HISTORY 2029

Digital Methods and Primary Sources

Graduate Seminar
Spring 2023

Gabe Pizzorno
pizzorno@fas.harvard.edu
Office: Robinson Hall 210
Office Hours: Tuesdays, 10:00am-noon, and by appointment

Course Canvas site: https://canvas.harvard.edu/courses/117634
Time: Mondays, noon-2:45pm
Location: Robinson 107

Please note: this is a 3-hour seminar, although we won't necessarily use the full allotted time every week. In addition, we will occasionally meet in locations other than the assigned classroom, such as libraries and archives, which will be announced in class.

Overview

This seminar will teach participants how to use digital tools to design and implement data collection strategies that anticipate their future analytical needs and maximize the research potential of the sources being collected. The course will focus on the design and implementation of effective workflows for digitization and data mining of primary sources, data modelling, and advanced computational approaches to the management of large collections and the extraction of information from a range of visual and textual sources.

Description

The use of digital evidence, methods, and tools has had a huge impact on nearly every aspect of historical work including the ways in which we collect, organize, and prepare primary sources for analysis. This seminar will teach participants how to use digital tools to design and implement data collection strategies that anticipate their future analytical needs and maximize the research potential of the sources being collected.

Specifically, we will cover the process of conducting initial evaluations of collections, determining their suitability to address specific historical questions, and implementing effective workflows for their digitization or data mining that follow established standards and best practices. Participants will also learn how to design storage strategies that take into account the kind of data model that best suits their data and its intended usage as well as advanced approaches to the management of large collections and the extraction of pertinent information from different types of sources such as documents, correspondence, tables, photographs, cartography, and so on. We will also discuss tools and strategies for project management, quality control, and documentation, as well as common pitfalls in these areas.

The seminar is designed to provide participants with working knowledge of methods, techniques, and tools to enhance their research and will include significant hands-on work. It is crucial, however, to see past the instrumental aspects of tools and platforms and focus on how they can be adapted, in a theoretically informed manner, to the unique ways in which scholarly disciplines approach their subject. As such, the seminar will also focus heavily upon key aspects related to the production of historical knowledge, such as the nature and types of historical evidence and fundamental principles of source criticism, particularly as it pertains born digital and digitized primary sources.

Objectives

Participants in the seminar will:

- Learn to design and implement sound data collection strategies that carefully integrate digital methodologies into their research, anticipates analytical needs, minimize data reconfiguration, and maximize the flexibility of the sources.
- Become conversant across a spectrum of relevant digital techniques so that they can critically evaluate their potential and limitations as it pertains specific research applications, and integrate them into their work in a careful, theoretically informed way.
- Learn to communicate effectively with technical experts to facilitate collaboration and integration into interdisciplinary research frameworks

Requirements and Grading

45% Informed participation in all seminar meetings, as well as the preparation of occasional discussion outlines (based on a schedule to be determined in the first week of classes). The articles and book chapters assigned for each week will be posted online. For other materials, such as books from the library, we will work out a system for sharing them amongst ourselves as necessary.

20% Archive/collection assessment (ca. 1,000 words). An essay presenting a full evaluation of an archive or source collection along the lines discussed in class.

35% Final project (format TBD in consultation with the instructor). This would ideally include all tools (databases, scrapers, DAMs, etc) and workflows designed to process a specific source collection alongside detailed documentation that follows the criteria discussed in class.

In the case of both the collection assessment and the final project, participants are strongly encouraged to focus on materials related to their own work (e.g. dissertation, in-progress paper, etc).

Prerequisites

There are no formal requirements for the seminar although participants are expected to either possess a minimal degree of comfort with basic data-processing techniques or be willing to some additional work during the semester to acquire it. If you wish to gain some experience in this area ahead of the term, I would recommend the workshops that form the common core requirements of the Digital Scholarship Certificate offered by Harvard's Digital Scholarship Group, which are offered regularly throughout the year.

This seminar is not designed to provide a general introduction to digital scholarship or the use of digital methods for historical research. If you are looking for such a course, please consider taking HIST 1993: Introduction to Digital History, which is offered regularly.

Schedule

Week 1: Formation Processes of the Historical Record
Importance of understanding how archives and other primary source collections are assembled and the structural bias those processes introduce. Strategies for evaluating source collections, including their suitability to address specific historical questions. Relevant issues of nomenclature and classification, including source taxonomy and typology.

> Akin, Marjorie. 'Passionate Possession: The Formation of Private Collections'. In *Learning from Things: Method and Theory of Material Culture Studies*, edited by W. D. Kingery, 102–28. Washington, D.C.: Smithsonian Institution Press, 1996.
>
> Parezo, Nancy J. 'The Formation of Anthropological Archival Records'. In *Learning from Things: Method and Theory of Material Culture Studies*, edited by W. D. Kingery, 145–72. Washington, D.C.: Smithsonian Institution Press, 1996.
>
> Schnapp, Alain. 'The Antiquarian, the Collector, and the Cultural History of the Material World'. In *Cultural Histories of the Material World*, edited by Peter N. Miller, 144–50.

The Bard Graduate Center Cultural Histories of the Material World. Ann Arbor, MI: University of Michigan Press, 2013.

Smith, Jason Scott. 'The Strange History of the Decade: Modernity, Nostalgia, and the Perils of Periodization'. *Journal of Social History* 32, no. 2 (1998): 263–85.

Welch, Todd. '"Green" Archivism: The Archival Response to Environmental Research'. *The American Archivist* 62, no. 1 (1999): 74–94.

Week 2: Sources as Data and Data as Source

Definitions of data and metadata, data formats, types, and structures. Introduction to data modelling and relational logic. Potential and perils of data as a historical source. Practical aspects of storing historical data, spreadsheets, databases, and unstructured approaches.

Doulton, David, and Arno Kitts. 'The Storing and Processing of Historical Data'. In *History and Computing II*, edited by Peter Denley, Stefan Fogelvik, and Charles Harvey, 81–87. Manchester; New York: Manchester University Press, 1989.

Greenstein, Daniel. I. 'A Source-Oriented Approach to History and Computing: The Relational Database'. *Historical Social Research / Historische Sozialforschung* 14, no. 3 (51) (1989): 9–16.

Hartland, Philip, and Charles Harvey. 'Information Engineering and Historical Databases'. In *History and Computing II*, edited by Peter Denley, Stefan Fogelvik, and Charles Harvey, 44–62. Manchester; New York: Manchester University Press, 1989.

Owens, Trevor. 'Defining Data for Humanists: Text, Artifact, Information or Evidence?' *Journal of Digital Humanities* 1, no. 1 (2011).

Schöch, Christof. 'Big? Smart? Clean? Messy? Data in the Humanities'. *Journal of Digital Humanities* 2, no. 3 (2013).

Swafford, Joanna. 'Messy Data and Faulty Tools'. In *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein. Univ Of Minnesota Press, 2016.

Week 3: Analogue Sources

How to design efficient digitization workflows. Practical considerations such as access, discoverability, sequencing, space, and light. Discussion of specifics related to different source types as well as existing standards and best practices. Issues of scale and organization, digital assets managers.

Pizzorno, Gabriel H. and Gareth Darbyshire. 'Building Digital Gordion: Coping with the Past in the 21st Century'. *Expedition* 51, no. 2 (Summer 2009): 23–30.

Vajcner, Mark. 'The Importance of Context for Digitized Archival Collections'. *Journal of the Association for History and Computing*, April 2008.

Whitelaw, Mitchell. 'Generous Interfaces for Digital Cultural Collections'. *Digital Humanities Quarterly* 9, no. 1 (2015).

Week 4: Born-digital and Mixed Sources
Finding and using online data sources and local repositories. Scraping, APIs, tagging, and metadata generation and management.

Flanders, Julia, and Trevor Muñoz. 'An Introduction to Humanities Data Curation'. *DH Curation Guide* (blog), 28 March 2014.

Maristella Agosti and Nicola Orio. 'User Requirements for Effective Access to Digital Archives of Manuscripts'. *Journal of Multimedia* 7, no. 2 (2012): 217–22.

Week 5: Workshop I

Week 6: Project Design and Management
Sampling and collection strategies, management tools, quality control, and documentation.

Owens, Trevor. 'Please Write It Down: Design and Research in Digital Humanities'. *Journal of Digital Humanities* 1, no. 1 (2011).

Reed, Ashley. 'Managing an Established Digital Humanities Project: Principles and Practices from the Twentieth Year of the William Blake Archive'. *Digital Humanities Quarterly* 8, no. 1 (2014).

Theimer, Kate. *Web 2.0 Tools and Strategies for Archives and Local History Collections*. New York: Neal-Schuman Publishers, 2010.

Week 7: Image as Image
Working with photographs, negatives, slides, cartography, and other types of images. Strategies for managing and interrogating large collections. Advanced image analysis and data extraction methods.

Budzise-Weaver, Tina. 'Developing a Qualitative Coding Analysis of Visual Artwork for Humanities Research'. *Digital Humanities Quarterly* 10, no. 4 (2016).

Deal, Laura. ‘Visualizing Digital Collections’. *Technical Services Quarterly* 32, no. 1 (2 January 2015): 14–34.

Drucker, Johanna. ‘Image, Interpretation, and Interface’. In *Graphesis: Visual Forms of Knowledge Production*, 17–63. MetaLABprojects. Cambridge, Massachusetts: Harvard University Press, 2014.

Evens, Aden. ‘Web 2.0 and the Ontology of the Digital’. *Digital Humanities Quarterly* 6, no. 2 (2012).

Theibault, John. ‘Visualizations and Historical Arguments’. In *Writing History in the Digital Age*, edited by Jack Dougherty and Kristen Nawrotzki, 173–85. Digital Humanities. University of Michigan Press, 2013.

Week 8: Image as Text

How to work with images that contain text. Optical Character Recognition, transcription standards, text encoding.

Binder, Jeffrey M. ‘Alien Reading: Text Mining, Language Standardization, and the Humanities’. In *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein. Univ Of Minnesota Press, 2016.

Clement, Tanya. ‘The Ground Truth of DH Text Mining’. In *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein. Univ Of Minnesota Press, 2016.

Engel, Deena, and Marion Thain. ‘Textual Artifacts and Their Digital Representations: Teaching Graduate Students to Build Online Archives’. *Digital Humanities Quarterly* 9, no. 1 (2015).

Renear, Allen H. ‘Text Encoding’. In *Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, Hardcover. Blackwell Companions to Literature and Culture. Oxford: Blackwell Publishing Professional, 2004.

Week 9: Text as Data

Approaches to the cleaning and manipulation of large corpora of texts. Data extraction for textual sources. Special types of documents: correspondence, tabular and quantitative data. Discussion of close and distant reading.

Engel, Deena, and Marion Thain. ‘Textual Artifacts and Their Digital Representations: Teaching Graduate Students to Build Online Archives’. *Digital Humanities Quarterly* 9, no. 1 (2015).

Hancher, Michael. ‘Re: Search and Close Reading’. In *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein. Univ Of Minnesota Press, 2016.

Jockers, Matthew. ‘On Distant Reading and Macroanalysis’. Accessed 30 August 2017.

Kuhn, Jonas. ‘Computational Text Analysis within the Humanities: How to Combine Working Practices from the Contributing Fields?’ *Language Resources and Evaluation* 53, no. 4 (1 December 2019): 565–602.

Underwood, Ted. ‘Distant Reading and Recent Intellectual History’. In *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein. Univ Of Minnesota Press, 2016.

Week 10: Analytical Pathways

Introduction to a range of analytical methods that take advantage of different types of digital primary sources. Networks analysis and prosopography, Natural Language Processing, topic modelling and sentiment analysis, named entity extraction, computer vision approaches.

Clement, Tanya, David Tcheng, Loretta Auvil, Boris Capitanu, and Megan Monroe. ‘Sounding for Meaning: Using Theories of Knowledge Representation to Analyze Aural Patterns in Texts’. *Digital Humanities Quarterly* 7, no. 1 (2013).

Hedges, Mark. ‘Grid-Enabling Humanities Datasets’. *Digital Humanities Quarterly* 3, no. 4 (2009).

Hunter, Lynette. ‘Fact—Information—Data—Knowledge: Databases as a Way of Organizing Knowledge’. *Literary and Linguistic Computing* 5, no. 1 (1 January 1990): 49–57.

Flanders, Julia. ‘Data and Wisdom: Electronic Editing and the Quantification of Knowledge’. *Literary and Linguistic Computing* 24, no. 1 (1 April 2009): 53–62.

Sweetnam, Mark S., and Barbara A. Fennell. ‘Natural Language Processing and Early-Modern Dirty Data: Applying IBM Languageware to the 1641 Depositions’. *Literary and Linguistic Computing* 27, no. 1 (1 April 2012): 39–54.

Weingart, Scott B. ‘Demystifying Networks, Parts I & II’. *Journal of Digital Humanities* 1, no. 1 (2011).

Week 11: Workshop II

Week 12: Source Criticism in a Digital World
Discussion of aspects of source criticism that are specific to digital sources (e.g. algorithmic bias).

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. First Edition. New York, NY: StMartin's Press, 2018.

Hine, Christine. 'Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work'. *Social Studies of Science* 36, no. 2 (2006): 269–98.

Sculley, D., and Bradley M. Pasanek. 'Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities'. *Literary and Linguistic Computing* 23, no. 4 (1 December 2008): 409–24.

Swinger, Nathaniel, Maria De-Arteaga, I. V. Neil Thomas Heffernan, Mark D. M Leiserson, and A. Kalai. 'What Are the Biases in My Word Embedding?' *AIES*, 2019.

Week 13: Digital Sources and the Historian's Craft
Participants project presentations and general discussion.

Boonstra, Onno, Leen Breure, and Peter Doorn. 'Past, Present and Future of Historical Information Science'. *Historical Social Research / Historische Sozialforschung* 29, no. 2 (108) (2004): 4–132.

Galarza, Alex. 'A Call to Redefine Historical Scholarship in the Digital Turn'. *Journal of Digital Humanities* 1, no. 4 (2012).

Gibbs, Fred, and Trevor Owens. 'The Hermeneutics of Data and Historical Writing'. In *Writing History in the Digital Age*, edited by Jack Dougherty and Kristen Nawrotzki, 159–70. Digital Humanities. University of Michigan Press, 2013.

Tenen, Dennis. 'Blunt Instrumentalism: On Tools and Methods'. In *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein. Univ Of Minnesota Press, 2016.

Thaller, Manfred. 'The Need for a Theory of Historical Computing'. In *History and Computing II*, edited by Peter Denley, Stefan Fogelvik, and Charles Harvey, 1–11. Manchester; New York: Manchester University Press, 1989.

Thomas, William G. ‘Computing and the Historical Imagination’. In *Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, Hardcover. Blackwell Companions to Literature and Culture. Oxford: Blackwell Publishing Professional, 2004.

Wolff, Robert S. ‘The Historian’s Craft, Popular Memory, and Wikipedia’. In *Writing History in the Digital Age*, edited by Jack Dougherty and Kristen Nawrotzki, 64–74. Digital Humanities. University of Michigan Press, 2013.