scope / Out of Scope

List the type of materials utilized for this project. This includes primary sources, digital platforms, and locations (physical or virtual) that will be associated with this project. Describe features, services, and products, if any, that will result from this project. Please include local, regional, or national affiliations, collaborators, target audiences, and functional requirements. Below that, define the boundaries of the project and what will not be included (if any).

Deliverables

Define the intended end product(s) of the digital project.

For any of the deliverables, you may choose to describe any software that will be used in creating the end product(s). For example, a scholarly publication or presentation reporting the state of the project's topic; the digital method of spatial analysis could use the tool/software of QGIS, Story Maps, StorymapJS, or CartoDB; or the digital method of text analysis could use the tool/software Voyant or R. In-depth responses are not necessary; simply listing the method and software is enough.

Intended Platform for Delivery

With the above deliverables in mind, select which platforms you intend to use for dissemination.

Project Goals

Describe the overall goals of the project, including how it will be utilized. This could include the general public, specific group, or an academic institution, etc.

[2] Available at http://library.ucla.edu/special-collections/programs-projects/digital-projects-special-collections

Potential Risks

Several items may prevent a project from moving forward (funding, staffing, timeline, etc). List any limitations or concerns for the project, this may include collaborators, stakeholders, intellectual property concerns, and ethics or privacy issues (especially when working with students). If possible, list approaches to help mitigate these risks.

Success Factors

Describe how the project will be measured for success. For example: use of web analytics, alt-metrics, scholarly citations, conference presentations, or other means. See the Metrics Toolkit[3] for more examples.

Roles and Responsibilities

Describe the roles and responsibilities of each collaborator, including but not limited to the project owner, sponsor, and team members (students or professionals). This might be achieved by creating, for example, a chart.

Project Timeline

List project milestones chronologically. Describe any factors that may affect the timeline, including funding, grant mandates, student availability, resource availability, conference travel, etc.

Project Requirements

Identify what this project needs that is not already understood or in use. For example: outsourcing digitization, transcription, or website development; required team member training; tools or skills needed for project completion; item purchases or travel/fieldwork.

Funding

Describe any grant-funded objectives or mandate, as well as timeframes for the grant award.

Collaborators Agreement

A set of parameters or code of conduct that governs the project, with acknowledgement from the collaborators.

Note: Best practices for project management include the use of a collaborators agreement and a project charter. The above section is a sample of a limited Project Charter. For a collaborators agreement, see the Media Commons Press version of Collaborators Bill of Rights,[4] the UCLA HUMTECH Student Collaborator's Bill of Rights[5] or the UCLA Library's Collaborators Agreement.[6] It is critical to provide proper acknowledgement to all project participants, and a collaborators agreement can assist with this job, in addition to keeping track of individual responsibilities. Additionally, the collaborators agreement should acknowledge the right to privacy when developing public-facing scholarship, especially if the project is for course credit. Students should have

[3] Metrics Toolkit is available at http://www.metrics-toolkit.org/

[4] http://mcpress.media-commons.org/offthetracks/part-one-models-for-collaboration-career-paths-acquiring-institutional-support-and-transformation-in-the-field/a-collaboration/collaborators%E2%80%99-bill-of-rights

[5] https://humtech.ucla.edu/news/a-student-collaborators-bill-of-rights/

[6] https://www.library.ucla.edu/sites/default/files/Template_CollaboratorsAgreement_1.pdf

the right to alternative assignments or identification anonymity in any published project. This is important to note for the Preservation Plan, as certain elements may be placed on public-facing platforms.

For a fillable template of a limited Project Charter with the suggested sections above, see Appendix A.

[end of Project Charter section]

Digital File Inventory

This Digital File Inventory is perhaps best suited for use in a spreadsheet, especially if there are many files associated with the project. Keeping an inventory of files is the critical part. However a project owner chooses to proceed, keep file descriptions consistent. See Appendix B for a one-page PDF fillable template with these recommendations. There are a few different ways the Appendix on Digital File Inventory (DFI) could be used:

1. For use in projects with a few files. Instead of using a spreadsheet for inventory of all files associated with the digital project, fill out the DFI template for each associated file. Recommended for project with less than 20 files.
2. For projects with large amounts of files, the DFI template could become time consuming and tedious to document each file in the manner describe above in point (1). That is why the use of a spreadsheet for inventory is suggested for a project with many files. For these larger projects, perhaps use the DFI template for the final published version(s) of the digital project; or for only a selection of the items from the spreadsheet inventory. For example, just the final versions of a file or a single set of items to be preserved.
3. As a guide to what type of fields to include in the spreadsheet for your own inventory purposes.

README File

Along with an inventory of files associated with the digital project, creating a README file is also recommended. The readme file is a term taken from computer science, and it is a form of documentation for software (Wikipedia, 2018) that describes the files in a directory or information that is beneficial to understanding why the software is valuable and how it can be used. There are examples of how to write this type of traditional readme file online[7] that may be helpful to understanding the purpose. For DH projects, a readme file might document how each file is connected to the digital project altogether (in addition to using it for the traditional purposes if your DH project produces its own code). A DH readme file puts the development process into words, helping to ensure files are correctly interpreted by the creators today or others who may pick up the project in the future (Cornell University, n.d.). The readme file can also be used for documenting decisions made in the project, such as methods that worked and those that did not; the files are also a productive way to describe the rational behind project decisions. Rockwell (2014) describes this DH readme file as a type of "deposit package." Preserving the development process of a digital project with a readme file is primarily for internal purposes. However, certain elements may be helpful to project end-users, and as a result a selection may also be used for external purposes.

The details of a readme file depend on the type of digital project being created. The purpose is to document how all the files being preserved relate or differentiate from one another, and the processes (technical, financial or administrative) involved in the project's actualization. Although there is no set template for a DH readme file, it is recommended to review the Guide to Writing "readme" Style Metadata[8] and DH styled readme file examples such as the ones for Globalization and Autonomy Online Compendium[9] or The Poetess Archive.[10] The latter uses the concept of a readme file to describe the terms, database, decisions, workflow,

[7] https://www.wikihow.com/Write-a-Read-Me
[8] https://data.research.cornell.edu/content/readme#introductory
[9] https://doi.org/10.7939/R3TH8BN81
[10] http://www.poetessarchive.org/about/index.html

and resources used in the creation of this digital archive, but instead of referring to it as a readme file, this information is listed as a narrative summary on the About page of its website. Therefore, DH projects may well be applying this readme file method without realizing it.

Digital File Inventory Recommendations

Regardless of the system for inventory, there are several suggested items to document that will help preserve your project and its individual components. These components will vary and the use of a category classification is recommended for organization. For example, Images-Original and Images-Edited; alternatively, if the project is grant funded, the project owner may choose to have a category specific to Grant or Travel.

A Digital File Inventory Template is available in Appendix B. Again, the use of a spreadsheet may be better for projects that contain many files. Below are suggested fields to include for each file in the inventory:

Title of Document

A purposeful name for the document; how you interpret it the file. For example, "Interview Notes," "Travel Arrangements," or "Final Network Analysis Visualization."

File Name with Extension (original)

The actual name of the file including the extension (.txt, .pdf, .mp4, etc) and the file location (where it is saved: include the hard drive location and file structure/directory location and/or link to cloud storage if applicable). Best practices recommend saving a file in three separate locations. For example (1) local machine, (2) external hard drive, and (3) cloud or additional off-site external hard drive. Third-party managed cloud storage is not recommended for sensitive data (such as student data).

Example: (1) InterviewNotes.txt saved on my computer>Documents>NetworkAnalysisProject>Interview Notes; (2) same file saved on external harddrive #1 located in office of Dr. Smith; (3) same file saved in institutional Dropbox account accessible at https://dropbox.com/work/forexamplepurposesonly.txt

Category

A classification used to group similar items together. This category could correspond to folders you create for organizing the files. For example, categories of Images, Essays, Data, Code, Travel, Administrative, and Grant. The categories will vary by project.

Creator

The person that drafted the file, or the first version of the file. If multiple people wrote or edited the document, that could also be indicated here.

Date Created

Original date the file was created.

Date Last Modified

Over time the file may change, with new versions, or a new name. Document the date of such changes. This is helpful in case you need to revisit drafts. For code management, use a code repository such as GitHub.[11]

[11] https://github.com

Type or Format

The form of the data file; this may include text, numeric, audiovisual, models/computer code, etc.

Preservation Copies

The preservation copy may be different from the original file copy which is described above in the "File Name with Extension" field. It is best to preserve a file in both the original format and an alternative or open format (see note below). For instance, tabular data created in Excel should be saved in Excel, but also in an open format such as CSV. See NDSA (2013) for more information on the different levels of digital preservation. A preservation copy is like the "hybrid archiving" scenario described in the 2014 NDSA report[12] on the PDF/A- 3 file format. Best practices for data archiving include considerations of location (on-site, off-site), file formats, responsibility, accessibility, frequency (hourly, daily, weekly, monthly, annually), retention (months, years), confidential/sensitive data security, intellectual property concerns, and testing the archive plan. When developing the first stages of a DH project, focus on preserving what is possible in that moment.

Preservation Copies Saved Location(s)

Document where the preservation copies are saved. This can be in the same system as the original file or a different file altogether. This can also include archives or repositories with permanent identifiers (URLS) such as institutional repositories, subject repositories, code repositories, and internal servers. Record the *preservation copy* last saved date as preservation copies may also have different versions and modifications.

Association with an Approvals Plan or Sensitive Data

Make a note if the file has confidential information such as student data, intellectual property concerns, or has an agreement such as Institutional Review Board (IRB) approvals, contract agreements and Memorandum of Understanding (MOU) associated with collaborators, vendors, or third-parties. Indicate which files of your project fit into this segment and document it in the Digital File Inventory list and the internal readme file.

Note: Closed vs Open File Formats

Different file formats are one of the most common issues when dealing with digital preservation. Many formats are software specific and will not be easy to maintain for the long-term. For long-term preservation, it is best to keep data associated with a DH project in an uncompressed non-proprietary file that can be opened using a variety of software. If using proprietary formats, be sure to document the software package, version, vendor and/or native platform.[13]

For example, spreadsheets are frequently used in digital projects. Although they are particularly helpful with organization, formats change frequently; even the software packages (Excel, Sheets) are not guaranteed to last forever.[14] In an instance where a spreadsheet was created in Excel, save that Excel copy. Additionally, save the spreadsheet file as a Comma-Separated-Value (CSV), and make both the Excel and CSV copies part of the Preservation Plan. The CSV files are preferred for long-term preservation and are considered an open format.

[12] Available at https://ndsa.org/documents/NDSA_PDF_A3_report_final022014.pdf

[13] For more information on all formats, see the content categories section of the Sustainability of Digital Formats: Planning for Library of Congress Collections Last updated 3/10/2017 at https://www.loc.gov/preservation/digital/formats/index.shtml

[14] DataONE. (n.d.). Preserve information: keep your raw data raw. https://www.dataone.org/best-practices/preserve-information-keep-your-raw-data-raw

Examples of closed or proprietary formats would be products by Microsoft (Excel), Apple (Pages), Adobe (Flash), Google (Sheets), ESRI (StoryMaps), and so forth. Whenever possible, use open, uncompressed, non-proprietary formats for production and preservation files. Examples of open formats are in the table below.

File Type	*Open Format Suggestion*	**File Type**	*Open Format Suggestion*
Text	*ODF, PDF, TXT, HTML*	**Database**	*DBF, XML, Base*
Language	*HTML, XML, TEI*	**Tabular data**	*CSV*
Geospatial	*GML*	**Presentations**	*PDF*
Images	*TIFF, PNG, JPG*	**Video/Audio**	*MP4, WAV, AIFF*
Containers	*ZIP, TAR*		

For a fillable template of a Digital File Inventory with the suggested sections above, see Appendix B. The headings in the PDF template can also be placed as fields in a spreadsheet to serve a similar inventory list purpose.

[end of Digital File Inventory section]

Additional Considerations

In addition to the project charter and digital file inventory recommendations from the preceding sections, there are other aspects to consider for digital project preservation. The below sample were selected because of their importance for showcasing, archiving, and indexing digital projects, as well as the imperative need for educating users as updates to digital education pedagogy and research developments emerge.

Collaboration

Collaborating on best practices with colleagues from various disciplines and roles. Experimenting with resources and documenting experiences with case studies and use of screencasts, screenshots, interviews, and prototypes.

Data and Digital Literacy Education

Primarily a focus on educating students and scholars developing DH projects for the classroom. Topics could include how the Web operates, basic HTML coding, and importance of file structures, file versioning, and file storage, and final backup methods and best practices. Libraries and universities offer workshops on these topics, hoping to fill the gap on digital competencies, while larger initiatives are looking to reshape the data curriculum (Nelson, 2017). The Data Information Literacy (DIL) project is an example of a multi-university initiative on such a curriculum, which produced a DIL Guide on how to develop a data curriculum.[15]

Server Space

It is important to identify the needs of a project early in the planning stages, including where a digital project will live virtually. This means server space, domains and web hosting. It is critical that server space is addressed early and the parameters of its use are agreed upon. For example, if the project is based out of a university, request server space from the institution in advance of project production. Discuss the terms of institutional space for hosting the project including access and duration, and use a collaborators agreement to document the responsibilities of those involved. Other server space and hosting options include purchasing third-party managed web hosting,[16] or use of flexible hosting plans from code-based repositories such as GitHub.[17] Regardless of your server and hosting choice for a digital project, understand the terms for using that choice and request it early.

Institutional Sandboxes

A sandbox is a common term for research and development. For DH, a sandbox is a place to access and play with technology, in both physical or virtual environments. Access to a variety of software and hardware can help in the early stages of a project, helping to alleviate software installation challenges or help determining which software best suits a project. A physical sandbox location[18] can serve as an interdisciplinary computing facility that maintains a variety of software and hardware, and the development space needed for digital projects. Virtual sandbox environments allow its users to create and share educational projects, including

[15] Available at http://www.datainfolit.org

[16] One example is Reclaim Hosting https://reclaimhosting.com

[17] GitHub plans range from individual and team-based; and from $0 to $7 or more a month. https://github.com/pricing

[18] For a physical sandbox space example, see: Institute for Digital Research and Education (IDRE) at UCLA http://idre.ucla.edu/technology-sandbox; or the Teaching and Learning Commons at West Virginia University http://tlcommons.wvu.edu/sandbox.

media-rich websites. Commons in a Box[19] and Academic Computing Environments[20] are examples of digital sandboxes in use at CUNY and Fordham University (respectively) that enable users to experiment with university provided or open-source software such as Wordpress, Omeka, and more. Similarly, some institutions provide Web space and domains upon request, giving practitioners the ability to manager their own institutional Web space for projects.[21] If electing to use an institutional Web space, keep in mind the terms of using that space (access to it for the long-term).

Repository Options for Archiving

A repository provides digital preservation through documentation of the digital files collected by technical and administrative metadata on the file types, sizes, dates of file additions and the name of the person depositing the data. Common repository options include institutional repository , subject repositories , open data repositories, or other forms of free or fee-based repositories such as GitHub (public repositories are free; limited free private repositories; fee-based options), Dropbox, and Google Drive. Repositories (and certain websites) use persistent identifiers (URL, URI, or handle) that allow for access to and discovery of a project. Metadata is important in a repository as it is data describing the file deposited. This in turn helps the file to get index and discovered. Some code repositories, such as Github, do no use clear metadata for author or affiliation identifiers. In this case, consider using CodeMeta , a new standard format for software metadata. By including the corresponding JSON file in the Github page, the CodeMeta will help cite the file with the author, keyword, and other metadata associated with the project profile.

Archiving Dynamic Objects and Executables

A 2016 NSDA survey on web archiving practices in the United States[25] reported several external service providers that carry out Web archiving: Heritrix, HTTrack, Webrecorder, Web Curator Tool, and Wget, among others. Although some of these providers are able to capture dynamic content, such as Webrecorder, it only captures the web pages you visit, it does not automatically obtain non-visited pages of a website nor any content from links on those pages.[26] With an online game or tutorial (dynamic content) based on user choices, this means it is possible not all pages are captured; and it is also time consuming to go through all possible combinations of paths a user may choose in a game. Additionally, whether archiving static or dynamic content, built upon previously established standards and practices by leading organizations. For example, the National Digital Stewardship Alliance (NDSA) created the Levels of Digital Preservation, a rubric to help organizations manage preservation risks of digital materials (NDSA, 2013). Beyond websites and data visualizations, dynamic objects also include 3D objects, photogrammetry, and augmented, and virtual reality experiences, which should share a common agenda of digital curation (CLIR, 2019).

---

[19] https://commonsinabox.org

[20] https://www.fordham.edu/info/25009/faculty_technology_services/1426/academic_computing_environment_ace/1

[21] See example at the University of Minnesota http://dash.um.edu/dash-domains or Humtech at UCLA http://humtech.ucla.edu/support.

[22] Unique to each institution such as a university by preserving the scholarship produced at the institutions. See one example, https://jewlscholar.mtsu.edu

[23] See the "Using Other Repositories" section at http://www.lib.berkeley.edu/scholarly-communication/publishing/open-access-publishing/deposit-preserve

[24] See http://oad.simmons.edu/oadwiki/Data_repositories or http://www.share-research.org/

[25] https://codemeta.org/github.io

[26] NSDA Web Archiving Survey Working Group. (2017). Web Archiving in the United States: A 2016 Survey. Report available at https://ndsa.org/documents/WebArchivingintheUnitedStates_A2016Survey.pdf

[27] https://guide.webrecorder.io/

Project Profile

A project profile is a descriptive summary of the project once it is complete and made available to the public. Components of the project profile may include the project title, staff, keywords, suggested citation, and software used. As oppose to the project charter, which is used in the planning stages of a project, the project profile is used to help highlight and index the project in its final form (such as dissemination in a publication[27] or on a website). Some projects list the Project Profile elements in an "About" page on a website. An example of a good "About" page is featured by Northeastern University Libraries' *Our Marathon* project.[28] The Project Profile can also include post-project development items, including references of the project in the media, scholarship produced with the project (peer reviewed articles, case studies, use in the classroom, etc.), team reflections (blog post or anecdotal feedback from collaborators who worked on the project), and additional/related projects. Project Profiles also provide a medium for announcing credit and reuse rights such as Creative Commons or other licensing. A Project Profile is a way to catalog key elements of the project that will result in better indexing (with tags or metadata) and discoverability in search engines.[29] See Appendix C for a fillable PDF template.

Collaborators Web Publishing Agreement

The Collaborators Web Publishing Agreement allows each project collaborator to grant or deny permission to have their name affiliated with the published project online, as described in the Project Profile. The agreement signals name-project affiliation in public-facing dissemination venues, including but not limited to the project website, press releases, news articles, or publications. *Note: this is different from a collaborators agreement used in project planning.* See Appendix D for a fillable PDF template.

Accessibility and Universal Design

Ng (2017) explains how universal design (UD) benefits everyone, not just those with disabilities. In her practical guide, Ng covers writing for the web, proper usage of links, audio and visual content including embedding third-party content, and several other helpful design considerations. Accessibility and universal design application for digital products, including documents, presentations, spreadsheets, video, audio, software and websites will help enable the preservation and wider use of such products for the long-term[30].

DH projects often use images as part of their design and dissemination on websites. These images should include captions and alternative tags (abbreviated "alt" tags). These alt tags describe the image and its functions (W3C, 2017); and tutorials[31] for creating alt tags for different types of images are available online to help demonstrate the process by which to provide appropriate descriptions based on image type. Additional considerations should include the accessibility of non-HTML content such as embedded or downloadable PDF or text documents, spreadsheets, charts, graphs, and presentations; this includes the use of captions and transcripts for videos or audio (ARL, 2013) used in digital projects online. Other universal design considerations should encompass visual design with adequate spacing, mobile platform compatibility, use of headings, contrast with colors for text, and ease in locating and navigation. The objective is to prioritize simple design over complex. See Appendix E: Universal Design Checklist for a fillable PDF template.

[27] Publications, such as book, include this profile information in the publication notes section, which is typically after the title page. See *page i* of this Digital Project Preservation Plan as an example.

[28] https://marathon.library.northeastern.edu/home/about

[29] See the project profiles of the DH Lab at Yale University http://dhlab.yale.edu/projects.html

[30] See Section 508 to learn how to create and test for accessible digital products http://www.section508.gov/create [31] https://www.w3.org/WAI/tutorials/images/

DH Preservation Consortium

A key to preservation is understanding that failure is part of the preservation process. Discussing, analyzing and publishing on failures, data loss, or errors will benefit DH practitioners and the field at large (Dearborn & Meister, 2017). These discussions take place at local and national levels[32] but a preservation-centric subject repository for DH projects has yet to be developed although some strides in this area have been made.

Monitoring Developments

The Socio-Technical Sustainability Roadmap, an NEH (2016-2018) funded curriculum, is starting to see results from the project's workshop participants. Although this project was initially created to address sustainability of Medieval Art, the curriculum is transferable to other digital projects.

IIPC (2018) curated a list of widely-used web archiving tools and software, indicating which are stable, in development or deprecated. Experiments sometimes fail. Even organizations whose membership or missions change can also sunset due to preservation challenges. As of 2018, the Digital Preservation Network (DPN) decided to close operations due to insufficient participation and membership, which was essential to the community-based preservation model it practiced (DPN, 2019). Although DPN will cease to exist after March 2019, the resources it has developed in its tenure will remain accessible by preserving its web content in the Wayback Machine.[33] Organizations like the IIPC, NSDA, CNI, DLF, DCC, Educopia, LOC, and universities[34] that experiment with software and dynamic object archiving are entities to look to for future trends and developments in digital preservation. A 2018 survey conducted by ITHAKA S+R identified challenges and gaps in digital preservation needs (Rieger, 2018). The survey participants, 21 leaders with digital preservation leadership roles, indicated the state of confusion of preservation services and strategies as a major challenge. Ambiguity of responsibilities within organizations, storage costs, vulnerability of content, and insufficient tools for analysis and assessment of digital preservation are other challenges (Rieger, 2018). These challenges exist in addition to the unresolved solutions for dynamic object preservation which remains in a state of flux.

Emulation-as-a-Service (EaaS) model is currently being developed (2018-2021) by the Scaling Emulation and Software Preservation Infrastructure (EaaSI) program led by Yale University. The program is focusing on how technology and services can enable broader access to preserved software and digital objects, emulating software across a broach range of disciplines (Software Preservation Network, n.d. a). Other digital preservation research in-progress (2017-2020) is a project geared toward libraries, historical societies, and museums: Fostering Community of Practice: Software Preservation and Emulation Experts in Libraries and Archives (FCoP). FCoP is led by CalPoly University, with the goal of broadening participation in long-term preservation practices by establishing a community of practice and experimenting with web-based emulation sandboxes (Software Preservation Network, n.d. b).

[end of Additional Considerations section]

[32] Examples include Humanities Commons (HCommons) https://hcommons.org and HASTAC https://www.hastac.org
[33] Accessible at https://web.archive.org/web/*/dpn.org
[34] Universities are among the organizations experimenting with web archiving initiatives https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives

Project Owner:

Preservation Plan- A Summary and Checklist

DH practitioners and their institutions may have varying levels of infrastructure (human, technical, or financial). To help facilitate a checklist methodology, the guidance suggested in the preceding sections of this plan have been broken down into two levels of infrastructure: Minimal and Expanded. The following checklists offer suggestions, serving as a guide no matter your existing level of DH or preservation infrastructure. It is suggested that you choose which are applicable to your specific project, not necessarily to comply with each goal. See Appendix F for a fillable PDF version of the two checklists.

Minimal Infrastructure (or Short-term) Preservation Goals

- Build a list of collaborators of varying backgrounds, skills, and strengths.
- Use a Project Profile (description including human, financial and technical components of the project on the website and/or repository).
- Whenever possible, use open standards and formats in developing DH projects.
- Develop a Digital File Inventory (spreadsheet or other organized system of all files associated with the project).
- Document and deposit the process, decisions, and code (if applicable) in a readme file or narrative summary for **internal** purposes, preserving the project's reproducibility.
- Determine which elements (from point above) should be preserved and available for **external** purposes or public audiences, and place on the project's website and/or archival repository.
- Collect a Web Publishing Agreement from each collaborator.
- Use a web-archiving tool or repository to preserve a project's current static and dynamic functionality and project documentation (external).
- Use rights statements and/or licensing for attributing credit or reuse permissions for projects, research, code, and metadata developed. This is one way for a project to have a longer life, by giving permission for its use, including attribution and/or remixing allowances.
- Create suggested citations of the intellectual property for discovery, crediting and sustaining use. This can help encourage citation and lessen confusion on how to cite/attribute credit.
- Export and save versions (servers, code, development notes, and native file formats) in a structured organization system; and save copies in different locations.
- Encourage data and digital literacy education.

The following checklist offers suggestions for projects with an expanded infrastructure:

Expanded Infrastructure (or Long-term) Preservation Goals

- All the above minimal infrastructure goals above.
- Consider current migration, emulation, or other preservation capturing methods for archiving dynamic objects and executables.
- Reflect on the intended use of the project from the user's perspective and address user experience design questions and methods.

- Creation of an institutional space for student and faculty project development for both testing and production. The servers used for the sandbox environment should be backed up just as would a digital file.
- Employ accessibility standards for text, images, and multimedia elements in projects including but not limited to transcriptions, closed captioning, alt tags, site maps, and browse and search options.
- Use existing digitization standards, metadata schemas, encoding practices, and code/version management.
- Monitor research and development in preservation practices in various disciplines and create a preservation practices rubric to follow for consistency among projects.
- Collaborate with other campus or community units and experts on best practices and experiment with available resources.
- Develop a consortium of like-minded collaborators across departments and/or institutions for potential development of a DH preservation repository/database.

[end of Preservation Plan-A Summary and Checklist section]

References / Plan Resources

General Resources –Digital Humanities/ Digital Scholarship

Association of Research Libraries (ARL). (2011). SPEC Kit 326 Digital Humanities. https://www.arl.org/storage/documents/publications/spec-326-web.pdf

DH+lib. (n.d.). Readings. [DH reading list]. https://acrl.ala.org/dh/dh101/readings

EDUCAUSE. (2017). Building Capacity for Digital Humanities: A Framework for Institutional Planning. ECAR/CNI Working Group Paper. May.

Goldenberg-Hart, Diane. (2016). Planning a Digital Scholarship Center 2016. Report of a CNI-ARL Workshop. https://www.cni.org/wp-content/uploads/2016/08/report-DSCW16.pdf

Lewis, Vivian, Lisa Spiro, Xuemao Wang, and Jon E. Cawthorne. (2015). Building Expertise to Support Digital Scholarship: A Global Perspective. https://www.clir.org/pubs/reports/pub168

Lippincott, Joan, K. and Diane Goldenberg-Hart. (2016). Digital Scholarship Centers: Trends and Good Practice. CNI Workshop Report. https://www.cni.org/wp-content/uploads/2014/11/CNI-Digitial-Schol.-Centers-report-2014.web_.pdf

Digital Education, Methods, Tools, and Sandboxes

Brennan, Sheila A., Megan Brett, Sharon M. Leon. “Doing Digital Scholarship.” Digital Culture Program. *Social Science Research Council Labs*. https://labs.ssrc.org/dds/

DMPTool. (2019). Data Management General Guidance. University of California Curation Center. https://dmp.cdlib.org/general_guidance

Fordham University. (n.d). Fordham IT. Accessed February 25, 2019. https://www.fordham.edu/info/25009/faculty_technology_services/1426/academic_computing_environment_ace/1

Johnston, Ben. (2019). Digital Tools for Course-Based Digital Projects. The McGraw Center, Princeton University. January 9. [list of tools]. https://mcgrawect.princeton.edu/digital-tools-for-course-based-digital-projects

Khan Academy. (2019). http://khanacademy.org/computing

Metrics Toolkit. (n.d.). http://www.metrics-toolkit.org/

Nelson, Megam R. Snapp. “A Pilot Competency Matrix for Data Management Skills: A Step Toward the Development of Systematic Data Information Literacy Programs.” *Journal of eScience Librarianship*, vol. 6, no. 1, e1096, 2017.

Omeka. (n.d.). https://omeka.org

Programming Historian. https://programminghistorian.org

Scalar. (n.d.). Alliance for Networking Visual Culture. http://scalar.me/anvc/scalar

Voyant. (2019). http://voyant-tools.org

Organizations Focusing on Digital Preservation

centerNet. http://dhcenternet.org

Coalition for Networked Information (CNI). https://www.cni.org

Council on Library and Information Resources (CLIR). https://www.clir.org

Digital Curation Centre (DCC). http://www.dcc.ac.uk

Digital Library Federation (DLF). https://www.diglib.org

Educopia Institute. Research. https://educopia.org/research

Library of Congress. Digital Preservation. http://www.digitalpreservation.gov/index.php

National Digital Stewardship Alliance (NDSA). An initiative of the National Digital Information Infrastructure and Preservation Program. http://www.digitalpreservation.gov/ndsa/NDSAtoDLF.html

National Information Standards Organization (NISO). (2019). https://www.niso.org

SHARE. http://www.share-research.org

Project Planning

Burress, Theresa, and Chelcie Juliet Rowell. (2017). Project Management for Digital Projects with Collaborators Beyond the Library. College & Undergraduate Libraries, 24(2-3): 300-321

Cornell University. (n.d.). Guide to Writing a “readme” Style Metadata. Accessed February 18, 2019. https://data.research.cornell.edu/content/readme#introductory

National Information Standards Organization (NISO). (2007). A Framework of Guidance for Building Good Digital Collections. Available at https://www.niso.org/sites/default/files/2017-08/framework3.pdf

Rockwell, G. et al. (2014). Burying Dead Projects: Depositing the Globalization Compendium. *Digital Humanities Quarterly*, 8(2).

UCLA Library Special Collections. (2015). Digital Project Toolkit. Available at http://library.ucla.edu/special-collections/programs-projects/digital-projects-special-collections

UCSC University Library. (n.d.). Lifecycle of a Digital Project. Available at https://guides.library.ucsc.edu/digitalproject/projectplanning

Wikipedia. (2018). README. Last updated December 23, 2018. https://en.wikipedia.org/wiki/README

Developing Web Projects and Universal Design

Association of Research Libraries (ARL). (2013). Web Accessibility Toolkit. http://accessibility.arl.org

US General Services Administration (n.d.). Create Accessible Digital Products. Accessed February 19, 2019. http://www.section508.gov/create

Ng, C. (2017). A Practical Guide to Improving Web Accessibility. Weave: Journal of Library User Experience, 1(7). https://quod.lib.umich.edu/w/weave/12535642.0001.701?view=text;rgn=main

OER Accessibility Toolkit. (n.d.). University of British Columbia. Available at https://open.ubc.ca/teach/oer-accessibility-toolkit

Section 508. Available at http://www.section508.gov/create

W3C. (2017). Image Concepts. Web Accessibility Tutorials. Accessed February 19, 2019. https://www.w3.org/WAI/tutorials/images/

W3C. (2018). Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21

Wayback Machine. (2014). Internet Archive. Available at http://archive.org/web

WebAIM: Introduction to Web Accessibility. People with Disabilities on the Web. Available at https://webaim.org/intro/#people

Webrecorder. (n.d.). Rhizome. Available at https://webrecorder.io

Preservation Practices and Polices

Boston University Libraries. (n.d.). BU Libraries Digital Preservation Policy. http://www.bu.edu/dioa/openbu/boston-university-libraries-digital-preservation-policy

Dartmouth College. (2011). A Report from the Digital Projects and Infrastructure Group. Dartmouth Digital Library Program. https://www.dartmouth.edu/~library/digital/about/report.html

Dearborn, Carly, and Sam Meister. (2017). Failure as Process: Interrogating Disaster, Loss, and Recovery in Digital Preservation. Educopia Institute. Available at https://educopia.org/wp-content/uploads/2018/04/dearborn_meister_failureasprocess.pdf

Five Colleges Consortium. (2017). Digital Preservation: A Planning Guide for the Five Colleges. Accessed February 18, 2019. https://www.fivecolleges.edu/libraries/digital-preservation/digital-preservation-a-guide-for-the-five-colleges

National Digital Stewardship Alliance (NDSA). (2013). Levels of Digital Preservation. Available at https://ndsa.org/activities/levels-of-digital-preservation

National Information Standards Organization (NISO). (2010). Special Issue: Digital Preservation. Information Standards Quarterly, 22(2). http://www.digitizationguidelines.gov/audio-visual/documents/IP_Fleischhauer_AudioVisual_Reformatting_isqv22no2.pdf

NINCH. (2003). The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials. http://chnm.gmu.edu/digitalhistory/links/pdf/chapter1/1.17.pdf

- General overview but especially chapter 14 on Preservation, pages 198-213

Digital Preservation Handbook. (2019). Digital Preservation Coalition. https://dpconline.org/handbook

Rice University. (2015-last update). Digital Projects for Fondren Library: Documentation. Available at http://bit.ly/u5m0RB

- Sections especially helpful include but not limited to File Management and the following:
- Video: Born digital mov masters edited in Final Cut Pro to access steps (2017). https://tinyurl.com/y9678qbc
- Video and Audio content for IR Deposit (2018).
- Digital Curation Profile Template (2013). http://bit.ly/w3mVaC

Rieger, Oya Y. (2018). The State of Digital Preservation in 2018-A Snapshot of Challenges and Gaps. ITHAKA S+R. October 29. Accessed February 20, 2019. https://sr.ithaka.org/publications/the-state-of-digital-preservation-in-2018

University of Maryland Libraries (2014). Digital Preservation Policy. Available at http://hdl.handle.net/1903/14745

University of Wisconsin (2017). File Naming and Versioning. Available at http://researchdata.wisc.edu/file-naming-and-versioning/

Visual Media Workshop at the University of Pittsburgh. *The Socio-Technical Sustainability Roadmap*. Accessed [January 20, 2019. http://sustainingdh.net.

Web / Software Archiving

Archive-It. (2014). https://www.archive-it.org

Chassanoff, Alenandra et al., (2018). Software Curation in Research Libraries: Practice and Promise. Journal of Librarianship and Scholarly Communication, 6 (General Issue). https://doi.org/10.7710/2162-3309.2239

Council on Library and Information Resources (CLIR). (2019). 3D/VR in the Academic Library: Emerging Practices and trends. https://www.clir.org/wp-content/uploads/sites/6/2019/02/Pub-176.pdf

Code of Best Practices in Fair Use for Software Preservation (2018). Association of Research Libraries. https://www.arl.org/storage/documents/2019.2.28-software-preservation-code-revised.pdf

International Internet Preservation Consortium (IIPC). (2018). Tools & Software. http://netpreserve.org/web-archiving/tools-and-software

Mulligan, Rikk. (2016). SPEC Kit 350 Supporting Digital Scholarship. Association ion of Research Libraries. Available at https://publications.arl.org/Supporting-Digital-Scholarship-SPEC-Kit-350/

National Digital Stewardship Alliance (NSDA). (2012). Web Archiving Survey Report. Available at http://www.digitalpreservation.gov/documents/ndsa_web_archiving_survey_report_2012.pdf

- List of software tools and descriptions for web archiving starts on page 20.

Perma.cc. https://perma.cc/

Software Preservation Network. (n.d. a). About EaaSI. Accessed February 20, 2019. http://www.softwarepreservationnetwork.org/eaasi

Software Preservation Network. (n.d. b). About FCoP. Accessed February 20, 2019. http://www.softwarepreservationnetwork.org/fcop

Stanford University Libaries. (n.d.) Archivability. Available at https://library.stanford.edu/projects/web-archiving/archivability

Wayback Machine. (2014). Internet Archive. http://archive.org/web

Webrecorder. (n.d.). Rhizome. https://webrecorder.io

Zhou, Yongli. (2010). Are Your Digital Documents Web Friendly?: Making Scanned Documents Web Accessible. Information Technology and Libraries, 29(3): 151-160.

Appendices

Overview

There are six appendices identified as A-F. The appendices are included here, part of the Digital Project Preservation Plan as complete document. However, the appendices are also available individually, as separate PDF downloads.

It may be more helpful to select which appendices are applicable to a project and only use the necessary parts of this plan. The individual PDF downloads are fillable versions and can be accessed at https://jewlscholar.mtsu.edu/handle/mtsu/5761.

Project Owner:

Appendix A: Project Charter (Limited)

Project Description

Provide a brief description of the digital project.

Scope / Out of Scope

List the type of materials utilized for this project. This includes primary sources, digital platforms, and locations (physical or virtual) that will be associated with this project. Describe features, services, and products, if any, that will be an outcome of this project. Please include local, regional, or national affiliations, collaborators, target audiences, and functional requirements. Below that, define the boundaries of the project, what specifically, will not be included (if any).

Project Owner:

Deliverables

Define the intended end product(s) of the digital project (check all that apply):

- ☐ White paper
- ☐ Scholarly research publication
- ☐ Digital collection and/or exhibit
- ☐ Multimodal or interactive active narratives (ie: digital storytelling, digital games)
- ☐ Immersive media applications (ie: AR, VR)
- ☐ Digital objects (ie: 3D objects, images, audio, video, data visualizations)
- ☐ Documentation (ie: internal use, user guides, workflows, procedures, outreach materials)
- ☐ Other

For any of the deliverables above, please describe any software that will be used in creating the end product(s). For example, the digital method of spatial analysis could use tool/software of QGIS, Story Maps, StorymapJS, or CartoDB. The digital method of text analysis could use the tool/software Voyant or R. In depth, responses are not necessary; simply listing the method and software is enough.

Project Owner:

Intended Platform for Delivery

With the above deliverables in mind, select which platforms you intend to use for dissemination (check all that apply):

- ☐ Institutional Repository
- ☐ Open Access Journal
- ☐ Other scholarly journal
- ☐ CONTENTdm
- ☐ Website
- ☐ Other locally hosted software installations (Omeka, Curatescape, OMP, etc)
- ☐ Other externally hosted software installations (Scalar, Omeka, PressBooks, etc)
- ☐ Other

For any selections with "other," briefly describe:

Project Goals

Describe the overall goals of the project, including how it will be utilized. This could include the general public, specific group, or an academic institution, etc.

Project Owner:

Potential Risks

A number of items may prevent a project from moving forward (funding, staffing, timeline, etc). List any limitations or concerns for the project, which may include collaborators, stakeholders, intellectual property concerns, and ethics or privacy issues (especially if working with students). If possible, list approaches to help mitigate these risks.

Success Factors

Describe how the project will be measured for success.

Project Owner:

Roles and Responsibilities

TEAM MEMBER NAME	DEFINE ROLE AND RESPONSIBILITIES	PERCENTAGE OF TIME
Project Owner		
Project Sponsor (if any)		
Team Member 1		
Team Member 2		
Team Member 3		
Team Member 4		
Team Member 5		

Project Timeline

Describe any factors that may affect the timeline, including funding, grant mandates, student availability, resource availability, conference travel, etc.

Project Owner:

Project Requirements

Identify what this project needs that is not already understood or in use. For example: outsourcing digitization, transcription, or website development; team members requiring training; tools or skills needed for project completion; item purchases or travel/fieldwork.

Funding

Describe any grant-funded objectives or mandate, and the timeframe for the grant award.

Project Owner:

Appendix B: Digital File Inventory Item

Title of Document

File Name with Extension

Saved Location *(For best practices, save in three separate locations. For example (1) local machine, (2) external hard drive, and (3) cloud or additional off-site external hard drive. Cloud storage is not recommended for sensitive data because it is third-party storage)*

(1)

(2)

(3)

Creator(s)

Date Created: **Date Last Modified:**

Type/Format: ☐ *Text* ☐ *Numeric* ☐ *Image* ☐ *Audiovisual* ☐ *3D/AR/VR*
☐ *Code* ☐ Other _____

Preservation Copies (alternate/open format)
(For best practices, preserve the original format and open format (if not original). For example tabular data created in Excel, should be saved in Excel (above in Saved Location), but also in an open format such as CSV. Include the file name and location saved such as a hard drive, and/or archives or repositories with permanent identifiers)

Copy (1) File Name:

Location Saved:

If applicable:

Copy (2) File Name:

Location Saved:

Preservation Copies Date Last Saved [yyyy:mm:dd]

Associated with an Approvals Plan (ie: IRB, MOU, Agreement) or Sensitive Data:
☐ No ☐ Yes (indicate Title of Document below and include it in the Digital File Inventory list)

_____

Document described in the readme file? ☐ No ☐ Yes (indicate page number or section)

_____

Project Owner:

Appendix C: Project Profile

External / Public-facing scholarship description of a digital project after completion:

Title of Project

Brief Description

Creator(s) of Project

Date of Project **Publication Date**

Grant or Sponsor Information (if applicable)

Keywords

Identifier (URL, DOI, Handle) of Project

Affiliation and/or Related Links (if applicable)

Recommended Citation

INTERNAL USE ONLY:

Collect the **Collaborators Web Publishing Agreement** from each of the Creators, signally permission to have their name affiliated with the published project online, as described in the Project Profile.

Project Owner:

Appendix D: Collaborators Web Publishing Agreement

The purpose of the Collaborators Web Publishing Agreement is to acknowledge the final outcome of a digital project; and recognize collaborators listed in the Project Profile. The final outcome of the project may include scholarly essays, a website, data visualizations, screencasts, computer code, multimedia components, and/or a final report, among others. *Note: this is different from a collaborators agreement used in project planning.*

If this final outcome (or related components) includes publishing on the Web, permission is sought from each team member (creator) associated with the production of the project. The project owner is responsible for filling in the first half and distributing to each collaborator. Each collaborator fills in the second half (shaded sections) indicating permission status.

Final Project Title:

_____

_____

The above project is produced by _____

The above project will be placed on the website _____ and as a result may appear in other forms of dissemination including but not limited to press releases, news articles, or as a citation in other scholarly works, digital projects, or websites.

Collaborator:

Do you grant permission for your name to be listed as a creator or contributor to the above project?

☐ No, I do not want my name affiliated with this project. I understand the project will be made available online, without identifying me as a collaborator.

☐ Yes, I want my name listed as a collaborator

Name: **Date:**

If you later change your mind, contact the project owner to update the Project Profile and this agreement.

Optional:

*Permission must be attained from students whose independently created scholarly subcomponent(s) may have been submitted for a grade. Permission to use a version of this work must also be included. Use "Sub-Component Title" on this form to give consent to place a copy of the scholarly contribution on the project website. Fill out this form and the Digital File Inventory of the scholarly contribution and give both to your professor. These forms, along with a copy of the original file and open/preservation copy should be given to your professor for placement on the project website. **ENTER SUB-COMPONENT TITLE BELOW:**

_____

_____

Appendix E: Universal Design Checklist

Purpose

Digital projects often use a variety of elements on their webpages; this includes, images, charts, data visualizations, videos, audio, 3D objects, interactive or dynamic objects, and text in many forms such as PDF, word processing document, and schemas such as XML, HTML, or TEI. As part of digital project design and dissemination of projects on websites, it is paramount that steps be taken to assure these elements, both holistically and individually, are accessible and designed with universality in mind.

Universal design (UD) benefits everyone, not just those with disabilities.[36] In her practical guide, Ng covers writing for the web, proper usage of links, audio and visual content including embedding third-party content, among other helpful design considerations. Accessibility and universal design considerations for digital products help to enable the preservation and wider use of such products for the long-term. The objectives of this document is to present a basic, consolidated resource. It is not meant to be a definitive representation of all UD aspects.

Objectives

- To aid in the fulfillment of making a good faith effort at addressing accessibility and universal design considerations for digital products;
- Provide guidance on understanding and developing products with the user experience in mind;
- Continue the re-evaluation of workflow procedures as part of the continuous cycle of development for digital projects;
- Use the UD Checklist as a guiding practice until a professional Web Developer and/or Accessibility Specialist is hired or joins your project team;

Instructions

Apply the UD Checklist to your finished digital project, the final version that will be published on the Web. As the UD Checklist tasks are completed, mark the corresponding box with a check mark and indicate the date of completion. Resources of the various tasks on the UD Checklist are available at the References/Plan Resources section of Digital Project Preservation Plan (Full Plan Version 2).

Additional considerations should include the accessibility of non-HTML content such as embedded or downloadable PDF or text documents, spreadsheets, charts, graphs, and presentations placed online; and the use of captions and transcripts for videos or audio (ARL, 2013) used in digital projects. Other universal design considerations should include visual design with adequate spacing, mobile platform compatibility, use of headings, contrast with colors for text, and a focus that allows users to find and navigate the page with ease.

[36] Ng, C. (2017). A Practical Guide to Improving Web Accessibility. Weave: Journal of Library User Experience, 1(7). https://quod.lib.umich.edu/w/weave/12535642.0001.701?view=text;rgn=main

[37] Association of Research Libraries (ARL). (2013). Web Accessibility Toolkit. http://accessibility.arl.org

Project Owner:

Product Designation

From the list below, select which product type this checklist is based on:

☐ Website ☐ Document ☐ Image ☐ Audio/Video ☐ 3D Object

☐ AR/VR ☐ Other

Product Name [indicate the title of the associated project or file]

_____

Organization

☐ Heading and subheading elements are used throughout and in proper sequencing

☐ Correct placement and use of project branding, include logos, fonts and placement

☐ Consistent colors and font sizes. Use easy to read fonts such as Arial, Calibri

☐ Sufficient color contrast in both text and graphics

Date for completed tasks above: _____

Documents

☐ Created in an accessible form. See best practices #8 in the OER Accessibility Toolkit

Date for completed tasks above: _____

Images

Note: images can include photos, drawings, charts, graphs, and maps

☐ Use of Alternative Text (alt-text) to describe the image function (skip if image is purely decorative and does not convey contextual information)

☐ If image is placed on the webpage directly, make sure the HTML image source is responsive

Date for completed tasks above: _____

Links

☐ Link name is contextual information, not generic

☐ Link opens in the same window or tab; Or the new window or tab is mentioned in the link information

☐ Confirm there are no dead links; all links open as intended

Date for completed tasks above: _____

Project Owner:

Audio

☐ If audio files are used (without video or images), create an accessible transcript to accompany the audio files

Date for completed tasks above: _____

Video

☐ Create closed captions for video and a transcript of the complete narration. The closed captions should be embedded in the video and the transcript should accompany the video file.

Date for completed tasks above: _____

Testing

☐ Use the appropriate technology to test the digital product (for example, WAVE for websites; Screenreader device for websites and documents; Accessibility checkers for document creation)

☐ Use of small group of beta testers on digital product. Select a variety of users include student, faculty, and users with varying levels of visual, hearing, motor and cognitive abilities.

☐ Confirm any sensitive data is removed before publishing on the Web.

Date for completed tasks above: _____

Archival Copy

☐ Save original format file in three different locations: 1) local machine; 2) external hard drive and 3) off-site external hard drive or cloud
[indicate file name/location saved] :

☐ If original format is proprietary or closed, save an additional copy in an open format
[indicate file name/location saved]:

☐ If published online, use of Web archiving tool for digital product (such as Wayback, Webrecorder)
[indicate web archive location, if applicable]:

Date for completed tasks above: _____

Project Owner:

Appendix F: Preservation Guidance Checklist

The preservation goals for minimal and expanded infrastructures will vary by project. It may not be necessary or possible to comply with each goal on the checklist. Use as a starting point and modify.

Minimal Infrastructure (or Short-term) Preservation Goals

Check	Date	Goal
☐		Use open standards and open, uncompressed, non-proprietary formats (if possible)
☐		Build a list of collaborators of varying backgrounds, skills, and strengths
☐		Collect a Web Publishing Agreement from each collaborator
☐		Use a web-archiving tool to preserve the project's static functionality
☐		Use a Project Profile (on the website and/or repository)
☐		Document the process/decisions in a readme file or summary for **internal** purposes
☐		Determine what should be preserved and available for **external** purposes
☐		Use rights statements and/or licensing for attributing credit or reuse permissions for projects, research, code, and metadata developed
☐		Create suggested citations of the intellectual property
☐		Export and save versions (documents, code, development notes, and native file formats) in a structured organization or inventory system
☐		Save copies in three different locations
☐		Data and digital literacy education

Expanded Infrastructure (or Long-term) Preservation Goals

Check	Date	Goal
☐		All the above minimal infrastructure goals above
☐		Consider migration, emulation, or other new preservation capturing method
☐		Reflect the user's perspective; address user experience design methods
☐		Creation of an institutional space for student and faculty project development
☐		Employ accessibility and universal design standards
☐		Use standards for digitization, metadata schemas, encoding, code management
☐		Monitor research and development in preservation practices in various disciplines and create a preservation practices rubric to follow for consistency among projects
☐		Collaborate with other onsite units and experiment with available resources
☐		Develop a consortium of like-minded collaborators across departments and/or institutions for potential development of a DH preservation repository/database

Glossary

A selection of basic definitions for terms used in the Digital Project Preservation Plan

Accessibility: The qualities of a product, service or website that allow ease of use by all persons, regardless of the varying levels of abilities.

Capture: The process acquiring or collecting digital objects or web pages, including its content, elements, and features (HTML, CSS, Javascript, interactive features or embedded media).

CSV (Comma Separated Values)**:** File type for formatted data exported from a table or spreadsheet and separated by commas.

Data Curation: The practice of collecting data with a standardized approach that aids in the preservation, discovery and maintenance of data sets.

Deliverable: A planned ending, product or output of a digital project.

Digital File Inventory: An organized list of all files (image, text, code, video, etc) associated with a project.

Digital Humanities (DH): An interdisciplinary area of study that emphasizes the approach, experimentation, and design of using interactive technologies to expand the participation, the modes of access, and the dissemination of scholarship in the humanities disciplines; a subset of Digital Scholarship.

Digital Literacy: One's ability to recognize and use components of digital information in various platforms and applications, including how information is created and used in digital environments.

Digital Object: A piece of information (image, word processing file, audio file, etc.) in digital form. Several digital objects can make up a digital project.

Digital Project: A broad term encompassing varying levels (planning, preparation, production) and types of interrelated tasks and materials that collectively transform digital objects into a complete work.

Digital Preservation: A set of actions required to maintain access to digital objects or projects for long-term use; this includes advancing beyond the lifespan of technologies used in the original content's creation.

Digital Scholarship (DS): Part of the scholarly communication process, where scholarship is enhanced by the design of digital projects, incorporation of digital tools, collaboration among digital partners, and dissemination through digital platforms.

Dynamic Object or Resource: Digital object that requires interaction to function such as online games, interactive data visualizations, hyperlinked media, and executable software programs.

Emulation: The preservation of both the functionality of the software and the actual hardware used to run the software.

File Format: A standard way to encode a file so a computer can read, display, print and save the information of the file; this includes formats and naming conventions. An open format is made freely available to anyone. A closed or proprietary format is controlled by a company that decides how and if the file format can be used. See page 11 for more information on open and closed file formats.

Metadata: Data about data; this can include descriptive information for individual digital objects. Useful in cataloging digital objects to assist in indexing, retrieval, and preservation.

Migration: The transfer of digital information from an older hardware/software configuration to a newer one.

PDF (Portable Document Format)**:** A standardized open format for presenting documents.

Persistent Identifier: An actionable/linkable (via a browser), public and unique identifier or location that lasts over time. Examples include:

> **DOI** (Digital Object Identifier), popular with journal publishing, embedded in a URL as http://dx.doi.org
>
> **Handle** an identifier managed through an API or user interface, embedded in a URL as http://handle.net
>
> **URL** (Uniform Resource Locator) typical address for web content, beginning with "http://"

Plugin: An additional software needed to run a program.

Preservation Copy: A copy of the digital file that is better suited for long-term preservation with less chances of degradation or access to original file format (if proprietary or closed); serves as an alternative or open format of the original file.

Project Charter: A set of guidelines that govern a project including the goals, objectives, limitations, timeline, deliverables, team member responsibilities, etc.

Project Profile: Descriptions (metadata) of key elements for a project that help explain its purpose and aid in its discovery. Descriptors may include the title, creators, keywords, publisher, suggested citation, copyright, and Web links.

Project Owner: A creator, originator, or designated person with the highest or overall responsibility of a project.

Readme File: In general, a form of documentation describing the technical specifications needed to run software. For DH purposes, serves as a narrative summary of how each file is connected to the digital project. A DH readme file puts the development process into words; it can describe how different elements of a project fit together, including the process, workflow, and decisions made along the way. Variations can be created that prove useful for archiving a project's methodology internally or explaining a project's purpose to an outside audience.

Repository (Digital, Institutional, Subject)**:** A place or location that stores data with a dedicated technical infrastructure responsible for ingesting, maintaining and preserving digital objects deposited by users. Deposits are described with metadata, given permanent identifiers (URI, URL, Handle) and made accessible and discoverable in search engines and databases because its contents are indexed.

Server: A physical or virtual combination of hardware and software that carries out services for programs in a network (a location that hosts the digital files that make up a project).

Web Archive: A location that allows future access to captured content from Web archiving, which is the process of collecting digital objects from the Web with the intent to manage and preserve them for long-term storage and usability.

Universal Design: An approach to the design of products or services that accounts for the needs of all types of potential users.

Digital Project Preservation Plan

A Guide for Preserving Digital Humanities / Scholarship Projects

https://jewlscholar.mtsu.edu/handle/mtsu/5761

Visualizing Objects, Places, and Spaces: A Digital Project Handbook

Assessing User Experience using Digital Humanities Projects in a Design Classroom

Serenity Sutherland

Published on: Jun 27, 2022

DOI: https://doi.org/10.21428/51bee781.957d585e

Author & Project Role

Author: Serenity Sutherland, SUNY Oswego

Role: Assistant Professor, Department of Communication Studies

Assignment URLs and/or Files

Students sign up for DH projects to review.

Then, students write a review based on this prompt.

BRC223-Reviews.docx 19 KB

Learning Objectives

What did you want students to be able to do by completing this assignment?

1. Evaluate digital media as content consumers employing moderate to high media literacy.
2. Assess metadata features like alternative text for determining accessibility and principles of universal design.
3. Critique the user experience and design of digital media while simultaneously writing with web-based writing tools and strategies
4. Understand digital humanities projects as one genre of scholarly web-based writing and research

Technology-Dependent Learning Outcomes

Was there anything this assignment taught students that you felt they wouldn't have been able to learn through other types of class assignments?

Students learn to use tools like WAVE WebAim's Accessibility Checker to critique web content for universal design and web accessibility standards. They also learn how to create a blog post using their individual WordPress.org sites. And finally, they learn the difference between web-based writing and other types of academic genres.

Skill Level

What is the course title and level?

This is a 200-level course, required for all students in the Broadcasting and Mass Communication major. The course title is “Integrated Media: Writing and Design.” The course is intended to teach students how to use many types of digital and web-based media to tell a professional story about themselves. Students create websites as professional portfolios.

What kinds of prior knowledge is necessary to complete this assignment? How do students gain this knowledge?

Students do not require any prior disciplinary knowledge. Typical college-level digital media skill usage is necessary, such as how to use a computer and an internet browser.

Assignment Description

The course I use this assignment in is not a digital humanities course. It is instead intended to teach students basic design principles so they can make their own websites using a Content Management System (CMS) such as WordPress. Additionally, students develop their media literacy and web-based writing skills in the course. Evaluating a digital humanities project helps them achieve these skills and introduces them to a way of knowing and creating (the digital humanities) that many have never heard about before. I share this assignment with the DH community not because it is particularly “DH-heavy” or full of neat technological tricks, but rather because it is one way I’ve found to weave together my research interests with my teaching responsibilities, which fall outside of a traditional DH curriculum. Furthermore, the assignment makes sense given the goals of the class: to increase media literacy, practice writing web content, design ethical web-based media and use different types of media to tell a story.

Students select a project they are interested in from a curated Google Spreadsheet list of DH projects. The curated list spans many different topics, regions and histories, and students are generally enthusiastic to choose something. We take about ten minutes in class for the students to begin looking. I circulate around the room and comment on each project they are looking at - providing small tips like “oh, this project is a really fascinating story about canoes,” or “can you believe they do this with coins?” or “This is an archive where you can toggle the transcription between English and Dutch - how cool is that - do you know Dutch?” (Their answer is no, my answer is also no). In this way, students can ask questions if they are confused. They can decide right in class, or later on at home, which project to review.

Students then spend about a week and a half reviewing the site on their own. They are instructed to focus on key user experience (UX) elements that we've discussed in class: things as simple as aesthetics and color, to more complicated design features such as reviewing any available code or evaluating how well the site does with accessibility features, such as descriptive alternative text. The variety of digital humanities projects allows students to review and critique the affordances of certain types of digital projects for accessibility. For example, alternative text is essential for images, but what about a project utilizing dynamic data visualizations that change as the user engages with the content? They write their review in WordPress employing the same design features they have been critiquing on the digital humanities site.

Because this is also a writing class, students engage in peer review of two of their peers' WordPress posts reviewing a DH project. This way, students are exposed to about three digital humanities sites. After making the appropriate edits, students then post and publish their reviews on their sites and submit for a grade. The students study broadcasting within a Communication Studies department and for many this is their first, and possibly only, exposure to digital humanities as a research method and a way of telling stories on the web.

Time Needed

How much time did you allot to this project?

Because the goals of the assignment are heavily integrated into the course, I spend a lot of time preparing the students to be successful in general content creation and content consumption on the web. The assignment asks them to employ nearly everything we've learned up to this point, including one 50-minute lesson spent on each of the following:

- principles of good design - does the web navigation "make you think"? (Krug, 2014); and do these "everyday things" on the web tell good stories? (Norman, 2013).
- User experience and user interface - how much control do we have over these in CMS's like WordPress?
- Accessibility and Universal Design - how to use tools like WAVE WebAim to evaluate alternative text? How does a screen reader work?
- How to write posts and create content in WordPress? How to write/design as you create content? (Ball, Shepard, Arola, 2018).

I spend about 20 minutes in class the day I introduce the prompt detailing what the digital humanities is as a field of study, the goals of this type of web content, why

digital humanities projects are important, and how many DH projects are built for a general audience. I spend time setting DH up to give them a bit of background, but also to let them know that the students are the ideal audience in some ways because DH-researchers and creators often imagine, and feel excited about, a user "stumbling" upon their site with little subject expertise and contextual background.

Support & Training

What kinds of support or training did you provide to help students learn to use new techniques or specialized tools?

They need to know how to create posts in WordPress and how to take screenshots of websites in order to provide evidence for their claims. They also need to know tools such as accessibility evaluators. These are all tools taught directly in the course and so the assignment blends the "hands-on" use of the tools learned in class alongside a critique of digital artifacts.

Resources

Did you need any specialized equipment, tools, or human resources to make this assignment feasible? If so, please describe.

For this particular iteration in a design/web writing focused course, students do need to have a website. This means they have signed up for and use Reclaim Hosting and install WordPress. It's important to emphasize that the hosting and WordPress are related to other goals of the course (to create a professional portfolio/site where students practice design and web-based writing). This assignment could just as easily be used without any CMS and simply done as an essay in a word processor.

Assessment

How did you assess or grade this project?

When grading I refer to the prompt as my rubric. I assess the following:

- Is there an argument or main claim?
- Is the argument backed up with evidence, or if no argument, what kind of evidence is provided?
- Does the reviewer keep medium-specificity in mind while reviewing the site (i.e., not reviewing an interactive art project unfairly as a video game)?

- Has the student assessed accessibility and backed their claims up with evidence from an accessibility checker?
- Does the student discuss UX, what it is, why it matters, and how their chosen DH site meets standard UX?
- Does the student employ principles of good web design and web writing themselves in the writing/designing of their post?

Challenges & Opportunities

If you assigned this project again, would you change anything? If so, what?

I assign this project every semester. I make small tweaks here and there and I'm always adding (and taking away due to link rot or unplanned project "endings") from the curated list of DH projects. In more recent years I've started offering the caveat that students will find projects on the list that have wonderful UX and some that have poor, or flagging, UX - this is intentional as students can learn what to do and what not to do, and that while they may want to take inspiration from the sites they review, they should also learn lessons (such as monitoring their sites for "404 errors," reviewing their web tools for dying software such as Flash, and making sure hyperlinks work) on how to keep their content fresh and sustainable.

Describe any trouble spots or complications someone else might want to be aware of before trying a similar assignment in a course of their own.

Depending on the course goals, this assignment might be hard to employ in a DH-focused class as I take a very "big tent"/ "contact zone" approach to inclusion in what is DH (Svensson, 2012; Ortega, 2019). Some students in more advanced courses may wonder why certain projects are considered DH in this list, although this seems like a wonderful pedagogical moment to discuss the boundaries and fluidity of the digital humanities.

Occasionally, one difficulty with this assignment is that a very few students may have missed signing up for WordPress and may not have sites ready for this assignment. This is a very infrequent occurrence, though, as we sign up for WordPress in class, learn the tools in class, and I introduce the assignment (complete with previous examples) in class. If students need help with WordPress, they typically reach out to me for assistance.

Bibliography

Ball, Cheryl E., Jennifer Sheppard, Kristin L. Arola. *Writer/Designer: A Guide to Multimodal Projects* (Bedford/St. Martins, 2018, 2nd edition).

Krug, Steve. *Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability* (New Riders, 2014, 3rd edition).

Norman, Don. The Design of Everyday Things: Revised and Expanded Edition (Basic Books, 2013).

Ortega, Élika, "*Zonas de Contacto*: A Digital Humanities Ecology of Knowledges" in *Debates in the Digital Humanities*, Matthew K. Gold and Lauren F. Klein, eds. (University of Minnesota Press, 2019).

Svensson, Patrik, "Beyond the Big Tent" in *Debates in the Digital Humanities*, Matthew K. Gold, ed. (University of Minnesota Press, 2012).

CHAPTER 1

Studying users in digital humanities

Claire Warwick

Introduction

Until relatively recently, it was unusual to study users in digital humanities. It was often assumed that the resources created in digital humanities would be used by humanities scholars, who were not technically gifted or, perhaps, even luddites. Thus, there was little point asking them what they needed, because they would not know, or their opinion about how a resource functioned, because they would not care. It was also assumed that technical experts were the people who knew what digital resources should look like, what they should do and how they should work. If developers decided that a tool or resource was a good one, then their opinion was the one that counted, since they understood the details of programming, databases, XML and website building. The plan, then, was to provide good resources for users, tell them what to do and wait for them to adopt digital humanities methods.

Frustratingly, potential users seemed stubbornly to resist such logic. The uptake of digital resources in the humanities remained somewhat slower than in the sciences. As I have argued elsewhere, the numbers of articles in journals, such as Computers and the Humanities (CHUM) and Literary and Linguistic Computing (LLC) in the 1990s and early 2000s, complaining about why traditional humanities scholars did not use digital humanities techniques or suggesting techniques they might use, grew heavily to outnumber those reporting on the actual adoption of such techniques in the mainstream (Warwick, 2004). Lack of knowledge was sometimes advanced as a possible reason for lack of engagement. During this period, very large amounts of money were spent on initiatives to publicize digital resources for humanities research and teaching. In the UK, this included the Computers

and Teaching Initiative (CTI) (Martin, 1996), the Teaching and Learning with Technology Programme (TLTP) (Tiley, 1996) and the Teaching and Learning Technology Support Network (TLTSN) (Doughty, 2001); the Arts and Humanities Data Service (AHDS: www.ahds.ac.uk) also had an advice and outreach role, as well as its core function of data preservation. None of these are now in existence; funders did not feel they had proved sufficiently successful to continue supporting them. Many university libraries and computing services also offered training courses in the use of digital resources for humanities scholars. Yet, the rate of change remained stubbornly slow. Funding bodies also supported digital resources for humanities scholars, with little thought to, or predictions about, levels of possible use because they did not know how such predictions might be made. Such resources often cost hundreds of thousands of pounds, so there was a risk of a severe waste of money and of academic time and energy if a funded resource was then not adopted.

In the late 1990s, a few of us began to wonder if there might be another cause for the lack of adoption of digital humanities resources. Could it be that users did not adopt resources because they were not useful or did not fit what they would like to do as scholars? Could there be other reasons to do with design, content, presentation or documentation? Initially, I suggested that digital resources available in the late 1990s did not fit the predominant research method of humanities scholars, which is complex reading (Warwick, 2004). Later, empirical studies on the way humanities scholars interact, or fail to interact, with digital resources allowed us to test this hypothesis. This chapter presents an overview of the findings of such work, arranged thematically.

What we know about humanities users

Despite some erroneous perceptions in both digital humanities and the computer industry, we know a significant amount about how humanities scholars use information, whether digital or not. Since Stone's pioneering article in the early 1980s (Stone, 1982), numerous studies of information needs, and some of information behaviour, have been published, both of the humanities as a field and of individual disciplines (Warwick, Terras, et al., 2008). As we have argued in more detail elsewhere (Warwick, Terras, et al., 2008), these suggest that humanities scholars are not luddites; they simply behave differently from scientists, and many social scientists, when interacting with physical and digital information. Humanities scholars tend

to avoid performing systematic keyword searches, although most information systems and digital resources assume this. Instead, they will follow footnotes in texts they are reading (what Ellis calls chaining (1993)) or browse for information. They may even do what Bates calls 'berry picking' – in other words, select interesting pieces of information that are particularly germane to the argument they want to make, rather than citing everything written on the subject (Bates, 1989, 1). (We might speculate that this may also become more common in science in the future, when the sheer number of articles published every year exceeds the researcher's ability to read them all.) They also need a greater range of information, in terms of publication date and type: instead of reading journal articles from the last five years, they may need to consult printed books or manuscripts that are hundreds of years old, as well as images, film, music, maps, museum artefacts and various different types of historical source material (Barrett, 2005). They do not expect to solve a research question comprehensively, but to reinterpret the sources and revise the findings of others: after Crick and Watson, no one tried to redefine the structure of DNA, but articles about *Hamlet* will probably always be written. They often reread or re-examine sources in a complex, immersive way, rather than searching digital documents for factual information.

It is evident, therefore, that humanities scholars have different information needs, both on- and offline, than scientists. They are a problematic population to design for, and the field lacks the financial clout of Science, Technical Engineering and Medicine (STEM) subjects, so funding to create resources for their needs is less plentiful and may seem less profitable for commercial publishers. It is, therefore, not surprising that, until recently, most resources have been designed for the majority of users who are not from the humanities. Yet, we might argue that the way they use digital resources is, in fact, closer to the way that the average, non-academic user interacts with digital or printed information. Most of us read for pleasure, may consult a wide range of information resources and don't conduct systematic keyword searches of recently published scientific literature; thus, a study of humanities user needs may also produce important results relevant to non-professional digital resource use.

How to study users

There are numerous methods for studying users, most of which have been developed in the fields of Human-Computer Interaction and Information Studies. There are also many excellent texts describing, in detail, how these

may be carried out, for example, Shneiderman and Plaisant (2009), Blandford and Attfield (2010) and Ruecker, Radzikowska and Sinclair (2011). Our approach at UCLDH has been to use a variety of methods, most of them designed to be as naturalistic and unintrusive as possible. Our overall approach is to study use in context; that is, to study what people do in their real life or work activities. This means that we prefer to visit someone in their office (or, in one case, an archaeological dig) and ask them to carry out a real research activity using a digital resource, rather than asking them to perform a set task in an interaction lab. We have used task-based lab testing for some research projects, but, in general, prefer to adopt as naturalistic an approach as possible to avoid the user's behaviour being prejudiced by unfamiliar conditions.

Our approach to studying users is to involve them, if possible, from the beginning of the project. Too often user testing, both in academic projects and industry, is left until late in the project; users are only asked for their opinion when the resource is built and a prototype is being tested. This may work, if the users like what has been built for them. However, if they do not, and feedback suggests radical change is necessary, there may not be sufficient funding, time or goodwill from developers to make such modifications. In such cases, the resource either remains unmodified or different researchers may be called in to conduct other tests, in the hope that they will find what the developers want them to discover, not report what users actually need. This is a very dangerous strategy, for reasons that I shall discuss below.

Thinking about use before a resource is built means studying the users, not the resource; this may be achieved using various methods. We have used interviews to determine what scholars like and dislike about digital resources and how they use information, and we have observed them using existing digital resources. We have asked them to keep diaries of their use of information and digital technologies over periods varying from between a day and a week (Warwick, Terras et al., 2009). This allows us to identify patterns of, and problems with, information usage, about which we can subsequently interview users. We have used surveys and questionnaires about the use of existing resources. We have interviewed the creators of existing, successful resources to see whether it is possible to identify any common features, in terms of design, creation or documentation (this is an unusual approach, and we believe we are the only team to have employed it in digital humanities; but it is an approach that we found very instructive during the Log Analysis of Internet Resources in the Arts and Humanities

(LAIRAH) Project) (Warwick, Galina, et al., 2008). All of these methods allow us to build up a picture of what users like and dislike, what they want to do and what they currently cannot achieve. This is then fed back to design teams to inform initial design and prototype ideas.

When initial ideas are being developed, it is also possible to use models, such as Ruecker's affordance strength model (Ruecker, Radzikowska and Sinclair (2011): Chapter 3), which allows us to test the potential functionality of a prototype design against some possible uses. At a slightly later stage in development, we can use wire frames and design sketches to run user focus groups. We have also conducted workshops, where users are asked to investigate different digital resources, record their views on paper and then take part in a subsequent focus group discussion. During the LAIRAH project, for example, we presented users with a mixed sample of resources that were either known to be used or neglected, without identifying them, asked them to speculate on which ones where used and comment on their reasons for saying so. This was then followed by a focus-group discussion. This proved a useful way to limit the bias inherent in focus groups, when one or two vocal members of the group may dominate and, thus, skew results. Subsequent examination of the written responses showed that users were willing to be more positive about some resources in writing, than they were in group discussions.

This variation between what people may say to others and what they will record in private is the reason why it is important to use a variety of different methods in user studies. It is well known that interviewees may say what they think someone may wish to know; thus, they may be more forthcoming if asked to fill in a survey or write down responses to a hands-on workshop session (Smith and Hyman, 1950). This is also why we have used quantitative data from web log analysis, since reported use may differ from what logs record, which may also be attributable to the interviewer effect. In the days before logging software, such as Google Analytics, was routinely used, very few projects or, even, institutions, such as libraries, had any reliable indication of which resources were used. Log data allowed us to determine that up to one-third of digital resources in the humanities remained unused (a very similar level to that of printed material in libraries) (Warwick, Terras et al., 2008) and to indicate the kind of material most commonly searched for. Log analysis can also indicate whether certain parts of a resource are used more often than others and whether this is related to content or design problems (the more clicks away from the index page, the less likely it is that users may find material, for example) (Huntington et al., 2002).

Of course, conducting user studies adds to the cost of developing digital resources. The time required to undertake such activities, especially if they last throughout the project, is considerable. Some projects have, instead, chosen to make use of personae or use cases. Some designers create indicative personae of typical users, giving them names, ages and occupations, and thus suggesting the uses that such a person might make of a resource (Jane is a secondary school teacher in her 30s. She wants to use a museum website to construct some new assignments about Roman food for her year 11 class on classical civilization, for example) (Grudin and Pruitt, 2002). Personae can be a useful tool, if they are constructed as a result of the kind of user studies mentioned above. However, if they are used as a substitute, there is a danger of a kind of self-fulfilling prophecy of use, where functionality is designed for the kind of users the designers want or can imagine. Yet, they cannot be sure that this is the kind of user that the resource will actually attract or that these predicted difficulties are the kind of difficulties that imagined users might face.

Use cases consist of reports of how a user, or small group of users, is using a given resource or one that is very similar. These are often used to make the case to develop something new or to argue that certain types of interface or functionality may be useful. Once again, these may be used as part of a multimethod user study, as evidence of real usage (Keating and Teehan, 2010). However, if used in isolation, the picture of use may be very partial, unless a very large number of use cases are collected. The behaviour of expert users or early adopters may also be very different from that of a majority of users, yet it is often the interested experts who furnish the use cases. As a result, the need for complex, specialist functionality, or the general enthusiasm in the user population for the resource, may be overstated. Use cases and personae should, therefore, be used with care in a multimethod user study, and should never be a substitute for other, more time-intensive methods.

Luddites or critics?

Despite the popular image of the luddite humanities scholar who does not know what they need or how to use it, we have found that users have very complex models of their information needs and environment; they are thoughtful and critical about the affordances of physical and digital resources. This may help to explain why e-journals have been such a success, and e-monographs are still not widely used. Users are aware that a journal

article and a book are used in very different ways, even if they do not articulate this until asked. Thus, most of us still prefer to read a book in print, because it is more convenient, but are happy to read a short article on screen or print out a longer one. We also found that humanities users had complex ways of evaluating physical information resources and could tell, simply from the design, publisher or even size of a book, whether it was likely to be useful. It is still difficult for users to find digital analogies for such skills, however, and it remains an important challenge for creators of large digital resources to design tools that will allow users to orientate themselves digitally as well as they can in a physical library. This is the reason for tools such as Amazon's 'user recommendation systems' (users who bought this, also bought …), but it is far more difficult to deploy such metaphors in an academic setting. Even the question of extent of collections is problematic; physical library users can see how big the shelf is that they are looking at and how many of them there are in a library. It is still very difficult for users to estimate how large a digital resource is and, thus, how comprehensive the results set from their search may be and how much further they need to explore. This is important for humanities users, who value recall over precision and expect to find about 90% of the results from a given search familiar. Nevertheless, we should not assume that humanities users always prefer physical to digital information resources. Users we have studied have found the convenience of digital information delivery as important as those in any other discipline and expressed considerable enthusiasm for the use of digital resources and methods. Difficulties caused by a badly designed interface to a digital collection were no more significant than a library or archive that was cold, cramped, dark or uncomfortable or an unhelpful member of staff. However, they were more likely to put up with difficult physical conditions than persist with a disappointing digital resource. It would seem ridiculous to a humanities scholar to refuse to return to a physical library if a book they hoped to find was not stocked, yet I have often heard digital resources dismissed outright if the contents were not as expected. It is difficult to tell whether this is something inherent in the nature of physical and digital information resources or whether, like the question of transferring information skills from physical to digital libraries, it is a problem of relative unfamiliarity on the part of users and signals the need for further refinement of the digital resource design. We may only find the answer to this question by repeating studies over time and trying to determine whether, and how fast, attitudes change. It is not, however, necessarily a function of being a digital native or immigrant; indeed, a recent

study suggests that there is no empirical basis for such assumptions. Rowland's research suggests that the information literacy of even, what he calls, the 'Google generation' is relatively unimpressive (Rowlands et al., 2008, 1). We found that even students who have been trained in information-seeking skills will give up as soon as they have a minimal level of information to complete a task and that they use their relative expertise to determine how little searching is necessary in a given situation, rather than conducting more complex searches to find a more complete result set. If the results of a search seem too complex to evaluate, they may even alter their query to achieve a simpler, less demanding answer. We cannot, therefore, assume that once a younger generation of scholars arrives, their ability to interact with complex digital information will necessarily improve.

Finding and using digital information seems to have something to do with how important it is to users. Students may not gain expertise gradually, however. The difference noted in the skill levels of young legal professionals, who may only be a few years older than our student sample, is probably because the information tasks they faced at work were more complex and urgent and forced them to suddenly acquire more expertise. However, it does help to explain an interesting phenomenon we found during the LAIRAH project, when we discovered that humanities scholars could be very easily deterred from using digital resources. Numerous factors caused this: confusing interfaces, problems with navigation or searching, a need to download data and use it with another application, content that was incomplete, not extensive enough, of poor quality or not as expected (for example, if a literary resource did not contain appropriate editions, it was considered unacceptable to many users). Yet, we found that if a research task is vital to the individual, and they are convinced that a resource will deliver high quality information, they will persist with a digital resource and force themselves to learn new skills or struggle with a difficult interface or functionality. Thus, we found that some linguistic resources were reported to be very useful, even if poorly designed, dated and difficult to use, because there was nothing better available for specialists in that field. The problem is that the proportion of such determined and persistent users appears to be quite small.

It has become clear to us, however, that most users will be quick to abandon resources whose quality they are concerned about. This is partly as a result of minor problems that could be relatively easily avoided. Our study of successful digital resources, during the LAIRAH project, suggested that even the name of the resource could make a difference to use. If someone is searching for census data, they may not also think to use the term

'enumerator returns', and, unless there is very complex metadata, or semantic searching is possible, a resource with a confusing or unusual title may not, therefore, be found. The possible uses of digital resources designed by technical or academic experts were often not evident to potential users. Not everyone, for example, knows what Geographical Information Systems (GIS) are or how they might be employed. This is not a problem for the dedicated expert user, but may mean that a, potentially, much larger audience fails to understand the potential use of some resources (Warwick, Terras, et al., 2008). This is part of what we have called the 'designer as user problem' (Warwick, Terras et al., 2008). If digital resources are created by academic or technical experts and user testing is not carried out, the assumption tends to be made that the users of the resource will be just like the creators. The academic creator may assume that everyone will understand what the resource is for and what it contains without much explanation, because it is obvious to them. They may also assume (possibly abetted by technical staff) that complex functionality and search capability is needed to make the resource usable and if they can learn to use such functionality, then anyone can. It may be, however, that most people do not need, or perhaps even like, the complicated functionality and, perhaps, difficult interface necessary to make this possible (Warwick, Galina, et al., 2008). The simple, Google-like search box has become a standard way that users expect to interrogate most collections of information; this is partly because it works. Most users, especially humanities academic users, do not want to have to be trained to use digital resources, regarding it as a waste of time. Some librarians have even alleged, strictly off the record, that they suspect academics do not want to admit ignorance, especially in front of their students, and that this may be a more profound reason for their antipathy to training (for obvious reasons, a source for this cannot be cited). Most technicians, librarians and commercial publishers who market resources at librarians seem to believe that it is important that all resources must have an advanced search function. In fact, numerous studies have shown that most people never use this function (Rieger, 2009). It is, therefore, clear why the model of designer-as-user is not advisable. It may lead to the creation of a resource that is needlessly complex, expensive in developer time, potentially not what users want, and, therefore, at serious risk of being under-used as a result.

CASE STUDY The LAIRAH Project: log analysis of internet resources in the arts and humanities

Claire Warwick, Melissa Terras, Paul Huntington, Nicoleta Pappa, and Isabel Galina, UCL Department of Information Studies

The aim of the LAIRAH survey is to discover what influences the long-term sustainability and use of digital resources in the humanities through the analysis and evaluation of real-time use. We utilized deep log analysis techniques to provide comprehensive, qualitative and robust indicators of digital resource effectiveness. Our aims were to increase understanding of usage patterns of digital humanities resources, aid in the selection of projects for future funding and enable us to develop evaluation measures for new projects. Until we carried out our research, evidence of actual use of projects was anecdotal; no systematic survey had been undertaken, and the characteristics of a project that might predispose it for sustained use had never been studied.

Methods

Phase 1: log analysis

The first phase of the project was deep log analysis: we were the first team ever to analyse web transaction logs to measure user behaviour within digital humanities resources. Transaction and search log files were provided by three online archives that were supported by the Arts and Humanities Research Board (AHRB) (now the Arts and Humanities Research Council (AHRC)): the Arts and Humanities Data Service (AHDS) Arts and Humanities Collection, Humbul Humanities Hub and the Artefact Database for the Creative and Performing Arts. These provided rich data for comparing metrics between subject and resource type. The search logs showed which resources users were interested in and which ones users subsequently visited.

We analysed at least a year's worth of transaction log data (a record of webpage use automatically collected by servers) from each resource. This data provided a relatively accurate picture of actual usage, providing: information on the words searched (search logs), the pages viewed (user logs), the website that the user has come from (referrer logs) and basic, but anonymous, user identification tags, time and date stamps.

Phase 2: case studies

We selected a sample of 21 projects that the log analysis indicated to have

varying levels of use – chosen to give us coverage of different subject disciplines – to be studied in greater depth. We classified projects as 'well used' if the server log data from the AHDS and Humbul portals showed that they had been repeatedly and frequently accessed by a variety of users. We also mounted a questionnaire on these sites and asked which digital resources respondents found most useful. Although most users nominated information resources, such as libraries, archives and reference collections, such as the eDNB, three publicly funded UK research resources were mentioned, and, thus, we added them to the study. We also asked representatives of each AHDS centre to name which resources in their collections they believed were most used. In the case of Sheffield University, the logs showed that a large number of digital projects accessed were based at the Humanities Research Institute (HRI). We therefore conducted interviews about the HRI and its role in fostering the creation of digital humanities resources.

The projects were studied in detail, including any documentation and reports that could be found on the project's website, and a representative of each project was interviewed about project development, aims, objectives and their knowledge of subsequent usage. We analysed each project's content, structure and design. We asked whether it undertook any outreach or user surveys and how the results of surveys were integrated into project design. We also asked what kind of technical advice the project received, whether from institutional support people, from humanities computing centres or from central bodies, like the AHDS. All these measures are intended to determine whether there are any characteristics shared between 'well used' projects.

We also studied projects that appeared to be neglected or underused. A small group of humanities users were asked to investigate a sample of digital resources: half were well used and the others neglected, but their status was not initially revealed. A hands-on investigation was followed by a discussion of factors that might encourage or deter future use of such resources. We aimed to find out whether their lack of use was because users had not heard of a given resource or whether there were more fundamental problems of design or content that would make the resource unsuitable for academic work.

Findings

We found that roughly one-third of all projects appeared to be unused. When asked to evaluate unused resources, users were able to identify several problems with design and content. They were deterred from use because of unintuitive interfaces, the need to download data for use in another application, confusion

as to what the content might be used for and even a confusing name. They also needed more information about the content of resources, how and why it had been selected and the expertise of the project team.

Well used projects did share common features that predisposed them to success. The effect of institutional and disciplinary culture in the construction of digital humanities projects was significant. We found that critical mass was vital, as was prestige within a university or the acceptance of digital methods in a subject field. The importance of good project staff and the availability of technical support also proved vital. If a project is to be well used, it was also essential that information about it should be disseminated as widely as possible. The single most common factor in use of a project was a good dissemination strategy. Even amongst well used projects, however, we found areas that might be improved: these included organized user testing, the provision of, and easy access to, documentation, and the lack of updating and maintenance of many resources.

Recommendations

Digital humanities projects should undertake the following actions:

1. Keep documentation and make it available from the project website, making clear the extent, provenance and selection methods of materials for the resource.
2. Have a clear idea of whom the expected users might be; consult them as soon as possible and maintain contact through the project, via a dedicated e-mail list or website feedback.
3. Carry out formal user surveys and software and interface tests and integrate the results into project design.
4. Have access to good technical support, ideally from a centre of excellence in digital humanities.
5. Recruit staff who have both subject expertise and knowledge of digital humanities techniques, then train them in other specialist techniques as necessary.
6. Maintain and actively update the interface, content and functionality of the resource, and do not simply archive it.
7. Disseminate information widely, both within specialist subject domains and in digital humanities. ■

Trust

As we have seen, users need as much information about a resource as possible to understand what it might be useful for. However, underlying much of our research on users is the issue of trust in digital resources and technologies. The more information users can find about a resource, the more they are likely to trust it. As discussed above, humanities scholars have a complex repertoire of information skills that allow them to evaluate traditional information resources. These have grown up over several hundred years of the development of printed academic resources (Vandendorpe, 2009). A prestigious journal name or book publisher tells us that the content has been peer reviewed by other academic experts. Footnotes or references in the text reassure us that the writer has compared their findings with other work in the field and researched other sources. The academic affiliation of the author tells us about their expertise and standing in the field. The methodology of an article tells us how the work has been conducted, for example, how data was selected, sampled and analysed. Digital resources are only beginning to find ways to provide such information. In the LAIRAH report recommendations, we suggested that all digital resources should have a top-level link called 'About this Project', or something similar, under which creators should provide as much information as possible about its purpose and how it might be used; what its contents are and how comprehensive they are; if selections have been made from a larger corpus, how this has been done, why, and who has done so; who created the resource and where they are based; how technical decisions were made, for example, about the markup or metadata schema. The more effectively this is done, and the more easily it can be accessed, the more users are likely to trust digital resources. This is likely to become even more important in the near future. The UK's Research Excellence Framework will now allow digital resources to be submitted in all subject areas and not simply the publications written about them (Higher Education Funding Council for England, 2011). As a result, it will become even more vital that we gain a sense of the rationale for the choices made in the course of digital resource construction, so that assessors can make informed decisions about resource quality and impact in the wider world.

At present, however, trusted brands are very important. Many digital resources that are most familiar to users, such as e-journals or large digital reference collections, are produced by commercial publishers, who make significant investments in testing the appearance and functionality of their resources. This is also usually the case with digital resources in major

cultural heritage organizations, such as museums and galleries. This means that the standard of resource that academic users expect is often higher than most academic projects can manage, especially for interface design. This, coupled with the brand identity of museums and major publishers, reassures users about the quality of the content.

A pioneering study showed that visitors to websites make judgements about them in fractions of a second. We appear to make up our minds about digital resources too quickly to perform a conscious critical evaluation of it: our gut instinct tells us whether it looks 'right'. If users sense that something looks 'wrong' – which may simply mean that the interface looks unfamiliar, is difficult to use or lacks information about its creation and provenance – users may regard it as untrustworthy, neglect it and revert to more familiar resources, whether printed or digital (Warwick, Terras et al., 2008). This demonstrates why those creating digital projects must design a resource that works easily and looks as impressive as possible. The only way to do this, other than being lucky, is to carry out proper user testing.

One of the reasons users think that resources look 'wrong' is if they seem dated. If they try to use a resource and parts of it no longer work – links are broken, for example – they will lose yet more trust. Commercial resources are updated constantly, to make sure that information is current and the interface functional and consistent with current design trends. The problem for many digital resources based in the academic and cultural heritage sectors is that there may no longer be any funding to perform such updating if the content is freely available and was funded by a fixed-term grant. As we have seen, if users do not feel that a resource is to be trusted, because it appears to be dated, they are reluctant to use it. This is a waste of the (probably) very large amount of money that was spent on its creation. Institutions have only recently begun to develop strategies to deal with this problem.

This is especially serious for resources that involve crowdsourcing or web 2.0 technologies, where users become an integral part of the research process. For example, the award-winning Transcribe Bentham project, discussed in Chapter 2 of this volume, was funded by a short research grant. However, at the end of its funding period, over 1000 people had already taken part in transcribing manuscripts and become part of a thriving user community. Since this project is an important vector for engagement between UCL researchers and the public, to have closed it and locked out all our volunteers would have been disastrous and contrary to everything that UCL believes in, in terms of outreach and openness. As a result, short-term

internal funding had to be provided, while further external funds were sought. UCL recognizes the need both to maintain the infrastructure and to continue the activity with which it is engaging volunteers.

Longitudinal studies

Change over time is not something that is very often considered in terms of user studies. They are often carried out at particular points in time, and significant longitudinal studies are relatively rare. There tends to be an assumption, therefore, that user views of digital resources are somewhat fixed. In digital resource creation, one important principle that should be followed, if at all possible, is a cycle of user testing and feedback. Once tests have been carried out and modifications made, it is important to feed back to users what has been done in response to their views. This can either be done by direct communication, in the form of a change log or development blog on the website; an end of project workshop; or another written form of communication with the user community, such as an online newsletter or progress report. An iterative development cycle is, in itself, a useful way to communicate with users. If, for example, a focus group has been carried out to ascertain users' views of wireframes or design sketches, then a hands-on session with a prototypical system not only helps to indicate whether views initially expressed are true in a working version, but shows that the development decisions taken reflect initial users' views. Users like to be able to see that changes have been made as a result of their input and will often be very supportive of something that they helped to create. Our work on the VERA Project was an excellent example of this. The following case study gives the full details of the project, our part in which was to study the way that archaeologists use digital technologies in the field, especially to record what they have found.

CASE STUDY The VERA Project

Claire Fisher, British Museum, Melissa Terras, UCLDH, and Claire Warwick, UCLDH

The Virtual Environments for Research in Archaeology (VERA) Project was funded as part of the Joint Information Systems Committee (JISC) and involved the Department of Archaeology and the School of Systems Engineering at the University of Reading, the York Archaeological Trust and the School of Library, Archive and Information Studies at UCL. The project was based around the

University of Reading's well established excavation at the Roman site of Silchester. The Silchester *Insula IX* project (www.silchester.rdg.ac.uk/index.html) provided the ideal test-bed for a Virtual Research Environment (VRE) project, because key to the smooth running of this large, urban excavation was the Integrated Archaeological Database (IADB, www.iadb.org.uk/index.htm). Used for recording, analysis, archiving and online publication of archaeological finds, contexts and plans, the IADB allows integrated access to all aspects of the excavation record. It was used to populate the VRE with data generated by a complex urban excavation.

The VERA Project set out to:

1. Investigate how digital input could be used to enhance the process of documenting, utilizing and archiving excavation data.
2. Create a suitable Web portal to provide an enhanced user experience.
3. Develop tools that could be integrated with existing practices of research archaeologists unfamiliar with VREs.
4. Test the tools in a real world setting.

UCL's role was to ensure that the needs of the archaeologists and associated specialists remained at the heart of developments.

The VERA IADB usability study was carried out at the 2007 VERA winter workshop at Reading. The development of the IADB has always been driven by its users and has developed alongside their working practices. However, this was the first time that user reactions to the IADB had been formally documented. Participants at the workshop were divided into two groups:

- those with no (or little) experience of using the IADB, designated 'novice users'
- those who have experience of using the IADB in their work, designated 'experienced users'.

The usability study provided the team with useful information about user perceptions, plus details of the typical tasks carried out by archaeologists and associated specialists. The novice users felt that they could quite quickly get to grips with the system; the experienced users carried out a wide range of tasks using the IADB and used it at (almost) all stages of various projects.

The Silchester project utilizes the skills of a large and geographically dispersed group of specialists. Each specialist uses the IADB for varying purposes, and one of the aims of the VERA Project was to enhance the ways in which each

researcher uses it. Interviews were carried out to explore how the existing users organize their work; to discuss their experiences of working with the IADB; to find out to what extent the IADB met their needs; and if any changes might make their work easier. The results from these interviews were used and fed into IADB development.

Excavation data has traditionally been entered into the IADB through manual digitization, usually once the excavation season is over. A key aim of the VERA Project was to investigate the use of information technology (IT) within the context of a field excavation and to ascertain whether it may be appropriated to speed up the process of data recording, entry and access. From 2005 onwards, a number of field trials had been carried out at the Silchester excavation, using a variety of digital recording methods, including digital pens and notebooks and hand-held internet browsers. The 2008 field season focused on the use of digital pens for direct digital data gathering. We used fieldwork observations, user needs discussions, a diary study and an end-of-season questionnaire to analyse user reactions to the digital pens.

We aimed to observe how well the digital pens fitted in to the workflow of the site and to record user feedback. The discussions provided the framework for creating the end-of-season review for the digital pens. A diary study was used to gather information about different archaeological roles and the way that they are supported by both digital and analogue technologies. These studies allowed the VERA team to explore how the implementation of new technology affected the workflow of the site. Lessons learnt included the central role of the traditional context-recording sheet and the need for any new technology to integrate with existing workflows. ■

Responses to the new technology

Introducing new ways of working into well established systems can be problematic, especially if the changes include the introduction of unfamiliar technology. UCL's involvement in the VERA Project illustrates how user case studies, analysis and feedback were used to develop recording systems and virtual research environments that fit into the current workflow of archaeologists and associated specialists.

The digging season was short – six weeks in the summer of each year – and we studied the dig for three years. This gave us an unusual opportunity to study change over time. Initially, digital methods were only trialled in a small part of the site. We found that they therefore seemed risky and abnormal to most people, and, thus, the methods, and we, were treated with

suspicion. Most people were relatively negative about the use of digital technologies – such as digital pens and paper – for finds recording, as opposed to traditional printed context cards. They also felt they had suffered from a lack of training. The following year, digital technologies, predominantly digital pens, were used throughout the site, and we provided training in their use, as well as feedback on what we had learnt the previous year. Diggers became more positive and began to understand the aims of the study, becoming more open to possible changes. In the final year, further improvements were made to the way digital data was entered and maintained as a result of user feedback, and, when other technologies, such as GPS, were introduced, they were adopted much more readily than we might have expected. Users could understand how their feedback had been integrated into the use of technology and that while systems were not perfect, they had improved, and we had made every effort to act upon user comments as far as possible. As a result, they became noticeably more positive about the use of digital technologies in each year of the study. This shows how important it is that users can see how their feedback has been used to improve a system: if they can see progress, it appears that they will make an effort to support the system they created. If it is not exactly what they would have wished, they will make an effort to deal with pragmatic decisions, if they can understand the reasons for them. In the case of Silchester, they understood that the cost of producing a fully digital recording system would have been prohibitive and were, thus, willing to work with a compromise – a semi-digital solution, which, nevertheless, resulted in faster, more accurate data entry than had been possible using manual recording.

Conclusion

It is clear, therefore, that we cannot, and must not, try to tell users what they ought to like, need or use. We also cannot expect people to abandon working practices instantly when they have suited them well over many years and, in some humanities fields, generations. As we saw at Silchester, if users are consulted, and researchers take the time to understand their working culture and how digital resources fit into it, there is the possibility that attitudes to, and levels of, digital resource use may change. However, we must ensure that users know what they need, to complete their work successfully. If digital resources fit well with what they want to do with them, users will adopt them. Attitudes to digital resources have changed massively in the last

decade, with far greater use of the internet for information seeking and the widespread uptake of resources, such as digital reference resources and e-journals. This is surely because they fit well with what humanities academics would like to do. For example, e-books have recently become more popular, because a new generation of digital reading devices are as light as a paperback, with screens that are more comfortable to read from than earlier e-readers. Thus, users are far more likely to adopt them, because they fit well with their usual reading behaviour and have notable advantages, such as the ability to carry several hundred 'books' in a small, light device.

The aim of those of us designing resources in digital humanities, therefore, remains analogous to this. We must understand the needs and behaviours of users. As a result of this understanding, we must design resources that fit well with what our users already do, while providing advantages in terms of convenience, speed of access, storage capacity and innovative information tools that digital publication affords. If we do so, there is every chance that such resources will be used and will help to make possible new kinds of scholarship that would be inconceivable without digital content, tools and delivery mechanisms.

Bibliography

Attfield, S., Blandford, A. and Dowell, J. (2003) Information Seeking in the Context of Writing: a design psychology interpretation of the 'problematic situation', *Journal of Documentation*, **59** (4), 430–53.

Barrett, A. (2005) The Information Seeking Habits of Graduate Student Researchers in the Humanities, *The Journal of Academic Librarianship*, **31** (4), 324–31.

Bates, M. (1989) The Design of Browsing and Berrypicking Techniques for the Online Search Interface, *Online Information Review*, **13** (5), 407–24.

Bates, M. (1996) Document Familiarity in Relation to Relevance, Information Retrieval Theory, and Bradford's Law: the Getty Online Searching Project Report No. 5, *Information Processing & Management*, **32** (6), 697–707.

Bates, M. (2002) The Cascade of Interactions in the Digital Library Interface, *Information Processing & Management*, **38** (3), 381–400.

Blandford, A. and Attfield, S. (2010) *Interacting with Information*, Morgan & Claypool.

Blandford, A., Rimmer, J. and Warwick, C. (2006) *Experiences of the Library in the Digital Age*, http://www.uclic.ucl.ac.uk/annb/docs/abjrcwCCDTpreprint.pdf.

Brown, S., Nonneke, B., Ruecker, S. and Warwick, C. (2009) Studying Orlando's

Interfaces, paper presented at *SDH/SEMI 2009*, Carleton University, Ottawa, Canada.

Brown, S., Ross, R., Gerrard, D. and Greengrass, M. (2007) *RePAH A User Requirements Analysis for Portals in the Arts and Humanities*, Arts and Humanities Research Council.

Doughty, G. (2001) *Teaching and Learning Technology Support Network*, www.elec.gla.ac.uk/TLTSN/.

Ellis, D. (1993) A Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences, *Journal of Documentation*, **49** (4), 356–69.

Grudin, J. and Pruitt, J. (2002) Personas, Participatory Design and Product Development: an infrastructure for engagement, *Proceedings of Participation and Design Conference (PDC2002)*, pp. 144–61, http://research.microsoft.com/en-us/um/people/jgrudin/.

Higher Education Funding Council for England (2011) *Assessment Framework and Guidance on Submissions*, HEFCE.

Huntington, P., Nicholas, D., Williams, P. and Gunter, B. (2002) Characterising the Health Information Consumer: an examination of the health information sources used by digital television users, *Libri*, **52** (1), 16–27.

Jones, C., Ramanau, R., Cross, S. and Healing, G. (2010) Net Generation or Digital Natives: is there a distinct new generation entering university?, *Computers and Education*, **54** (3), 722–32.

Keating, J. and Teehan, A. (2010) Appropriate Use Case Modeling for Humanities Documents, *Literary and Linguistic Computing*, **25** (4), 381–91.

Lindgaard, G., Dudek, C., Fernandes, G. and Brown, J. (2005) Attention Web Designers: you have 50 milliseconds to make a good first impression, *Behaviour & Information Technology*, **25** (2), 115–26.

Makri, S. Blandford, A., Gow, J., Rimmer, J., Warwick, C. and Buchanan, G. (2007) A Library or Just Another Information Resource? A case study of users' mental models of traditional and digital libraries, *JASIST*, **58** (3), 433–45.

Martin, J. (1996) The Computers in Teaching Initiative, *Ariadne*, Issue 5, www.ariadne.ac.uk/issue5/cti/.

O'Hara, K., Smith, F., Newman, W. and Sellen, A. (1998) *Student Readers' Use of Library Documents: implications for library technologies*, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Los Angeles, CA*, pp. 233–40, New York, ACM Press.

Rieger, O. (2009) Search Engine Use Behavior of Students and Faculty: user perceptions and implications for future research, *First Monday*, **14** (12).

Rimmer, J., Warwick, C., Blandford, A., Gow, J. and Buchanan, G. (2008) An Examination of the Physical and Digital Qualities of Humanities Research,

Information Processing and Management, **44** (3), 1374–92.

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H., Dobrowolski, T. and Tenopir, C. (2008) The Google Generation: the information behaviour of the researcher of the future, *Aslib Proceedings*, **60** (4), 290–310.

Ruecker, S., Radzikowska, M. and Sinclair, S. (2011) *Visual Interface Design for Digital Cultural Heritage*, Ashgate.

Shneiderman, B. and Plaisant, C. (2009) *Designing the User Interface*, Pearson.

Smith, H. and Hyman, R. (1950) The Biasing Effect of Interviewer Expectations on Survey Results, *Public Opinion Quarterly*, **14** (3), 491–506.

Stone, S. (1982) Humanities Scholars: information needs and uses, *Journal of Documentation*, **38** (4), 292–313.

Tiley, J. (1996) *TLTP: The Teaching and Learning with Technology Programme*, http://www.ariadne.ac.uk/issue4/tltp.

Vandendorpe, C. (2009) *From Papyrus to Hypertext: toward a universal digital library*, University of Illinois Press.

Warwick, C. (2004) Print Scholarship and Digital Resources. In Schreibman, S., Siemens, R. and Unsworth, J. (eds), *A Companion to Digital Humanities*, Blackwell.

Warwick, C., Terras, M., Huntington, P., Pappa, N. and Galina, I. (2007) *The LAIRAH Project: log analysis of digital resources in the arts and humanities. Final report to the Arts and Humanities Research Council*, www.ucl.ac.uk/infostudies/claire-warwick/publications/LAIRAHreport.pdf.

Warwick, C., Galina, I., Terras, M., Huntington, P. and Pappa, N. (2008) The Master Builders: LAIRAH research on good practice in the construction of digital humanities projects, *Literary and Linguistic Computing*, **23** (3), 383–96.

Warwick, C., Terras, M., Huntington, P. and Pappa, N. (2008) If You Build It Will They Come? The LAIRAH Study: quantifying the use of online resources in the arts and humanities through statistical analysis of user log data, *Literary and Linguistic Computing*, **23** (1), 85–102.

Warwick, C., Rimmer, J., Blandford, A., Gow, J. and Buchanan, G. (2009) Cognitive Economy and Satisficing in Information Seeking: a longitudinal study of undergraduate information behaviour, *Journal of the American Society for Information Science and Technology*, **60** (12), 2402–15.

Warwick, C., Terras, M., Fisher, C., Baker, M., O'Riordan, E., Grove, M. et al. (2009) iTrench: a study of user reactions to the use of information technology in field archaeology, *Literary and Linguistic Computing*, **24** (2), 211–24.

Interfaces, ephemera, and identity: A study of the historical presentation of digital humanities resources

Claire Warwick
Department of English Studies, Durham University, UK

Abstract

This article reports on a study of interfaces to long-lived digital humanities (DH) resources using an innovative combination of research methods from book history, interface design, and digital preservation and curation to investigate how interfaces to DH resources have changed over time. To do this, we used the Internet Archive's Wayback machine to investigate the original presentation and all subsequent changes to the interfaces of a small sample of projects. The study addresses the following questions: What can we learn from a study of interfaces to DH material? How have interfaces to DH materials changed over the course of their existence? Do these changes affect the way the resource is used, and the way it conveys meaning? Should we preserve interfaces for future scholarship? We show that a valuable information may be derived from the interfaces of long-lived projects. Visual design can communicate subtle messages about the way the resource was originally conceived by its creators and subsequent changes show how knowledge of user behaviour developed in the DH community. Interfaces provide information about the intellectual context of early digital projects. They can also provide information about the changing place of DH projects in local and national infrastructures, and the way that projects have sought to survive in challenging funding environments.

Correspondence:
Claire Warwick, Department of English Studies, Durham University, Hallgarth House, Hallgarth Road, Durham, DH1 3AY, UK.
E-mail:
c.l.h.warwick@durham.ac.uk

1 Introduction

The nature of the much-speculated-upon contents of the Cambridge University Library tower has recently been revealed. It is not, contrary to persistent undergraduate speculation, filled with Victorian pornography; instead, it contains a fascinating collection of ephemera—books considered insufficiently serious or academic to form part of the main collection (Harper, 2018). These books were stacked in order of the dates they arrived, complete with original dust jackets which, for many years were routinely discarded, as was common in most academic libraries at the time (Tanselle, 1971, 2003). Yet this long-disregarded material now provides historians with information about the commercial and social context of such texts, complete with an incontrovertible timeline against which they can be dated. The current status of the digital interface is similarly precarious. Like dust jackets, they are often regarded as functional ephemera to be discarded when worn or outdated. This is in some ways understandable. As Drucker (2013) argues, the goals of Human–Computer Interface (HCI)

Digital Scholarship in the Humanities, Vol. 35, No. 4, 2020.

doi:10.1093/llc/fqz081 Advance Access published on 9 December 2019

research are to render the interface invisible, the better to facilitate access to digital content. Yet she insists that interfaces themselves are legitimate objects of study in digital humanities (DH).

The following article reports on a study of interfaces to long-lived DH resources and argues that such interfaces, designed when the web was new, also provide information about the nature of the material they surround. Original interfaces, often the products of bespoke design, demonstrate how DH projects visually communicated the meaning and importance of their material. Subsequent changes show how DH led, or responded to, advances in web technologies and interface design conventions.

However, it is impossible to study an object if, as was the case with book jackets, it is not considered worthy of preservation. Yet, while a great deal of attention is paid to digital preservation and curation, both in the DH and Information Studies communities, the question of how, or whether, interfaces should be preserved remains unjustly neglected. As a result, we risk losing a valuable resource for the study of how DH has developed over the last 25 years.

The following study therefore uses an innovative combination of research methods from book history, interface design, and digital preservation and curation to investigate how interfaces to DH resources have changed over time and argues that the preservation of interfaces is as crucial as that of the content to which they provide access.

The study addresses the following questions:

- What can we learn from a study of interfaces to DH material?
- How have interfaces to DH materials changed over the course of their existence?
- Do these changes affect the way the resource is used, and the way it conveys meaning?
- Should we preserve interfaces for future scholarship?

2 Methodological Context

There is, clearly, a profound intellectual resonance for DH scholarship of ideas from textual bibliography and the history of the book. Such scholarship addresses the physical presentation of a text, including type styles, images, white space, paper, and the way it is gathered and sewn together to make the book. Yet for many years the study of literary criticism and the theory of dematerialized text was perceived as separate and perhaps inimical to that of books as objects. Literary scholars such as Geoffrey Keynes once regarded textual bibliographers as antiquarians, who fetishized books as objects, simply cataloging arcane features of printing or binding without any care for the meaning of their contents (Tanselle, 1992, p. 14). Yet, as McKenzie (1999) demonstrates, a study of the materiality of the book provides vital information about its economic and social milieu, and thus about the meaning of the literary text within. Such ideas were a profound influence on Jerome McGann, who applied theories of the New Bibliography and the Social Edition to digital editing. The Rosetti Archive, was, in many ways, a digital instantiation of such theories (McGann, 1983, p. 84).

DH scholars have demonstrated that techniques from the history of the book and textual bibliography may be applied to studies of the materiality of digital texts. Drucker (2002), Hayles (2003) and Kirschenbaum (2001), for example, insist that digital resources have a material nature and that, although digital text is more mutable than print, it should not be thought of as entirely virtual. The way that we consume such resources is, they argue, affected by the physical characteristics and affordances of the machines that deliver them or the server blades or discs on which they are stored. The intellectual links between digital scholarship and book studies are especially well demonstrated by Galey, who uses the phenomenological methods of textual bibliography to analyse the detail of a digital resource by removing 'the veil of code' as he terms it (Galey, 2012).

Drucker (2013) believes that the interface itself should be an object of study. She argues that the goal of HCI and user experience research is to make the interface invisible, allowing users to move through it to the digital content itself. However, just as Tanselle stresses the importance of the physical aspects of printed books, including dust jackets, so does Drucker, who has conducted research both

on the materiality of digital objects and the design of printed books, argue that interfaces are far more than content facilitation devices. The interface functions like the frame of a picture: it may enhance the presentation of the content but may also be a beautiful object in itself. Thus, we should, as critics, look at the interface, not simply through it: we must study not simply the contents of a digital resource, but also the details of its presentation, from the code to the interface that surrounds it.

2.1 Digital preservation and interfaces

The literature on digital preservation and curation[1] is vast and complex (Poole, 2015).[2] Questions of how to preserve the interface and design of the original resources have, however, been largely ignored in such debates, which are dominated by the need to make the content accessible. Despite the involvement of several DH scholars in community-based digital curation initiatives, discussion remains dominated by the paradigm of large data sets in science and social science that, once completed, may be archived, or reused with a generic interface.

This is not an appropriate solution for DH resources. As Galey and Ruecker (2010) have argued, the design of a DH resource is an intellectual argument; its arrangement in digital space represents a particular view of the data. In DH, therefore, it can be argued that presentation is interpretation, thus the design of the front end and of the search experience is as deliberate as that of a physical museum (Schofield *et al.*, 2017). Interface design therefore dictates not only how a digital resource looks but also how it works, and how information may be accessed and comprehended by users. It may also provide clues as to how the resource fits into the longer history of DH, and before that humanities computing, or even literary and linguistic computing.

However, Maron *et al.* (2013) argue that few universities have a strategy for the maintenance and preservation of DH content, not least because it may exist in numerous different places: on individuals' computers; a departmental server; in a library or archive, but far too rarely preserved in an institutional repository. Despite the growth of relevant programmes at iSchools, there is still a relative lack of information professionals with sufficient skills in data curation and preservation. Few DH centres have resources to maintain legacy projects: both they and libraries face a difficult trade-off of time spent on preservation, against that on new, or current projects, which may bring in additional revenue (Open Research Data Taskforce and Jubb, 2017).

As we shall see below, DH interfaces and content often evolve and must remain useable over a period of time; yet most are developed, initially, as a result of time-limited external funding. When this ceases, maintenance and updating of content must be funded by institutional resources or unremunerated academic time (Maron *et al.*, 2013). Both commodities tend to be scarce to non-existent. Updating may also require technical skills that many researchers do not possess. They may also not be aware of where to find relevant technical help if it exists in their institution.

When an academic hands over the final manuscript of her book to the publisher her work is at an end. But the long-term survival of a DH resource continues to impose demands on its creators. The need to find funding both for further development and to ensure existing material remains accessible becomes a constant task for project teams and PIs. Crowd-sourced, or user-generated, content places a further burden on resource creators and institutions, creating an implicit assumption that such a resource is maintained and kept accessible to external contributors. Thus, closing it down becomes a very public act, with concomitant risks to institutional reputation. The projects discussed below are relatively unusual examples of high-profile, long-lived projects. However, many smaller-scale digital projects have been subject to periods where use was low, and funding lower (Warwick *et al.*, 2007). These projects either struggled to survive or underwent what Nowviskie and Porter (2010) have called 'graceful degradation'.

The interfaces to those resources that do survive long term may have to be redesigned many times, to ensure their continued usability. However, this implies that, to return to Drucker's arguments, interfaces are purely utilitarian—features through which, as opposed to at which, we look. The assumption is that old interfaces can be discarded

once they appear dated. This is reasonable in most cases: users distrust digital resources that look dated and want to use the most current technologies (Warwick *et al.*, 2007), thus updating is essential.

A recent account of the problems that King's Digital Lab has faced in securing the future of over a hundred legacy projects is especially instructive in this context. Smithies *et al.* (2019) describe a landscape in which many projects are over 5 years old; most have no further funding to make possible updating and maintenance; and those which are built on aging systems and servers may represent a critical risk to the security of university networks under increasing threat of cyber intrusion. The obvious solution in cybersecurity, and financial terms, is to turn off such resources, but to do so means losing valuable, perhaps unique, digital content. The authors describe the complex process by which they evaluated each project, and outline their rationale for deciding how and whether to preserve its content. They have found that each solution must be bespoke: in some cases emulation is proposed, in others preservation of the data alone. This is akin to rebinding a book to preserve its content when the original binding is beyond repair. Librarians are also aware that most readers simply want efficient access to the book, of which thousands of identical copies may exist. However, as books become older and rarer, some scholars wish to use the book in its original binding if possible, with the dust jacket (if there was one) intact. Hence, Tanselle (1990, pp. 9–12) warns librarians about the implications of heedless rebinding because once discarded the information that original bindings and jackets provide about their historical, social, and literary context may be lost forever.

Thus, we might argue that the original design, and subsequent reskinnings, of a digital resource should also be preserved and studied because of the information that they provide about the social and intellectual context of the digital resource. This means that it must be possible for users who are interested in investigating earlier interfaces to be able to find them. It must also be possible to determine how a contemporary digital interface may differ from its original design or functionality.

Hitherto, this has been neglected in discussions on digital preservation. For example, at no point in their otherwise fascinating article do Smithies et al. (2019) discuss whether they considered it important to try to rescue or preserve original interfaces, or what is lost if this proves impossible. This is hardly surprising since digital resources have such a relatively short history. Similarly, it was not until the nineteenth century that scholars became aware of the importance of the physical presentation of rare books and began to be interested in historical bibliography (Tanselle, 1992). With the relative maturity of digital resources in the humanities and cultural heritage, we are now reaching an analogous moment in digital preservation.

3 Experimental Design

The following research adopts a case study approach to a study of the interfaces to DH resources, analysing a sample of projects and their progress over time, in detail. The sample is as follows:

- The Women Writers Project—Brown University and subsequently Northeastern University
- The Valley of the Shadow Project—University of Virginia
- The William Blake Archive—University of Virginia, University of North Carolina
- Proceedings of the Old Bailey Online—Sheffield University and Hertfordshire University
- Digital Images of Mediaeval Music—Kings College London and Oxford University
- The Oxford Text Archive—Oxford University
- Virtual Seminars for Teaching Literature—Oxford University

The reasons for choosing these projects are largely pragmatic: to reach a detailed understanding of interface development over an extended period it was important that resources had as long a lifespan as possible but remained accessible and usable. The above projects were established in the 1990s or early 2000s and are still accessible, even if in a somewhat different form; relatively few DH projects with such a long history are still easily available. Nevertheless, this study is not intended to represent a

comprehensive audit of all such surviving projects. Undertaking a larger study of this type could represent the next development of this research, were funding to be granted to do so.

Although DH is now a global field, its antecedents in literary and linguistic computing were largely Anglo-American, based in a small number of universities, some of which are represented in the sample above. Thus, the sample is skewed towards English language resources. It was also important to have fluency in the language of the resources, to gain the most complex possible understanding of them, and their accompanying documentation. However, important work was being done in humanities computing in countries such as Finland, Germany, and Italy during the same time. Thus, future work could be carried out on a sample of projects in collaboration with researchers fluent in such languages.

The method of analysis is influenced by the work of Vela *et al.* (2014) who used the Internet Archive's Wayback Machine to investigate the design history of the Perseus Project. The Wayback Machine was therefore used to identify the original presentation of, then track every significant design change to, the websites in the sample. It is impossible to be certain when every change was made to the resources because, especially in its early days, Wayback Machine captures were relatively infrequent. Nevertheless, this method provides the most comprehensive insight currently possible into interface change over time. Each website and all significant design changes were examined in detail, not only in terms of their visual design but also of their technical functionality, encoding, and markup.

Jakob Nielsen's work has been used throughout the following discussion as context and comparator for the changing design features of the various sites. Nielson began conducting usability studies and publishing articles on web design in 1994, just before the oldest project in this sample was established (Nielson, 1997). His views quickly became highly respected—considered one of the major references for good practice in web design and usability—and remain so today. The articles on Nielson's website therefore provide a consistent reference point against which to compare developments found in the sample projects.

4 Findings

The following section presents the results of the analysis. It takes a dual approach to the presentation of findings, which highlights both change over time and cross-cutting, recurrent areas of thematic interest. The analysis discusses notable aspects of individual projects, but, in doing so, highlights common themes which recur throughout the sample. These are: navigational metaphors; use of colour; questions of scholarly legitimacy; changes in the original user experience; establishing and maintaining visual identity; search and presentation of a complex resource; and the means to ensure the survival of projects amid infrastructural change. A brief account of each project's origins and history is presented the first time it is discussed.

4.1 Physical metaphors for navigation

The Valley of the Shadow Project is one of the most remarkable early digital resources in terms of interface design. It was established by Edward Ayers in conjunction with the University of Virginia Institute for Advanced Technology in the Humanities (IATH), and later, the Virginia Center for Digital History. An experimental website was created in 1994, only a year after the Mosaic image-based web browser was released. The first archived version is dated 1999. It describes itself as "a digital archive of primary sources that document the lives of people in Augusta County, Virginia, and Franklin County, Pennsylvania, during the era of the American Civil War". (http://valley.lib.virginia.edu/VoS/usingvalley/valleyguide.html)

The site's creators were aware that users might be unfamiliar with scholarly websites and sought ways to support this process. On the homepage, they provide instructions which visitors are expected to read before progressing further. This text is placed lower on the page than the main navigational links, and thus requires users to scroll down to find it, suggesting that advice from usability experts to avoid doing so may not yet been widely known (Nielson, 1997).

The most innovative device to aid novice users is found once the user has clicked down another level to an intermediate gateway page. On the left-hand

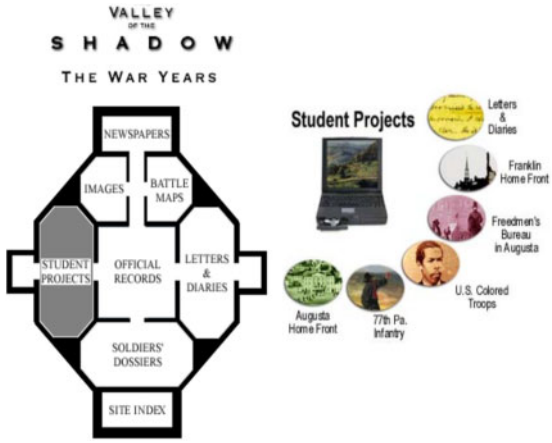


Fig. 1 The floorplan of Valley of the shadow

side of the page is the image of a plan of a building, in which each room is labelled with a different topic (Fig. 1).

This is, one might argue, the ultimate in skeuomorphic design, where the digital interface mimics the appearance of a physical phenomenon. While many digital resources are called archives, in this case, the design metaphor is that of a physical archive building. The rationale for this choice is explained as follows:

> Though people's access to the Web continued to improve, the great majority of people reached the Valley Project through a telephone line and we could not burden the site with large images or elaborate navigation. Will and Michael Mullins, a visiting graduate student from Australia, suggested that we use a floor plan as a way to convey to visitors the sense that they were working with an archive with different "rooms." With a single black-and-white octagonal image, a floor plan by Thomas Jefferson, we provided an overview of the entire archive. The octagon immediately became the most visible symbol of the Valley Project. (http://valley.lib.virginia.edu/VoS/usingvalley/valleystory2.html)

They extend the metaphor further, encouraging visitors to 'take a walking tour' of the archive (at http://valley.vcdh.virginia.edu:80/cwtour.html) to gain insight into its contents. On clicking this floorplan image, the user can find links to a variety of resources, including those created by students. In all subsequent pages, the original navigational image returns and is often complimented by other types of navigation images, for example, that of a compass rose. By 2003, the image has become more complex. The project's opening page now comprises three floors of the archival plan, each in a different colour, as an organizational device to present material from before, during, and after the Civil War (Fig. 2).

The project was archived by the University Library at the end of active development, in 2008. It remains accessible and largely in working order, but the navigation image functionality ceased to work in 2009 when accessed via the Wayback Machine, doubtless because the links to the original image maps were no longer accessible once the resource was moved to a different server. The GIS based animated battle maps, which must have been technically very advanced when they were introduced in 2003, are also no longer accessible. Thus, users can no longer experience the full functionality of the interface in the way that was first intended. We will return to further discussion of this problem below.

4.2 Innovative use of colour

The Women Writers Project Online was established at Brown University in the late 1980s and provided digital texts by female authors, many of which were very difficult to access at that time (https://www.wwp.northeastern.edu/about/history/).

Not surprisingly, the webpage for the Women Writers Project, first captured by the Wayback Machine in 1996 was relatively basic, the only decorative features being black and white images of woodblock prints of the type that might have decorated a Renaissance text. These appear behind the title and decorating the bulleted list that provided navigation (Fig. 3).

However, this visual link to the past of printed books was abandoned relatively quickly; the woodblock-print images have disappeared by mid-1999. The site then adopted a visually experimental design: the navigational hyperlinks are placed at the bottom of the page and highlighted in different

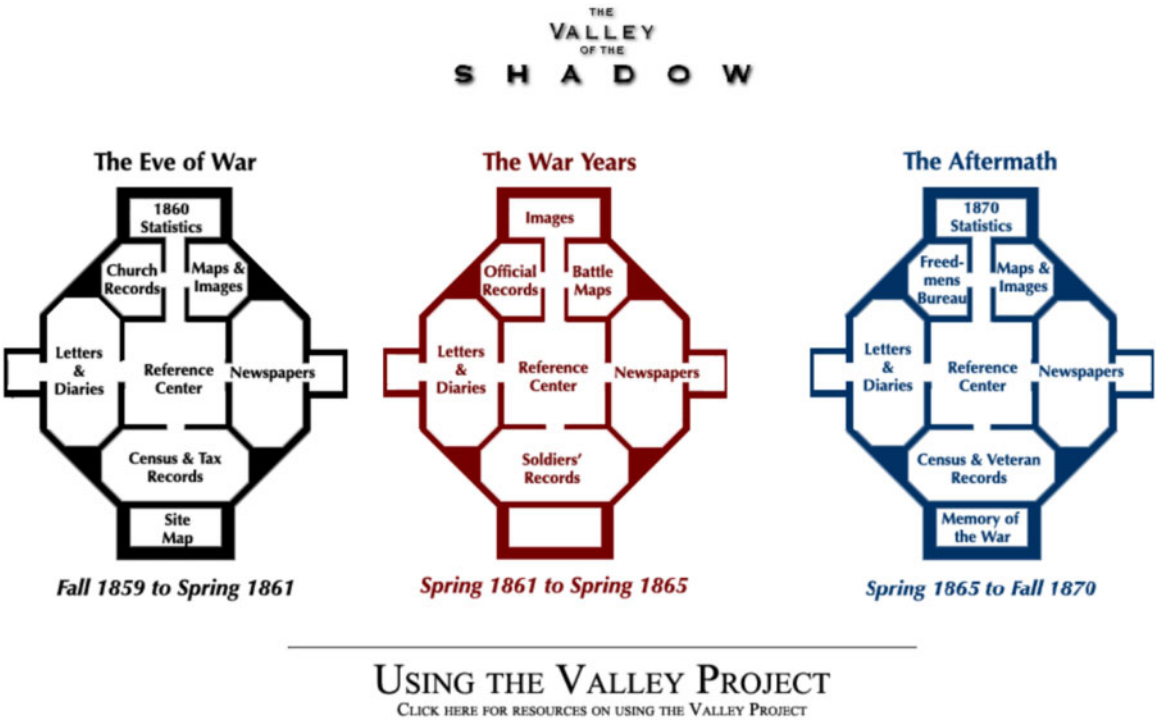


Fig. 2 The extended floorplan

Fig. 3 The original homepage

colours. Any additional hyperlinks on the rest of the site are in the same colour (for example, links to the texts are always dark red) (Fig. 4).

Fig. 4 The coloured navigation menu

By 2006 the page had been redesigned, using Cascading Stylesheets. Instead of being at the bottom of the page, the links, still in different colours, are diagonally stepped across the front page (Fig. 5).

A coloured square marks each link, and the initially black text of the hyperlink changes colour to match the square, on mouseover. Such a use of colour is evidently intended to help users unfamiliar with webpages to distinguish between different parts of the collection.

However, in some cases, innovative designs, intended to improve the user experience for visitors unfamiliar with digital formats, conflicted with the growing standardization in user interface design, especially that of websites: numerous studies demonstrated that users found it easier to follow familiar

patterns (Nielson, 2004b). As a result, projects had to decide whether to abandon such innovative design features.

Text colour, particularly that used for hyperlinks, is an interesting example of this. Some projects in the sample, such as the Blake Archive, initially used the standard of underlined, blue hyperlinks. Others, such as Virtual Seminars, preferred to use coloured text for decorative purposes, often with no functional consistency: this was common for the early web, where individuals designed sites in a way that seemed attractive to them. As users became more accustomed to the concept of clicking on menu items, blue, underlined hyperlinks became less prevalent as a form of navigational signposting (Nielson, 2004a). So, in the case of the Women Writers Project, the use of colour for navigational links was abandoned in 2011, when the resource was radically redesigned.

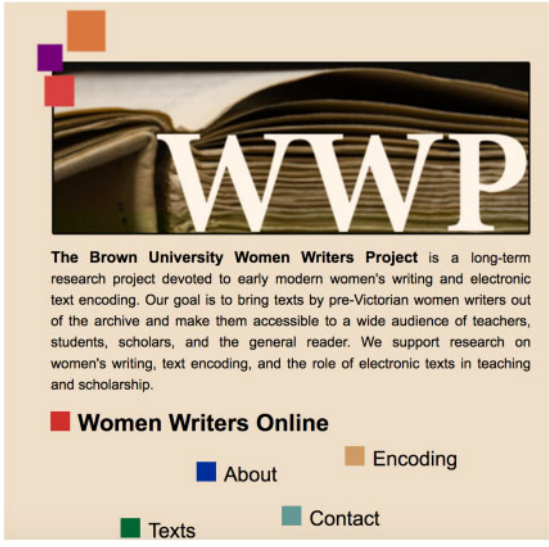


Fig. 5 The 2006 homepage

This new site looks more conventional, in terms of interface design. A horizontal navigation bar provides the main menu beneath the title of the resource, and hyperlinks are now in black text on a pale grey background—the text becomes paler grey on mouseover to indicate its navigational function (Fig. 6).

However, a visual echo of the previous design is retained in the form of a small logo graphic of four different coloured diamonds, at the top left of the page, next to the main title; each word of the title is also in a different colour. This is intriguing, because the use of grey, yellow, and paler grey text on very pale grey background breaches accessibility conventions, because it is difficult to read, especially for partially sighted users (Sherwin, 2015). Thus, tensions between usability conventions and visual identity are played out in the redesign of the site.

The 2011 redesign also features three scrolling images, demonstrating the innovative navigation and visualization features of the interface for the

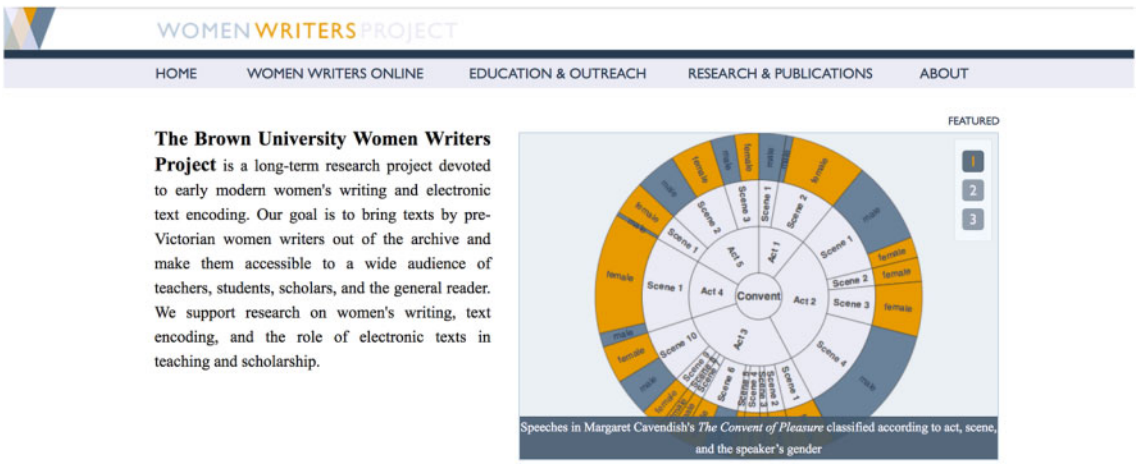


Fig. 6 The 2011 homepage

texts in Women Writers Online. Scenes of a play are displayed in a wheel-like visualization; word frequencies are displayed in a graph reminiscent of a stalactite; and a text features a vertical bar tracking word frequency. All use colour as part of the information presentation. Even if their webpage now appears more conventional, such images demonstrate that the Women Writers Project remains keen to innovate in the interface to its texts.

The developing use of colour as interface design conventions changed can also be seen in design changes made to the Old Bailey Proceedings website. The Old Bailey Proceedings Online project was founded in 2000 as a collaboration between the Universities of Sheffield and Hertfordshire. It 'makes available a fully searchable, digitised collection of all surviving editions of the Old Bailey Proceedings from 1674 to 1913, and of the Ordinary of Newgate's Accounts between 1676 and 1772' (https://www.oldbaileyonline.org/static/Project.jsp). The first Wayback machine capture is in 2003, and thus its presentation is, unsurprisingly, less idiosyncratic than that of older resources in the sample. It is an excellent example of how the navigation of a complex resource has changed. The initial landing page is dominated by blue hyperlink text both from a bulleted top menu and making secondary-level links. The predominance of such textual links over image content encourages the user to explore multiple aspects of this complex resource, either by keyword or theme (Fig. 7).

Perhaps because the original page was so dominated by blue text, when the site was redesigned in 2008, the colour palate had entirely changed (Fig. 8).

Blue is now replaced by red body text, an unusual choice, given the growing awareness of accessibility requirements for those with red-green colour blindness (Nielson, 2004a). It also reverses usual hyperlink convention, making coloured text turn black on mouseover.

4.3 Navigation of a complex resource

Despite changes in the colour, ease of navigation of the resource is always stressed; the same menu appears both across the top of the page and on the left-hand side. The sense of complexity and richness continues when the search link is clicked. As Fig. 7 shows, the original website includes search in the top-level navigation menu: there are nine main top-level links, with a further four secondary links to resources about different communities in London. When the page was redesigned (Fig. 8), a simple search box appeared on the top right of the

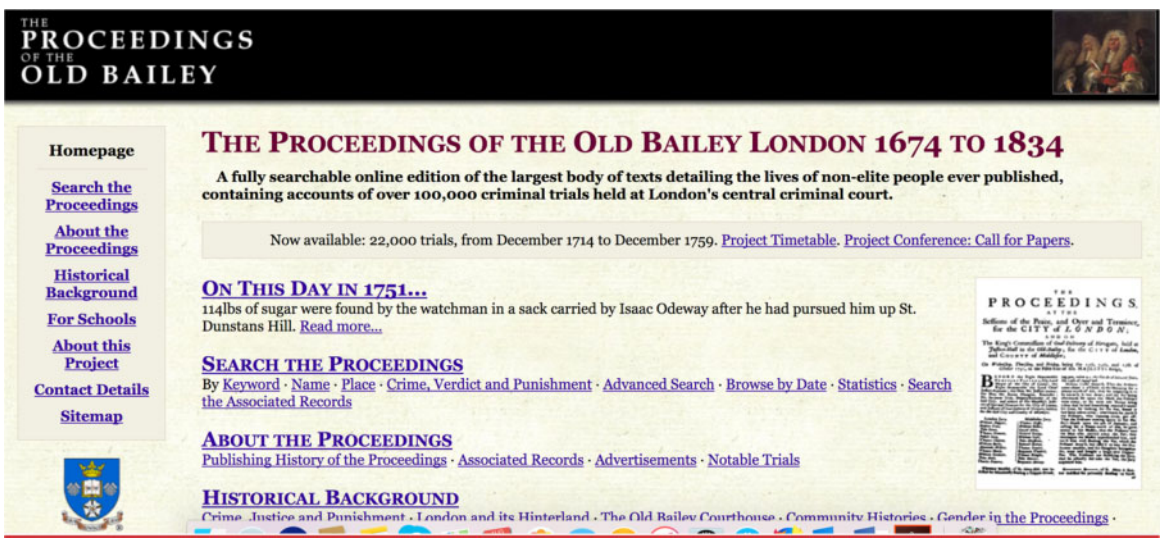


Fig. 7 The original homepage

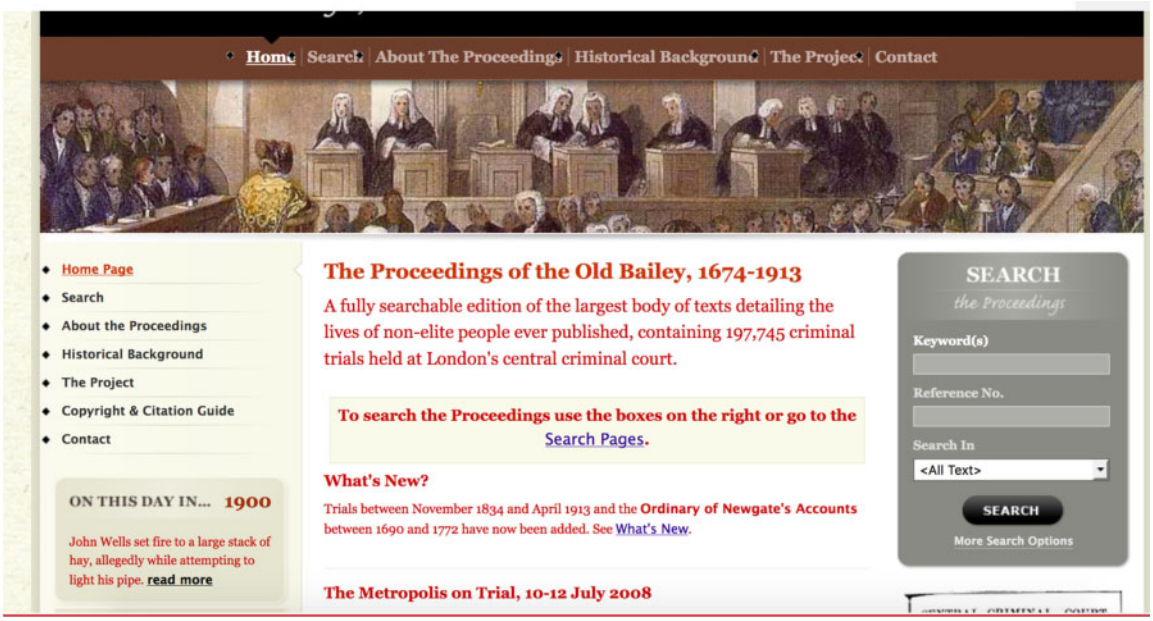


Fig. 8 The 2008 redesign

main page as well as providing a menu link to the very complex advanced search box, with ten different categories made possible by JavaScript enabled search boxes. It also provided help links for each category and a link to a tutorial and help for the whole site (Fig. 9).

An advanced search function can be found in almost all the projects in this sample. But the complexity and visual dominance of search in the Old Bailey Proceedings Online is unusual. In the early period of web design, interfaces and search functionality were usually designed by the project team, many of whom were academic experts in their field. They tended to assume that users would carry out the same complex queries as expert researchers, although, by the mid-2000s, research on DH resources suggested that most users welcomed simplicity, and rarely used advanced searches (Rimmer *et al.*, 2008; Warwick *et al.*, 2008). Nevertheless, the search function of Old Bailey Proceedings Online remains highly complex. The message communicated initially by the multiple blue hyperlinks thus persists, despite changes in interface design. If this reflects a wish to make the variety and complexity of the content accessible, then the decision taken is a good one: Old Bailey Online is one of the most widely used resources in DH.

4.4 Scholarly legitimacy

The Blake Archive was founded in 1996 to provide access to digitized versions of the poetry and images created by William Blake. Given the technical challenges of displaying images on the web at that time, this was an especially pioneering project. Like the Valley of the Shadow project, it was initially produced in collaboration with the IATH and subsequently with the Carolina Digital Library and Archives (CDLA) (http://www.blakearchive.org/staticpage/archiveataglance). The first archived capture appears on the Wayback Machine in 2001.

The opening page is a typical static webpage of its time; it is long and requires the user to scroll through quite dense and substantial amounts of information. Thumbnail images from Blake's original texts are used at the bottom of pages, framed in blue; presumably, this is to provide a visual hint to users of their navigational purpose (Fig. 10).

The landing page initially presents the user with detailed information about the project and its

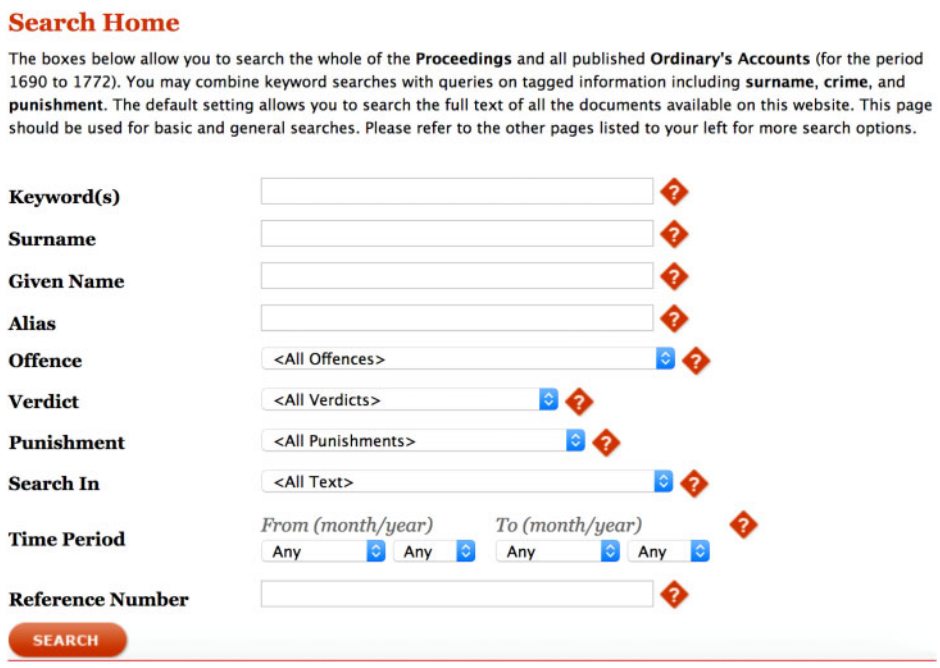


Fig. 9 The advanced search interface

creators, stressing their academic affiliations and track record and providing links to project funders. Although this might seem peculiar to modern users, this information appears before the list of items in the navigation menu (Fig. 11).

Users must also agree to terms and conditions and submit a permission form before they are able to access image content. These represent quite significant barriers to entry to the resource. The image thus created is that of a resource intended for expert, academic users, rather than the interested public.

Having scrolled down the page to the bulleted navigation menu, the first link the user finds is to the information about the archive, rather than to its contents (Fig. 12).

The implication is that the user must be informed about the credentials of those producing the resource, before she will feel comfortable about using it. Somewhat unusually for resources of this age, the main menu of the Blake archive also presents users with a comment link, encouraging them to engage with the editors. This, again, could indicate the assumption of a relatively small community of academic users, who may wish to contact the editors with their opinions about variants and textual presentation.

Although this project is described as an archive, its initial presentation is akin to a printed scholarly edition, in which the editors' names appear on the title page along with information about the scholarly press that produced the text, usually followed by a scholarly introduction to text and the editorial methods used. In a printed book this information is presented before the reader is able to access the text itself and provides reassurance about its scholarly legitimacy and authority.

In common with the older projects in the sample, the design of the Blake Archive provides evidence of its creators grappling with problems of how to establish academic credibility for work presented via what was, at the time, a new medium, of questionable intellectual value. By 2002 the menu had been rearranged, and the hyperlink to the archive itself was top of the menu. This suggests a growing confidence on the part of the Archive's technical team that users

Fig. 10 Navigation menu, showing thumbnail images with blue borders

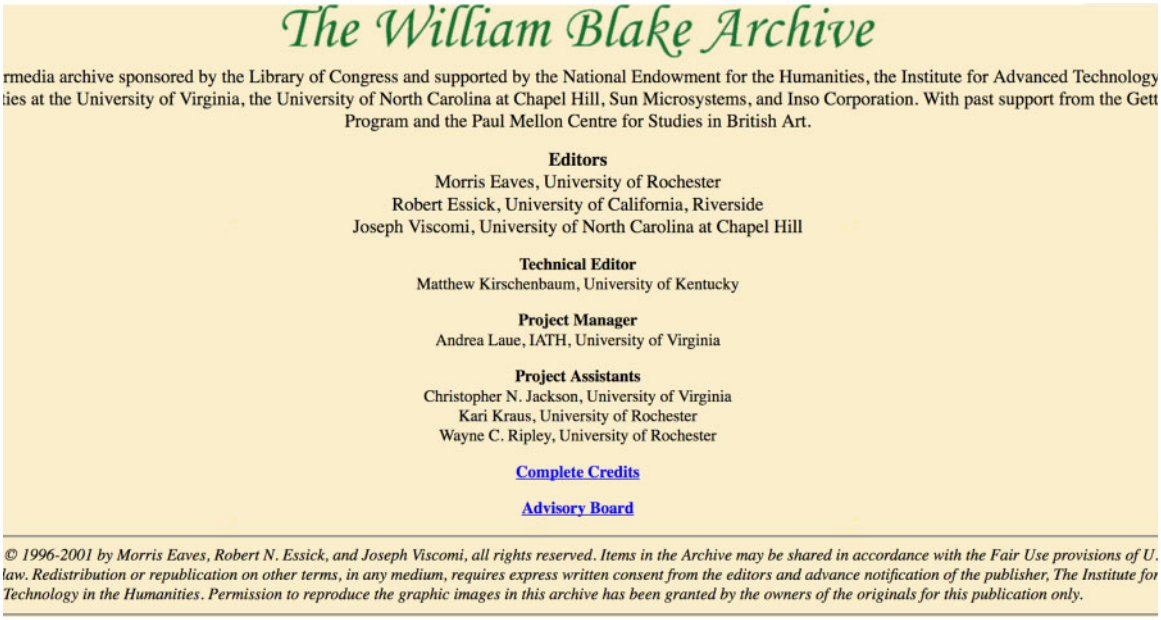


Fig. 11 Blake Archive original homepage

Fig. 12 The navigation menu for the original Blake Archive page

would be familiar with the use of digital resources and need less reassurance about scholarly quality.

Similar evidence of what might be termed scholarly status anxiety is found in several of the projects in the sample. Designers of such early websites stressed scholarly credibility by using the tropes of book publication and design to link their digital scholarship visually to the printed book as a prestigious publication medium. As we have seen, the Women Writers Project briefly used images of woodblock printed capital letters to establish a visual link to the book (Fig. 3 above), but this is even more evident in the initial design of the Valley of the Shadow project. The appearance of the landing page immediately recalls the title page of a book, with text, an image, the name of the author and the publication date centred on an otherwise blank page. The viewer must then click on text underneath the image, to access the main navigation page (Fig. 13).

When Valley of the Shadow was initially created, such splash screens were relatively common, although they fell out of use as design conventions

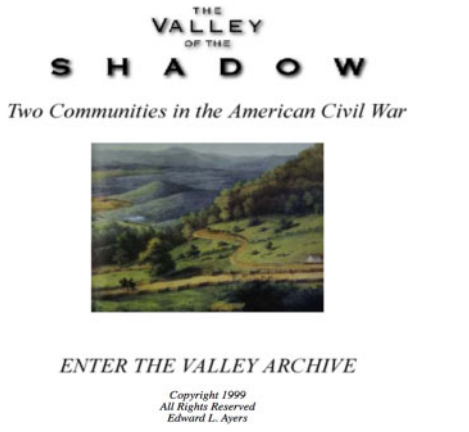


Fig. 13 the original splash screen

developed to avoid the need for users to click down more levels than strictly necessary (Nielson, 2001). Not surprisingly, therefore, the title screen had disappeared by 2002, meaning the link to book design norms was lost.

Once the main website has been reached, the opening page remains image heavy, with a title image featuring members of the valley community against a landscape background (Fig. 14).

The initial design of the project also used other visual devices to situate it within a context of scholarly legitimacy. The navigation menu—a bulleted list—provides prominent links to the university and to other IATH projects, and to a list of awards for digital projects, both from the scholarly community and the web more generally. The page features a lengthy 'story behind the project', which stresses its scholarly credentials, and includes links to a CD-ROM produced by a commercial academic publisher.

4.5 Techincal legitimacy

A similar wish to assert scholarly legitimacy was to lead to very different design decisions, in the case of Virtual Seminars for Teaching Literature. This project began in 1996; it grew from a small pilot project on Isaac Rosenberg's poem 'Break of Day in the Trenches' first launched on the web in 1995 (http://projects.oucs.ox.ac.uk/jtap/proposal.html). The Virtual Seminars webpage provides information about its creators, their contact information, and creation date in the bottom left corner of each page. This was a relatively common practice in early academic websites, although, as design conventions developed, this information, if provided at all, appeared only in the header tag metadata, and via a 'contact us' link.[3] Virtual Seminars also links to a webring, a device of the early web by which pages about similar subjects made mutual links. It also features prominently displayed logos of web awards.

At a time when the web was widely regarded as a junk medium, such devices were important means of establishing intellectual respectability, as was a named contact, clearly identified with their institution (the fact that one of the creators has a doctorate is noticeable). In the mid-1990s, relatively few people knew HTML, or SGML, coding; all pages

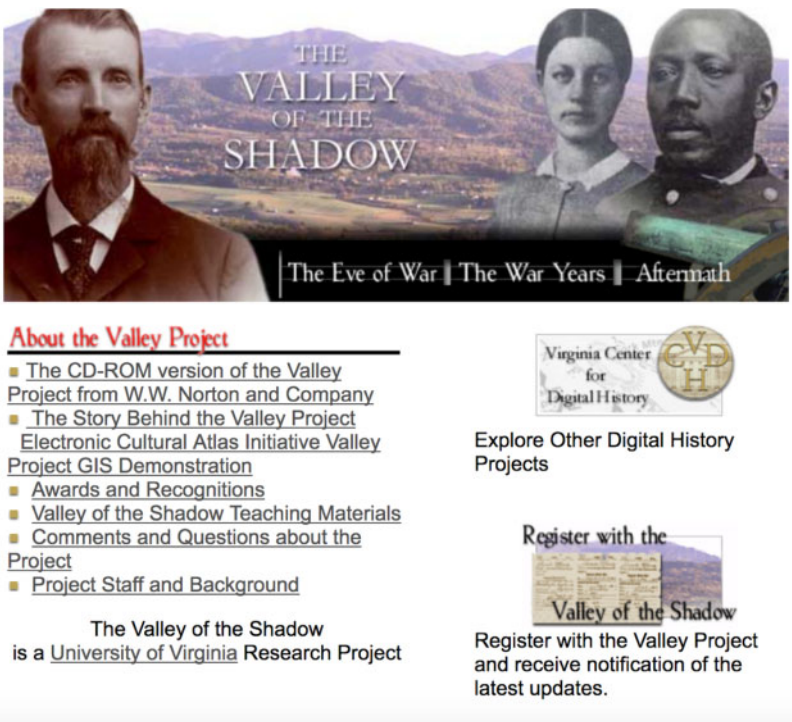


Fig. 14 the original homepage

were marked up by hand, and maintained individually. Even fewer people would have been capable of marrying this skill with the academic expertise to write tutorials on poetry. It is not surprising, therefore, that such individuals took credit for the resource they had created, on its front page.

From the design of the homepage onwards, Virtual Seminars show evidence of being a complex digital resource using highly innovative educational methods and technical functionality, unusual in the late 1990s. For example, it uses layout tables—then a very innovative technique—to organize the content. It is perhaps because use of tables was so new that the borders are still visible; layout tables were soon rendered transparent on most webpages. A table at the top of the page contains the title logo and the one in the middle of the page provides navigational hyperlinks to each seminar (Figs. 15 and 16).

Within these tables, instead of the more usual blue hyperlinked text, Virtual Seminars used images both for navigation and to indicate the content of each section. This practice may now seem commonplace but was highly innovative at the time. The green hyperlinks, which turn brown when clicked, are also part of a complex visual identity: navigation buttons and title images are in a green

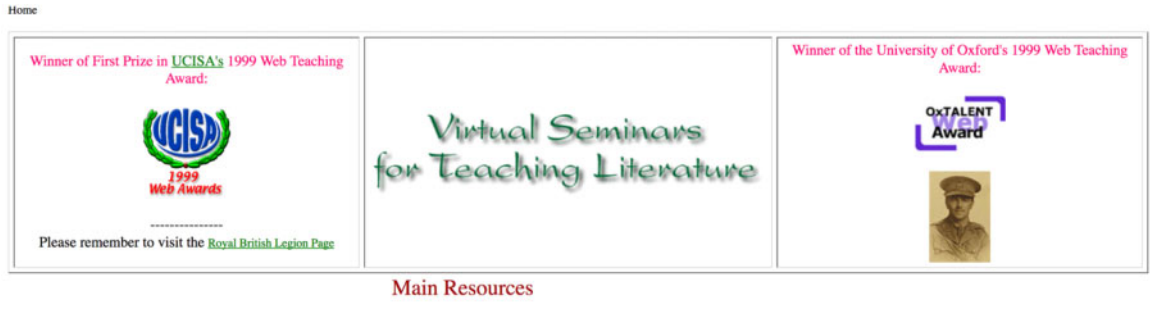


Fig. 15 The top menu table

Fig. 16 The central content organized in a table

Fig. 17 Camouflage patterned title image

and brown camouflage pattern providing a visual link to the subject matter (Fig. 17).

It is noticeable that, unlike many other projects discussed above, Virtual Seminars does not make references to conventions of book design. The devices they employ to assert intellectual respectability are those of developing web technologies themselves. This may be because the resource was designed by humanities computing practitioners, only some of whom were also academics. Thus, the community to which they looked for intellectual legitimacy in the early days of the web stressed expertise in new technologies rather than knowledge of previous tropes of print publication.

4.6 Changing interfaces and unrecoverable features

Digital Images of Medieval Music (DIAMM) is another project that has always been willing to innovate technically, thanks to its continuing relationship with a world-leading DH centre. However, unlike Virtual Seminars, which, having ceased to be updated in 2009, now looks somewhat dated, DIAMM has been constantly redeveloped, and thus provides a fascinating insight into changes in state-of-the-art design and functionality in DH projects.

DIAMM began in 1998: academic research was based at Oxford University, and the technical work was undertaken at the, then, Centre for Computing in the Humanities (CCH) at Kings College London (KCL) and the first version of the page archived by the Wayback Machine is from 2000. This used the latest navigational functionality: frames. This technique is now largely forgotten but allowed part of the page to remain static—usually the menu bar(s)—while the body scrolled. This meant that the user could always see the navigation menu, which sometimes proved extremely useful. Virtual Seminars, for example, used frames in its manuscript study pages, to allow users to display different versions of poems concurrently. Jakob Neilson, however, strongly counselled against the use of frames because of poor functionality and usability (Nielson, 1996).

Like Virtual Seminars, DIAMM initially made innovative use of coloured navigation buttons instead of more basic bulleted lists, as Fig. 18 demonstrates.[4]

The relatively complex colour scheme of the buttons which are rendered either in white, with black text, or in dark blue with red text helps to indicate their functions: for example, the distinctive design of the copyright button- white with blue text and a blue border- draws the user's attention to the need for permission to use the images. However, as Fig. 19 shows, not all of the buttons load well enough to be captured by the Wayback Machine.

This in itself is instructive. As the quotation from the Valley of the Shadow project above shows, in the early days of the web many users accessed webpages via a telephone line and modem, thus images could be slow to load (Nielson, 1997). The use of navigation buttons could therefore be problematic. However much current users complain about slow home broadband speeds, the phenomenon of having to wait while an image loads, sometimes literally pixel by pixel, is something that early Internet users may have forgotten, and more recent adopters never experienced. In this case, therefore, the Wayback Machine emulates the experience of using the early web.

Such an experience also helps to demonstrate some of the particular challenges of making digital images available in the early days of the web, even in compressed form. As we have seen in discussion of the Blake Archive, the original experience of using such websites was necessarily somewhat forbidding to a user. The old Bailey project welcomes users with different suggestions about how to navigate its, mainly textual, content. However, users of resources whose content consisted predominantly of images were compelled to use passwords and copyright permission forms to access rare and precious material, even in digital form. This seems almost contrary to the purpose of digital resource creation,

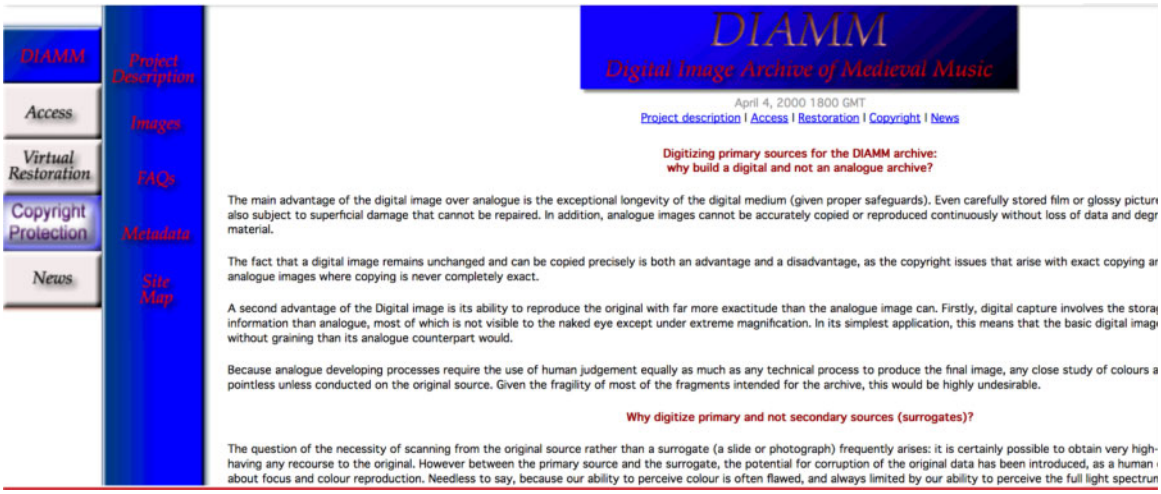


Fig. 18 The original homepage

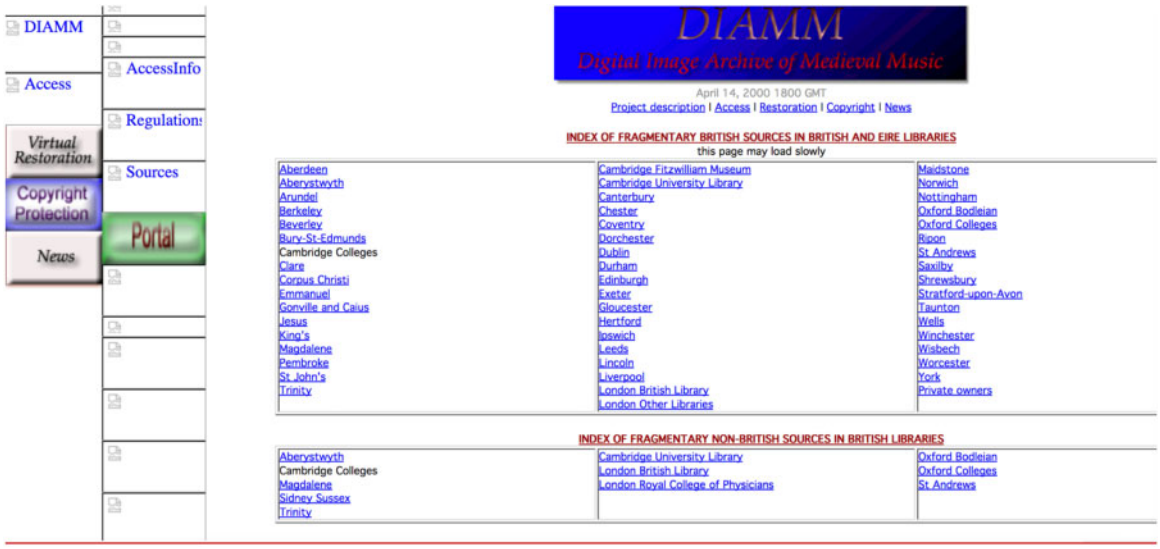


Fig. 19 Partially loaded navigation images

and to the advice Neilson was already providing about how to make commercial websites as easy to use as possible. However, websites such as the Blake Archive and DIAMM were providing expert users with unprecedented access to digital materials and so might reasonably expect them to persist, despite hindrances such as passwords and permissions (Warwick *et al.*, 2008). The only alternative would have been expensive and lengthy trips to see the original manuscripts *in situ*.

Nevertheless, no doubt because of usability considerations, by 2001 the DIAMM front page had lost its frames, although some of the navigation buttons remain, in the middle of the top page. For the first time a logo also appeared which links the project, visually, to its historical subject matter using a red-

and-white colour scheme and an image of illuminated lettering. A more significant redesign can be found in 2004. The page became much simpler, with the content organized using layout tables with black text on a white background (Fig. 20).

The logo is superimposed on a background of music manuscript and the red, illuminated A is accented by red stripes above and below the border. The navigation menu has become more conventional with links in a black, bold, sans-serif text. At the bottom of the menu the login link replaced the need to register for image content, making access to the resource far simpler.

As Fig. 20 shows, the table itself, seen via the Wayback Machine, renders incorrectly; the left-side navigation overlaps the title image. This demonstrates another historical problem of interface design—whether to make table dimensions fixed or variable. Fixed dimensions allowed designers more certainty about how their page would look but might annoy a user with a non-standard sized screen. The aim was to make the page as accessible as possible for the browsers then available, rather than to guarantee future-proofing. In 2004 few websites were designed for mobile access; nobody predicted the ubiquity of tablets or large screened smartphones. Both mobile phone screens and computer monitors were smaller than those of today, and of far lower resolution. Indeed, DIAMM advised users to check that their monitor was calibrated correctly to access a high definition image (the assumption being that it was not) (www.diamm.ac.uk/images). DIAMM's images were digitized according to the most rigorous technical standards, in an attempt to guarantee longevity (www.diamm.ac.uk/images), yet no such thought seems to have

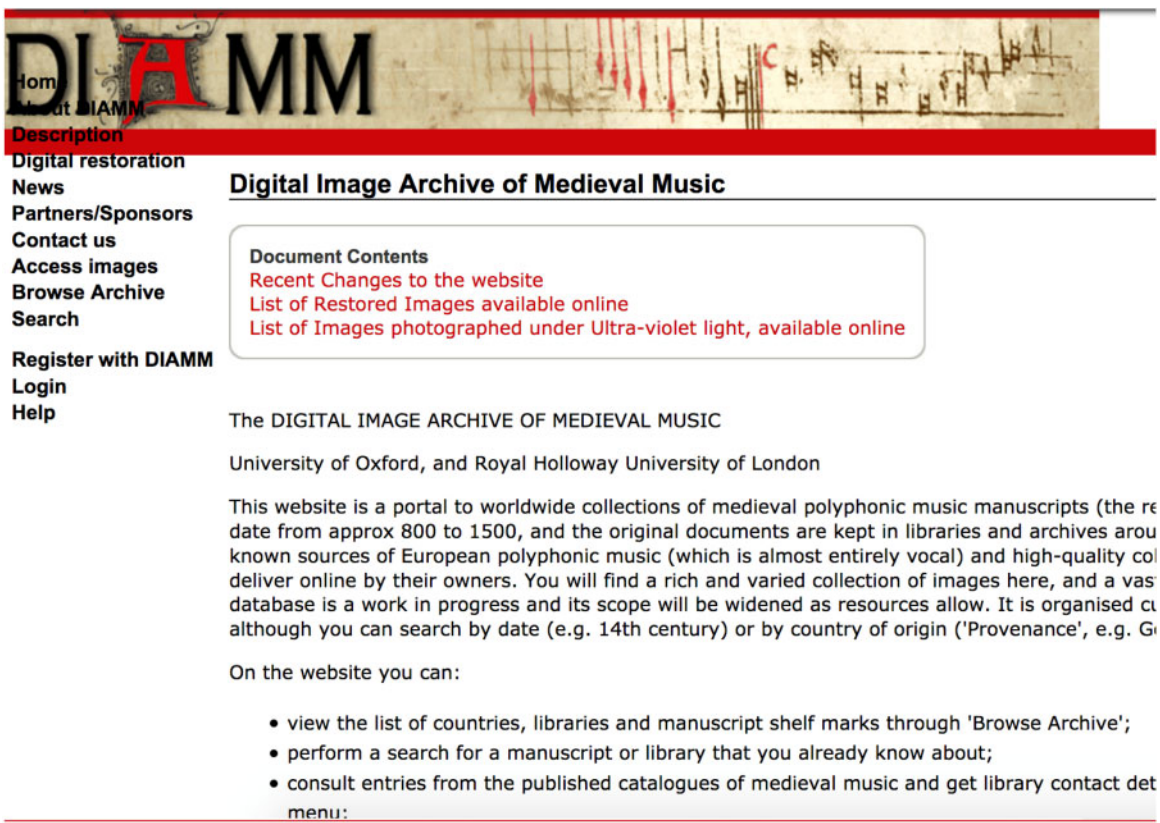


Fig. 20 The 2001 redesign

been given to interfaces in the case of this, or any other, digital project. As a result, we can never fully recreate the experience of viewing an old interface using a modern screen of a higher technical specification than those for which they were designed.

The Blake Archive also shows how the user experience for a long-lived resource may change, but in this case, this is likely to have been as a result of deliberate, radical redesign undertaken when the archive moved from IATH at the University of Virginia- to the University of North Carolina library (https://blog.blakearchive.org/2016/12/12/william-blake-archive-redesigned/). The new homepage is dominated by beautiful, impressively detailed reproductions of Blake's images. Users are immediately immersed in rich image content, in contrast to the original text-heavy pages from which images could be accessed only by completing a permissions form. The resource is now oriented predominately towards visual experience as opposed to that of reading the text (Fig. 21).

The top-level links, in a horizontal strip across the top of the page—a style familiar from corporate websites—guide the user towards complex image-viewing functionality. The links that once constituted the top menu, and those to the more traditional scholarly output of the quarterly journal, the edited print edition, and even the project blog, are now relegated to the bottom of the page, indicating lesser prominence. This suggests a move away from the original emphasis on scholarly discussion and editing, and the abandonment of reference to predominant print formats.

By 2016, the far greater technical capacity to present and manipulate images on the web means that the user experience is completely different from that of the original website. This resource is no longer the preserve of academic specialists; its new design allows any interested user to access digital images of once rare and precious print originals, thereby emphasizing the public mission of digital scholarship. The user of the redesigned Blake Archive must

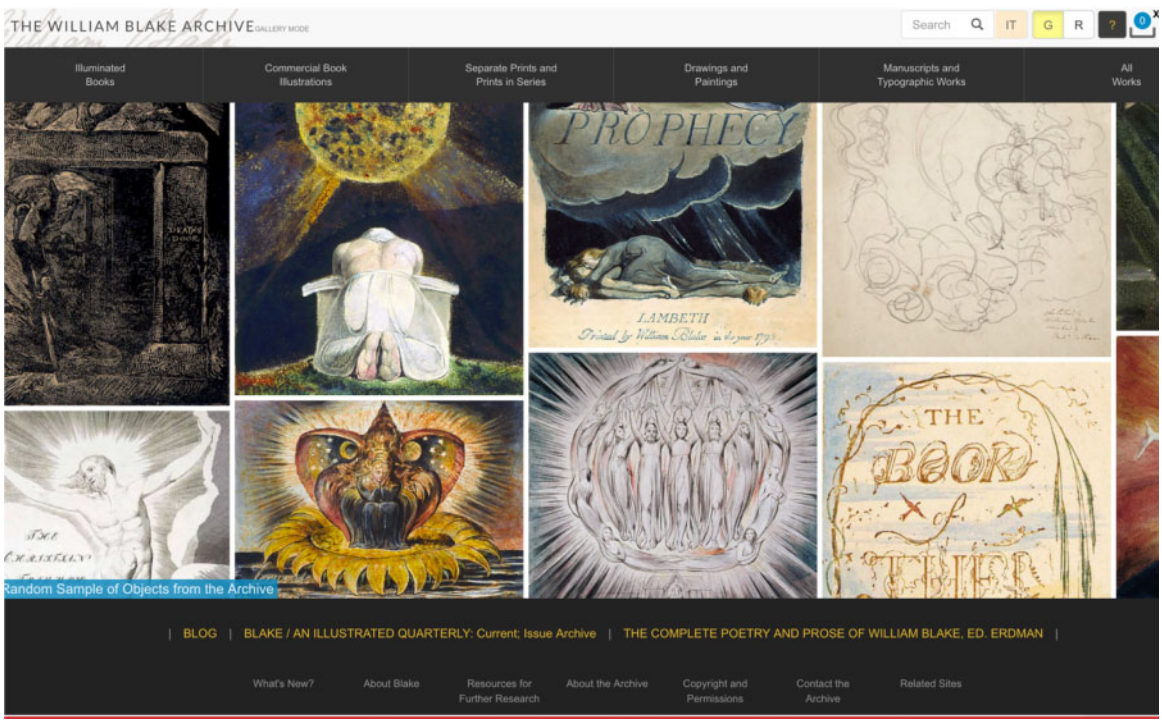


Fig. 21 The redesigned homepage

therefore approach its contents in a very different way from that of the original; it is no longer by and for academics but produced by a library with a strong emphasis on outreach and engagement.

4.7 Visual links to the past of web design

Although it did not undergo such a radical overhaul in visual presentation, the DIAMM website was redesigned four subsequent times, in 2009, 2011, 2012, and 2017. These changes made evident the difficulty of balancing attractive, innovative visual design with optimum usability, as conventions developed and technology changed. For example, all the redesigned homepages remain faithful to the visual identity established in 2004 by retaining the red illuminated capital in the logo. But in 2009 the navigation of the site was simplified using CSS stylesheets: navigation links are now across the top of the page, while the content itself is clearer (Fig. 22).

By late 2012, even more image content was used on the top page: a detailed image provides the background to the title logo, and that of an original manuscript features in the centre of the page. Rather unusually, images of manuscript are also used to frame the main textual content, recalling the earlier practice of using images as wallpaper on early websites. This practice had fallen out of favour at a time when images slowed page loading times. The greater use of images to decorate the 2012/2013 site, therefore, may be evidence of the designers' confidence that most users will now use a broadband connection. Instead of making reference to previous conventions of book design, the designers of DIAMM seem deliberately to be alluding visually to the more recent past of web design itself.

Another relatively unusual design decision is the use of a serif font in the main navigation menu, which provides a visual echo of the logo. Fonts such as Times New Roman, familiar from word-processing, were often used in early web design, but were subsequently abandoned in favour of sans-serif fonts, for greater legibility. However, by 2012, improvements in screen resolution were beginning to make this issue one of significant debate (Nielson, 2012). Thus, advances in display technology make possible, once again, the use of the apparently outdated conventions of serif fonts and wallpaper images as a deliberate link to the visual identity and longevity of the resource.

By 2017, however the balance had tipped back in favour of simpler design and greater usability: sans-serif fonts have reappeared; a simple search box is available from the top menu; the high contrast of white text on a black background improves (Fig. 23).

However, the visual link to the historical materials is enhanced: the background to the DIAMM logo now consists of a large panel of scrolling

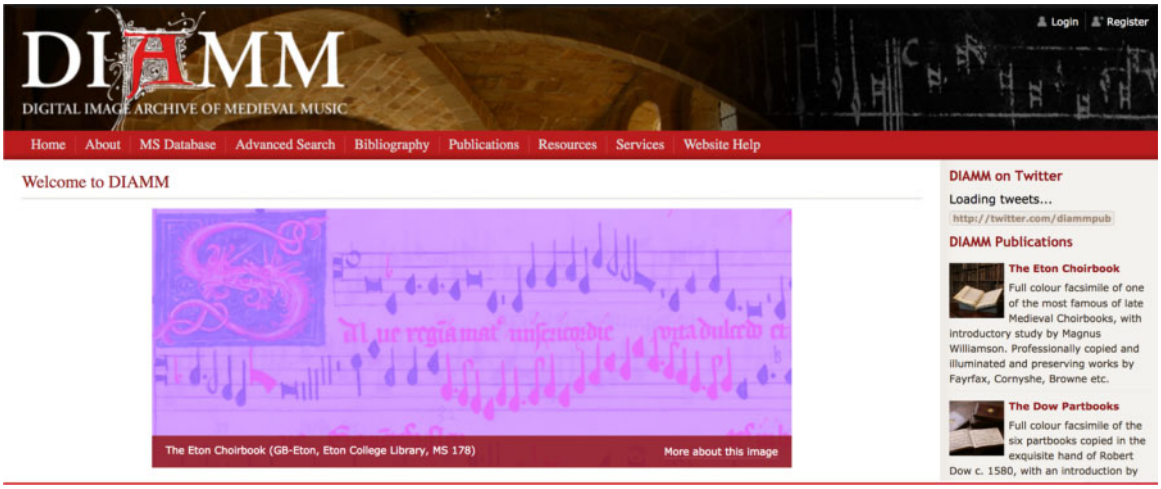


Fig. 22 The 2012 redesign

images, illustrating different aspects of the digitization process. The homepage provides immediate access to information about recent developments and recently added material, showing the influence of the blog style of information design. Red is now less prominent, used only for a line to divide the whitespace, and blue hyperlinks have reappeared—another nod towards the past of web design.

DIAMM therefore provides a fascinating example of a resource that has been regularly redesigned and updated to conform to the latest technical standards of interface design and XHTML coding. However, it is sufficiently confident in its visual design and functionality to make conscious references to its own longevity and part in the history of web design, while remaining true to its visual identity. The development of DIAMM demonstrates the difficulty of balancing often contradictory demands of functionality, usability, and attractive visual design—a balance that must constantly be renegotiated as software and hardware develop.

4.8 Sustainability and infrastructural change

DIAMM's history is, in many ways, what one might wish for all digital resources. It has continued to be funded and developed throughout its life. Unfortunately, this is rarely the case: digital resources may suffer a precarious existence, however well-funded the universities which host them might be. The visual presentation of a digital resource may provide information about the relationship between DH resources and local and national organizational infrastructures.

The Oxford Text Archive (OTA) was founded in 1976, as a repository of digital literary texts. It was based at the Oxford University Computing Services, and its website was first archived by the Wayback Machine in 1997. Its visual identity does not initially reference the university; the small crest in the top left-hand corner is the only Oxford University branding (Fig. 24).[5]

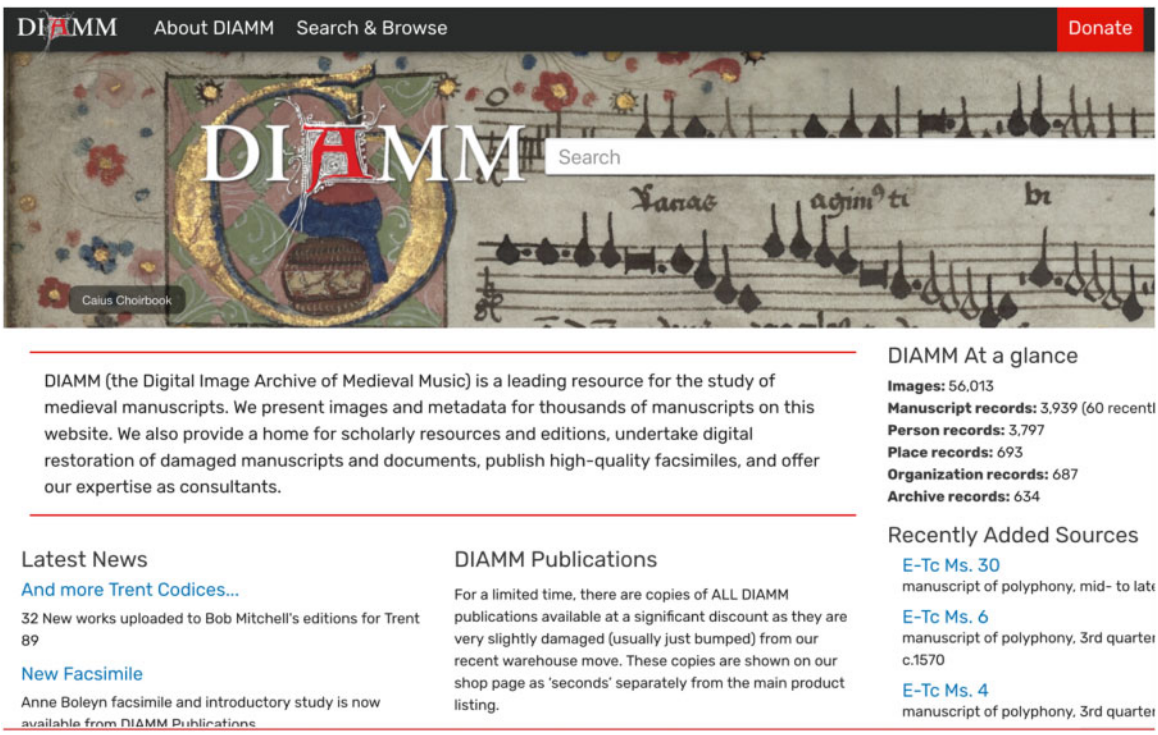


Fig. 23 The 2017 redesign

In 1996, the OTA had become part of the Arts and Humanities Data Service (AHDS)—a national service to preserve and make accessible the data from arts and humanities research, and its visual identity is clearly modelled on the AHDS, using the same font for the title graphic, and the red and black colour palate. The OTA identified itself more with a national infrastructural body, leading innovation in DH resources, rather than its university, with its more traditional brand. Perhaps, as a result, the OTA website adopted innovative design features, such as a circular navigation image, which appeared in 2001 (Fig. 25).

In 2008, following the demise of the AHDS, the OTA reverted to the Oxford University Computing Services, and its website was branded accordingly. As a result, the visual design becomes less innovative: the navigation image is replaced by an already outdated bulleted list and the logo rendered in dark blue, making it less visually arresting (Fig. 26).

In 2012, its name also changed to the University of Oxford Text Archive, emphasizing the OTA's place in the institution, and in 2016, following

Fig. 24 The original homepage

Fig. 25 The 2001 redesign, showing the circular navigation image

Fig. 26 The 2008 site, rebranded in Oxford University style

further reorganization of university IT services, the OTA became part of the Bodleian Library, taking on their brand. At this point top-level links to other DH projects in Oxford, and to the EU-funded CLARIN project appear, replacing those to generic IT services (Fig. 27).

The design therefore highlights the OTA's status as a DH project, rather than generic service, and its place in a larger infrastructural organization, this time on a supra-national scale. Such interface changes are a visual reminder of the vicissitudes that many DH projects have faced over long years of operation. The OTA can be seen developing from one individual's project, to part of a national infrastructure project, which proved to be ahead of its time technologically, then reverting to being part of university computing services and finally declaring its place as part of the library, locally, and an EU network, internationally.

The development of the material originally to be found on the Virtual Seminars resource is an example of another way in which digital resources can survive and be maintained and funded—as part of a new resource collection. The design of the tutorial on 'Break of day in the Trenches' is noticeably different from the other seminars—a simple white page decorated with a top and bottom border of poppies and barbed wire. This suggests that this material formed part of the pilot project from which Virtual Seminars developed. Sadly this is now only accessible as part of the later resource and even Virtual Seminars itself cannot easily be accessed via the Wayback Machine. A search returns the correct link, but, when clicked, the top-level domain of the relevant Oxford University webserver appears instead. This may be due to an automatic redirection script in the original pages. It is fortunate, therefore, that completed digital projects have been archived on a different server—projects.ox.ac.uk- which can be searched via Google. This provides a very good example of why it is inadvisable to rely purely on the Web Archive to preserve the interface history of digital projects. It is unlikely that many other universities have archived completed digital projects, as Oxford did. Had this not occurred the entire history of Virtual Seminars' development could have been lost.

Virtual Seminars themselves were subsequently to form part of the education section of the First World War Poetry Digital Archive: a vast, multifaceted collection of digitized material, which includes an archive of crowd-sourced content from thousands of members of the public. At first glance the design of the World War I Poetry Digital Archive may seem very different from the earlier resource.

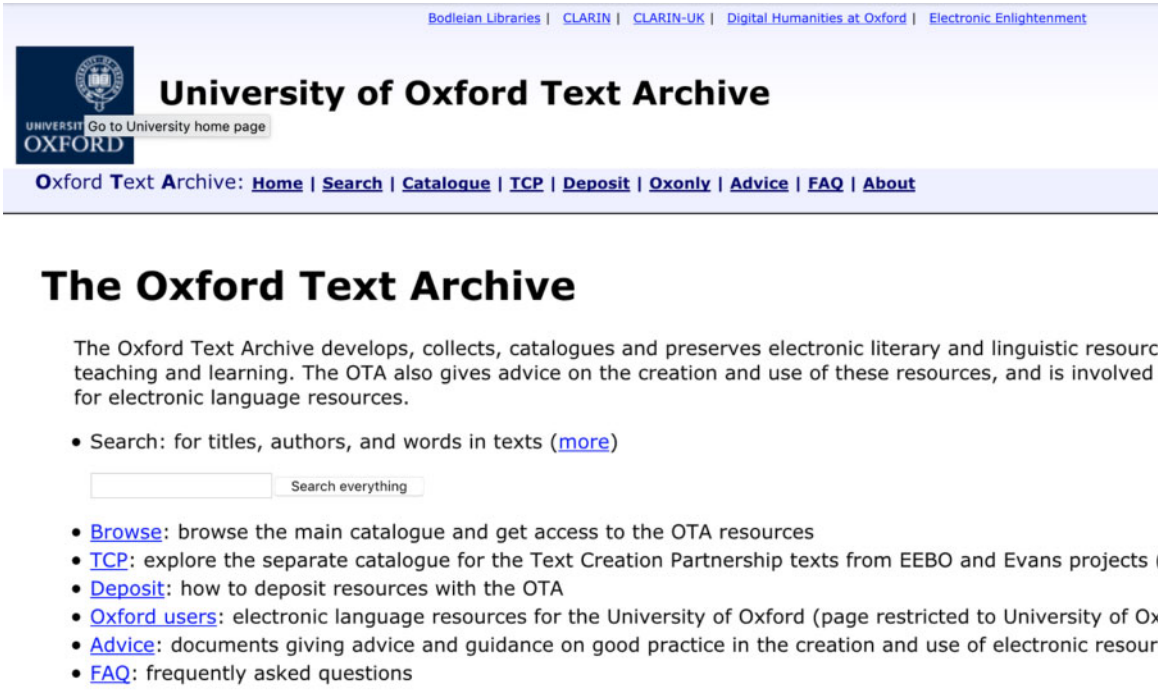


Fig. 27 The 2016 redesign- note the links at the very top of the page

Unlike the clean design of Virtual Seminars, the interface to the World War I Poetry Digital Archive is extremely complex and crowded. Thus, it is important that it has a simple unifying design feature—the use of different shades of pale blue, as background to the white or black text of the navigation menu (Fig. 28).

The homepage is framed by contemporary photographs, organized in window-like squares, some of which, to the bottom right of the page, act as navigation devices, while others, across the top of the page, provide a visual link to the resource's content. These grid structures are reminiscent of the layout tables of the Virtual Seminars project, an impression that is further reinforced when the section on the poets is accessed. Photos of each poet are arranged inside a square frame, which can be clicked to gain access to the works. The only real difference between this visual device and that of the original Virtual Seminars is that the frames surrounding the photographs are outlined in pale blue. Thus, however large and complex the new site may be, its design retains visual clues to its past. Rather than undergoing multiple redesigns of the same website, this content has survived and remained accessible as part of three different resources, each more complex than its predecessor. Ultimately, however, even the First World War Poetry Archive succumbed to the vicissitudes of project-based funding, and does not appear to have been actively updated for around a decade. The contrast between the fate of this project and that of the Oxford Text Archive provides an unwelcome reminder that however much agility digital resources may demonstrate in visual design and technical functionality, without sustainable institutional links, and thus funding, their survival is not guaranteed.

5 Conclusion

The books in the Cambridge University Library tower, complete with their dust jackets, provide a

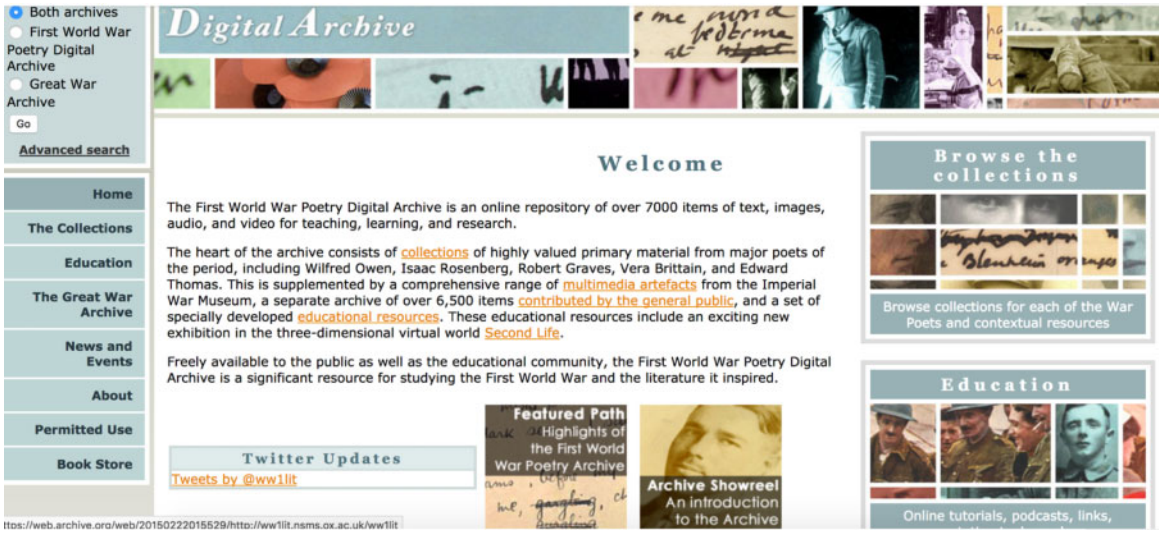


Fig. 28 The original homepage of the World War I Poetry Digital Archive

serendipitous timeline of information not only about book design but also of the changing economy and society in which they were produced. In similar fashion, the Wayback Machine's actual timeline enables us to derive valuable information about the development of DH projects, and the interfaces to them. However, original bindings and dust jackets were once discarded, even by academic libraries. There remains a risk that the same fate awaits early interfaces to digital projects, which may be disregarded as ephemeral wrappers of lesser significance than their content.

This would be deeply regrettable, since, as we have seen, a study of such interfaces provides valuable information about how long-lived digital resources have developed over time and responded to changing assumptions about the scholarly value of digital resources. Visual presentation can also provide information about the changing place of DH projects in local and national infrastructures, and the way that they have sought to survive in challenging funding environments.

A study of how the web presence of such projects has developed reminds us about how changing access conditions, technical standards, and hardware, especially in terms of image display, have affected web design. The projects discussed above were pioneering in their use of an experimental medium, once assumed to be of dubious intellectual merit; thus, it was important to establish their intellectual credibility in the scholarly community. It is perhaps not surprising, therefore, that the interface designs of many early websites referred back to the tradition of printed books. Other sites, such as those produced at Oxford, turn away from such conventions, and, even in their earliest iterations, stress the novelty of this new dissemination medium by showcasing innovative design features, not all of which have survived subsequent design iterations. All of the projects, however, offer detailed information about the intellectual and technical credibility of the project team, and some may feature awards and webrings. Such early websites therefore demonstrate the perceived challenges of convincing potential users of the intellectual credibility of digital resources.

The original versions of these websites were often visually experimental. Project teams could assume no knowledge on the user's part of how to use digital resources, and so were creative in the use of visual navigation devices such as colour, the arrangement of resources in tables or even an image of a floorplan of a physical archive building. Subsequent redesigns might make them look more

conventional because of growing awareness of good practice in user interface design. It is nevertheless significant that the interfaces to all the projects in the sample have maintained some form of visual links with their original identity, by means of fonts, logos, an original colour scheme, or imagery, or the arrangement of navigation and content on the page. However, some redesigns, such as that for the Blake Archive, may change the user experience, and the visual identity, very radically. In that case, it is even more vital that users should be able to access earlier versions of the site because the user experience implied by its design has been so radically altered.

It is still possible to find early versions of many digital resources using the Wayback Machine, and other web archiving applications.[6] However, this is not a perfect solution. Once-experimental functionality, such as imagemaps, frames or animations, or the setup of early web servers, may be incompatible with the Wayback Machine's harvesting technology. This means that some digital resources are already either wholly or partially inaccessible in their original form, and this may become even more of a challenge in future. It is vital that the final form of the project should be preserved, ideally in fully usable form, but it is also important that early versions remain accessible to scholars who wish to study their original presentation and functionality.

It is for all these reasons that we should be conscious of the need to preserve original interfaces and their significant iterations when redesigned. This is a responsibility that project creators will need to consider. Of course, resources must be kept up to date, and thus interfaces redesigned, but at the time of doing so, it would be relatively easy to archive the previous form of the interface, perhaps in an institutional repository. It is far easier, at least, than trying to reengineer a partially functioning former interface, or for researchers to have to make conjectures about how it might have worked, based on intermittent historical screen captures.

It is significant, in this context, that many of the projects in this sample, had, or still have, strong collaborations with libraries, where important work in digital preservation is already being undertaken. Given the relationships that already exist, there is therefore an excellent opportunity for the DH community to work with libraries to preserve original interfaces and their subsequent iterations. It is better to make conscious decisions to archive all versions of sites that are still accessible, as part of an agreed preservation strategy. Not to do so means that we risk losing a wealth of information about the development of the early web and the status of DH resources.

References

Drucker, J. (2002). Intimations of Immateriality: graphical form, textual sense, and the electronic environment. In Bergmann Loiseaux, E. and Fraistat, N. (eds), *Reimagining Textuality: Textual Studies in the Late Age of Print*. Madison: University of Wisconsin Press, pp. 152–77.

Drucker, J. (2013). Reading interface. *PMLA*, **128**(1): 213–20.

Galey, A. (2012). The enkindling reciter: e-books in the bibliographical imagination. *Book History*, **15**(1): 210–47.

Galey, A. and Ruecker, S. (2010). How a prototype argues. *Literary and Linguistic Computing*, **25**(4): 405–24.

Harper, F. (2018). Banished books: Exhibition reveals fabled Tower Collection at Cambridge University Library [Text]. Retrieved 11 December 2018, from http://www.lib.cam.ac.uk/news/banished-books-exhibition-reveals-fabled-tower-collection-cambridge-university-library (accessed 30 October 2019).

Hayles, K. (2003). Translating media: why we should rethink textuality. *Yale Journal of Criticism*, **16**(3): 263–90.

Kirschenbaum, M. (2001). *Materiality and Matter and Stuff: What Electronic Texts Are Made of. Electronic Book Review*(**12**). http://www.altx.com/ebr/riposte/rip12/rip12kir.htm (accessed 30 October 2019).

Maron, N. L., Yun, J., and Pickle, S. (2013). *Sustaining Our Digital Future: Institutional Strategies for Digital Content*. New York: Ithaka S+R, p. 91.

McGann, J. J. (1983). *A Critique of Modern Textual Criticism*. Chicago: University of Chicago Press.

McKenzie, D. F. (1999). *Bibliography and the Sociology of Texts*. Cambridge, UK; New York: Cambridge University Press.

Nielson, J. (1996). Why Frames Suck (Most of the Time). Retrieved 2 January 2019, from Nielsen Norman Group website: https://www.nngroup.com/articles/why-frames-suck-most-of-the-time/ (accessed 30 October 2019).

Nielson, J. (1997). Changes in Web Usability Since 1994. Retrieved 21 December 2018, from https://www.nngroup.com/articles/changes-in-web-usability-since-1994/ (accessed 30 October 2019).

Nielson, J. (2001). 113 Design Guidelines for Homepage Usability (Jakob Nielsen). Retrieved 7 December 2018, from Nielsen Norman Group website: https://www.nngroup.com/articles/113-design-guidelines-home-page-usability/ (accessed 30 October 2019).

Nielson, J. (2004a). Guidelines for Visualizing Links. Retrieved 7 December 2018, from Nielsen Norman Group website: https://www.nngroup.com/articles/guidelines-for-visualizing-links/ (accessed 30 October 2019).

Nielson, J. (2004b). The Need for Web Design Standards. Retrieved 7 December 2018, from Nielsen Norman Group website: https://www.nngroup.com/articles/the-need-for-web-design-standards/ (accessed 30 October 2019).

Nielson, J. (2012). Serif vs. Sans-Serif Fonts for HD Screens. Retrieved 7 December 2018, from https://www.nngroup.com/articles/serif-vs-sans-serif-fonts-hd-screens/ (accessed 30 October 2019).

Nowviskie, B. and Porter, D. (2010). 'The Graceful Degradation Survey: Managing Digital Humanities Projects Through Times of Transition and Decline' in *2010 Conference of the Alliance of Digital Humanities Organizations*. Presented at the 2010 Conference of the Alliance of Digital Humanities Organizations, Kings College London. http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-722.html (accessed 30 October 2019).

Open Research Data Taskforce and **Jubb, M.** (2017). *Research Data Infrastructures in the UK*. Universities UK.

Poole, A. H. (2015). How has your science data grown? Digital curation and the human factor: a critical literature review. *Archival Science*, **15**(2): 101–39.

Rimmer, J., Warwick, C., Blandford, A., Gow, J., and Buchanan, G. (2008). An examination of the physical and the digital qualities of humanities research. *Information Processing & Management*, **44**(3): 1374–92.

Schofield, T., Whitelaw, M., and Kirk, D. (2017). Research through design and digital humanities in practice: what, how and who in an archive research project. *Digital Scholarship in the Humanities*, **32**(Suppl_1): i103–i120.

Sherwin, K. (2015). Low-Contrast Text Is Not the Answer. Retrieved 21 December 2018, from Nielsen Norman Group website: https://www.nngroup.com/articles/low-contrast/ (accessed 30 October 2019).

Smithies, J., Westling, C., Sichani, A.-M., Mellen, P., and Ciula, A. (2019). Managing 100 digital humanities projects: digital scholarship & archiving in King's Digital Lab. *Digital Humanities Quarterly*, **13**(1). http://www.digitalhumanities.org/dhq/vol/13/1/000411/000411.html (accessed 30 October 2019).

Tanselle, G. T. (1971). Book-Jackets, blurbs, and bibliographers. *The Library*, **s5-XXVI**(2): 91–134.

Tanselle, G. T. (1990). *Textual Criticism and Scholarly Editing*. Charlottesville: University Press of Virginia (published for the Bibliographical Society of the University of Virginia).

Tanselle, G. T. (1992). A description of descriptive bibliography. *Studies in Bibliography*, **45**: 1–30.

Tanselle, G. T. (2003). Dust-jackets, dealers, and documentation. *Studies in Bibliography*, **56**, 45–140.

Vela, S., Cerrato, L., Ilovan, M., Li, T., Rockwell, G., and Ruecker, S. (2014). The Biography of an Interface: Perseus Digital Library. In *Canadian Society for Digital Humanities/Société Canadienne Des Humanités Numériques (CSDH/SCHN) Conference*. St Catherine's, Ontario.

Warwick, C., Galina, I., Terras, M., Huntington, P., and Pappa, N. (2008). The master builders: LAIRAH research on good practice in the construction of digital humanities projects. *Literary and Linguistic Computing*, **23**(3): 383–96. https://doi.org/10.1093/llc/fqn017 (accessed 30 October 2019).

Warwick, C., Terras, M., Huntington, P., and Pappa, N. (2007). If you build it will they come? The LAIRAH Study: quantifying the use of online resources in the arts and humanities through statistical analysis of user log data. *Literary and Linguistic Computing*, **23**(1): 85–102.

Notes

1 Digital preservation and curation are often used synonymously, but essentially, the former term refers to preserving and archiving data so that it may be used

again, and curation to a larger range of activities that enhance the data, such as the addition of metadata, migration, and emulation, which may make it easier to discover and use in future.

2 The Digital Curation Centre provides an excellent resources section, listing the various reports and recommendations at http://www.dcc.ac.uk/resources.

3 Nevertheless, DIAMM retains such information in a somewhat abbreviated form on its main page until 2003.

4 This was innovative, but not unique: similar buttons could be found on the British National Corpus website, established in 1996 and based at the Oxford University Computing services (https://web.archiv.org/web/20020808223725/http://www.natcorp.ox.ac.uk:80/).

5 Unfortunately, as with the DIAMM site above, the logo image loads too slowly to be captured by the Wayback machine.

6 A list of these may be found at http://www.dcc.ac.uk/resources/external/category/web-archiving.

WEB ACCESSIBILITY

BASICS & GUIDELINES

Why Is Accessibility Important?

Approximately 20 percent, or 1 in 5 people, in the United States has some form of disability. About 11 percent of college students have a disability. As an institution of higher learning, having an accessible website that provides equal access and equal opportunity for all is essential.

That's why there is a framework of federal laws and regulations, including the Americans with Disabilities Act (ADA) and Section 508 of the Rehabilitation Act of 1973, that require the University to have an accessible website.

Section 508 was updated in 2000 and again in 2017 to reflect the changes in technology and the need for accessibility in regards to the internet. In 2017, Section 508 was updated to align with the Web Content Accessibility Guidelines (WCAG) 2.0, AA success criteria. WCAG itself was updated to version 2.1 in 2018 and all content that conforms to version 2.1 also conforms to WCAG 2.0.

What Is WCAG?

WCAG is a set of technical standards developed under the guidance of the World Wide Web Consortium (W3C), which is the international standards organization for the web. **View a video introduction to Web Accessibility and W3C standards.**

WCAG is a set of 12 guidelines organized under a set of four principles:

Perceivable: provide text alternatives for non-text content; provide captions and other alternatives for multimedia content; ensure content can be presented in different ways with assistive technology without losing its meaning; and make it easier for web visitors to see and hear content.

Operable: all functionality available using a keyboard; providing enough time; content will not cause seizures; multiple ways to help users navigate and find content.

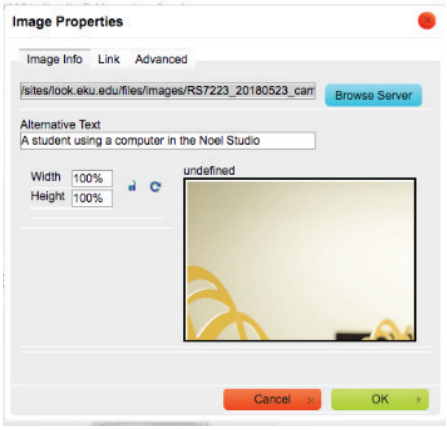
Understandable: text is readable and understandable; content operates in predictable ways; assistance to avoid and correct mistakes.

Robust: content is compatible with current and future user tools.

There are three levels of success criteria within WCAG — A, AA and AAA. Federal guidelines require compliance with level AA, although compliance with AAA success criteria is preferred if possible. W3C points out that even content conforming to a AAA compliance level will still not be accessible to all — there is no "perfect" system to ensure all content is accessible for all people.

Tips to Ensure Your Content is Accessible

IMAGES



All images must have alternative text (alt text). Alt text is required for publishing within the content management system (CMS).

Alt text serves a number of functions, the most important of which is to allow the purpose/presentation of the image to be accessible to those with visual or cognitive disabilities.

Alt text should:

- Equivalently describe the image.
- Describe it succinctly. Try to keep alt text to 16 words or less.
- Be unique. Make sure each alt text on a site is different.
- NOT be redundant (i.e. it should not duplicate already included content.)
- NOT include phrases such as "picture of..." or "image of..."
- NOT include copyright/source information of the image.
- Include supplementary information about a graphic that isn't included within the standard content.
- NOT be generic (i.e. "Student" or "Building")

ALTERNATE TEXT: *

Powell building reflected i

This text will be used by screen readers, search engines, or when the image cannot be loaded.

ALTERNATE TEXT: *

EKU Powell Building Image

This text will be used by screen readers, search engines, or when the image cannot be loaded.

In cases of purely decorative images — images that are not part of the content, but are simply design elements — null alt text (alt="") is accessible and the proper way to continue, since the expectation is that content must stand alone. Images being added by content creators would not meet the definition of "purely decorative" and MUST have alt text.

Charts and other complex images require more than the short description that can be given from alt text. But charts and complex images can actually increase accessibility.

See **webaim.org/techniques/alttext** for more details on alt text.

IMAGES OF TEXT/TEXT WITHIN IMAGES

Text within logos, when appropriate alt text is included, is acceptable. Generally, it's considered a poor practice to include text within an image, especially if the text is important to the understanding within the content. In these cases, the text within the image must be contained within the alt text. A purely image presentation of text or information should never be used.

ICONS

If you utilize icons, ensure that they are:

- Simple
- Are easily understood/well-designed
- Would not be misunderstood because of culture and/or language
- Are not dependent upon color
- Include appropriate alt text

Group Tours are offered for elementary, middle and high school students.

ANIMATION

It's generally considered a poor practice to utilize animation on the web, because animations simply act as a distraction. This includes animated .gif files and Adobe Flash. There's hardly a case for making content more accessible by including animations.

If you do include animation within content, it must:

- Be controllable by the user and/or very short in duration.
- Not move, blink, or scroll for more than five seconds without having the ability to pause, stop, or hidden by the user (**WCAG 2.0 Success Criterion 2.2.2 - Level A**)
- NOT flash more than three times per second.
- NOT be of the type of graphic that can cause a photo-epileptic seizure (**WCAG 2.0 Success Criterion 2.3.1 - Level A**).

COLOR

Because visitors may be colorblind or low-vision, a sufficient contrast between background and foreground (both text and graphical elements) must be present. There must be a contrast ratio of at least 4.5:1 except when:

- Text is rendered at 18pt or 14pt if bold. Here, a ratio of 3:1 applies.
- Text or image is incidental, such as on decorative images or not visible. Photographs also have no requirement for contrast ratio.
- Text is part of a logo/name brand.

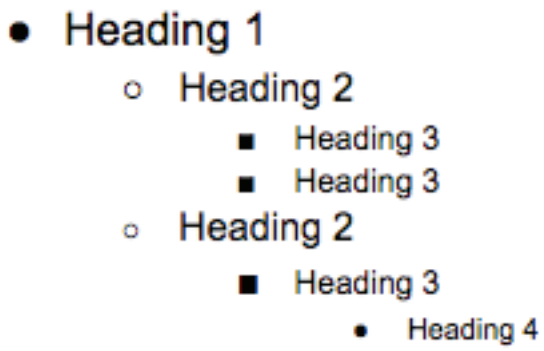
Use the **WebAim tool to check color contrast**.
Color should never be used as a method of communicating content.
Don't say things like "click on the red button."

PAGE TITLES

Concise, unique page titles will ensure all visitors can quickly understand the purpose of a webpage. Page titles are the first element announced by screen-readers.

HEADINGS

Headings within web pages provide structure and should always be correctly applied.
<H1> is the highest heading level and <H6> is the lowest heading level. The hierarchy should always be applied in order, like:



CAPTIONS/TRANSCRIPTS

Video content should always have accurate, synchronized captioning.
YouTube provides closed captioning by default, however, accuracy is not foolproof. Ensure the CC content provided by YouTube is accurate.

Audio content should always have accurate, full-text transcripts as an alternative representation.
If you're including audio-only files, also include a transcript.

LINK TEXT

Link text should be clear and meaningful, so avoid link phrasing that isn't easily understandable. Don't use link text that's unclear, such as:

- Click Here for today's weather
- More information

To visitors viewing the entirety of a site, those unclear links may seem perfectly clear. But to visitors using a screen reader or another alternative browsing method, they're patently unclear. By adjusting the text, a clearer picture can be given, such as:

- Today's weather
- Learn about EKU

Link text should never be empty.

For more information and/or details on structuring link text, see:
webaim.org/techniques/hypertext/link_text

Do not utilize underlines to place emphasis on text, or on text that is not a weblink.

ADOBE PORTABLE DOCUMENT FORMAT (PDF) FILES

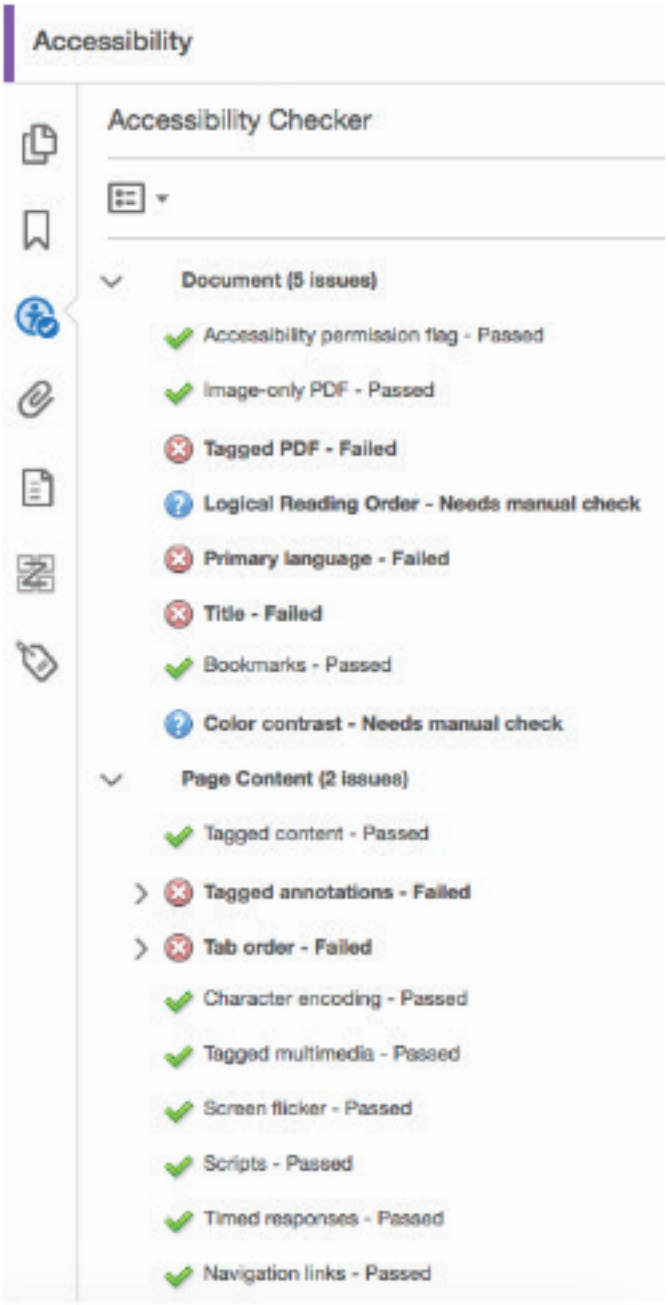
There are a number of reasons why you might want to include PDF files on your website, including:

To maintain design and formatting,
To protect the document from editing,
To allow wide distribution and/or printing in the original, intended format.

And like other content, Adobe PDF files must also meet accessibility guidelines when shared on the web.

DO:

Start with an accessible source document created in Microsoft Word or Adobe InDesign.
Ensure the PDF file has appropriate tags.
Use appropriate nesting in headings. The examples used in the "Headings" section of this document also apply to PDF files.
Include alt text for all images and charts within the PDF, unless they are decorative or redundant/are not part of the content.
Use the Accessibility Checker in Adobe Acrobat Pro before uploading your PDF file. Correct any accessibility issues it finds.
Define the Primary Language and that the file has a Title. The title is different from the filename.



DON'T:

Use a PDF document if there's no essential reason for doing so. If the information could be shared as a webpage, the use that format instead.
Scan documents and create image-only PDF files. Image files are NOT accessible as PDF documents, even if appropriate alt text is added.
Split rows of a table across pages.
Use ambiguous or unclear copy for links.
Use confusing navigation.
Use references to color or other sensory characteristics.
Use tiny fonts.
Use background images or watermarks.

For more information on creating WCAG 2.0 compliant PDF files, see **www.w3.org/TR/WCAG-TECHS/pdf.html**

For more information and/or details on Adobe PDF accessibility mitigation, see **helpx.adobe.com/acrobat/using/create-verify-pdf-accessibility.html**

For a more in-depth assortment of tools, see **www.w3.org/WAI/ER/tools/**

TABLES

Tables should never be used for layout. They should only be used to present tabular information in a grid or matrix format, and there should always be columns/rows that show the meaning of the information within the grid.

While not a requirement, tables can utilize a <caption> element that gives a brief description of the contents of the table.

For more information on utilizing tables, see **webaim.org/techniques/tables/data**.

WEB ACCESSIBILITY CHECKERS

There are a number of free web accessibility checkers and tools available for use to assess the accessibility of websites. They include:

- **AChecker**
- **ATester**
- **Colorblind Web Page Filter**
- **Color Contrast Checker**
- **SiteImprove Accessibility Checker Chrome Extension**
- **WAVE: Web Accessibility Evaluation Tool**

www.eku.edu