

Access, Books, and Digital Collections

On the things I've learnt from a few years teaching students how to "read" a digital instantiation of a book is that without some knowledge of the wider context of that digitization – the platform, database, collection, or archive – it can be difficult to understand why the way digitized books look the way they do. This worksheet enables students to begin to learn what's involved in the digitization of books. In particular, it aims to help students explore the various legal, technological, and economic factors involved in the creation of large-scale digital collections; and also the cultural contexts of representation and access against the hype of universal knowledge. The in-class task, a paper prototype of a digital archive, works particularly well to bring home the difficult choices between sometimes contradictory factors involved in real-life digitizations. The session is part of a second-year undergraduate module entitled "Literature and Digital Culture", following on from discussions in previous weeks about the digital medium, the digital divide and information privilege, and the representation of gender, race, sexuality, and intersectionality in Wikipedia articles.

PART ONE

Please read (see reading list below)

1. On the "universal library": EITHER Paul Hammond, chapter three; OR Marilyn Deegan & Kathryn Sutherland, chapter five.
2. Intersectionality and access: ONE of these: Amy Earhart / Jacqueline Wernimont & Julie Flanders / Adeline Koh.

How to read a digitisation

The questions below will help your digital literacy and will enable you to assess the differences between digital collections of books. This involves a bit of detective work, since some collections don't give full information about themselves, or make it difficult to find: just try to find out as much as you can. Some of this information you might be able to find by browsing the archive website (for example, look for pages labelled "About", "History", "FAQs"). Don't worry if you are not able to answer all these questions in relation to your chosen archive: this itself might be significant in relation to the question of access.

1. Is it a standalone collection or a collection that is accessed via a platform? (e.g. Jisc Historical Texts or Gale Primary Sources are platforms that give access to multiple individual collections or archives).
2. Try reading and/or downloading material to get a feel for each site.
3. Is it paywalled (e.g. via our library), or free to access? Is it commercial or scholarly? What kind of organisation is it? How is it funded or supported?
4. How is it digitised? Is it images or text, or both? Are the images colour or b&w? If it has text, how was the text created? Was it transcribed (copied by hand), or created by automated software (OCR)? Was it edited?
5. Does it tell you about the original source of the material (the library or archive in which the

- physical copy is held)? Does it give you any other details about the individual books or materials?
6. Is there any information about the copyright status of the book images or text?
 7. Is it delimited (e.g. by genre, period, geography, nationality, etc)? Does it tell you how the material was selected?
 8. What terms, language, or imagery does the collection use to describe its ethos or aims?

Choose one collection from these:

- Digital Public Library of America (DPLA)
- Early Caribbean Digital Archive
- Google Books
- Jisc Historical Texts – access to Eighteenth-century Collections Online (ECCO); or Early English Books Online (EEBO)
- Project Gutenberg
- Rosetti Archive
- UbuWeb
- Women Writers Project

PART TWO

The second part of the session is a paper prototype experiment, and enables you to experience the issues in the creation of digital collections, drawing on your research and analysis. In teams, imagine that you're either scholars, or university librarians, or a commercial digital publisher, or a collaboration (hint: it makes a difference), and you are in a position to plan a digital collection. What would it be? You can be as realistic or as utopian as you like! Answer these questions to help you plan and write a rationale for the project.

1. Who is it for?
2. How would it be funded?
3. What limits would you have to place on the collection? What kinds of material would you include (or not include)?
4. How would you present the books? (photographic images? text only? Both? OCR or hand-transcribed?). Why?
5. What kind of access to the material would you or could you allow (free or paywalled? Can users download material or not)?

Further Reading

Cohen, Daniel J. and Roy Rosenzweig, *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web* [detailed guide on what's involved in building a digital resource]

<http://chnm.gmu.edu/digitalhistory/digitizing/4.php>

Carr, Nicholas, "The Library of Utopia", *MIT Technology Review*, April 25, 2012
<https://www.technologyreview.com/s/427628/the-library-of-utopia/> [on Google Books]

Darnton, Robert, "Digitize, Democratize: Libraries and the Future of Books," *Columbia Journal of Law & the Arts*, 36:1 (2012), 1-20

Deegan, Marilyn, and Kathryn Sutherland. *Transferred Illusions: Digital Technology and the Forms of Print* (Farnham: Ashgate, 2009)

Earhart, Amy E., "Can Information Be Unfettered? Race and the New Digital Humanities Canon," in *Debates in the Digital Humanities*, 2012 <<https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfd1e/section/cf0af04d-73e3-4738-98d9-74c1ae3534e5>> [accessed 23 July 2020]

Findlay, Peter, "Commercial Digital Archival Collections and the Charges for Accessing Them," *Jisc Content and Digitisation*, 2019 [A report on costs to libraries to buy and access digital resources in the UK] <<https://digitisation.jiscinvolve.org/wp/2019/07/01/commercial-digital-archival-collections-and-the-charges-for-accessing-them/>> [accessed 29 October 2021]

Google Books – About <https://books.google.com/intl/en/googlebooks/about/index.html>

Gregg, Stephen H. *Old Books and Digital Publishing: Eighteenth Century Collections Online* (Cambridge: Cambridge University Press, 2020): <https://www.cambridge.org/core/elements/old-books-and-digital-publishing-eighteenthcentury-collections-online/058DB12DE06A4C00770B46DCFAE1D25E>

Hammond, Paul, *Literature in the Digital Age* (Cambridge: Cambridge University Press, 2015)

Kizhner, Inna, Melissa Terras, Maxim Rumyantsev, Valentina Khokhlova, Elisaveta Demeshkova, Ivan Rudov, and others, "Digital Cultural Colonialism: Measuring Bias in Aggregated Digitized Content Held in Google Arts and Culture," *Digital Scholarship in the Humanities*, 36.3 (2021), 607-640 <<https://doi.org/10.1093/lc/fqaa055>>

Koh, Adeline, "Inspecting the Nineteenth-Century Literary Digital Archive: Omissions of Empire," *Journal of Victorian Culture*, 19.3 (2014), 385-395

Mak, Bonnie, "Archaeology of a Digitization," *Journal of the Association for Information Science and Technology*, 65.8 (2014), 1515-1526 <<https://doi.org/10.1002/asi.23061>> [on EEBO]

Sherratt, Tim, "Unremembering the Forgotten," in *Debates in the Digital Humanities*, 2019 <<https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/be608100-95b6-4e48-bfd5-a82a588da8f1#ch12>> [accessed 21 July 2020]

Somers, James, "Torching the Modern-Day Library of Alexandria," *The Atlantic*, April 20, 2017, https://www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/?utm_source=atltw

Thylstrup, Nanna Bonde, *The Politics of Mass Digitization* (Cambridge, MA: MIT Press, 2018) [on Google Books, Europeana, DPLA, Monoskop]

Wells, H. G., *World Brain* (New York: Doubleday, 1938)

<https://archive.org/details/worldbrain00wells/page/n5>

Wernimont, Jacqueline, and Julia Flanders, "Feminism in the Age of Digital Archives: The Women Writers Project," *Tulsa Studies in Women's Literature*, 29.2 (2010), 425-35

Date Created

01/30/2022