Texts, Maps, and Networks: Readings and methods for digital history
Professor Benjamin Schmidt
benschmidt.org/dighist17
Email: b.schmidt@northeastern.edu
Office Hours: Wednesday, 1-3pm 413 Nightingale Hall

Overview

This course explores changing methods and practices of history in a digital age. It is designed to expose students to the wide variety of work being done computationally by historians and other humanists today, and equip them to be creative producers of new work and critical and constructive readers of existing projects. We'll look at how the historian's craft in three major units: the creation of digital sources, the algorithmic transformations that computers can enact on cultural materials like texts, and the new ecologies of publishing and scholarly communication made possible by new media.

In a new experiment, we're also going to try run a thread through the course that looks at how the general principles we've been using could be integrated with one of the newest areas of intersection between computers and the humanities; the digitization of historical photographs. We'll be particularly focused on thinking about what historical uses there may be for the millions of digital photographs collected by the Digital Public Library of America; insofar as is possible, we'll try to prototype and explore some of these possibilities.

Course Units

1. Digital Sources

What is a digital source? How do we create them? Looking at both "born-digital" sources and the processes of digitization, we'll explore the theory and practice of bringing artifacts from the analog world into the digital one, of interpreting secondary literature from fields that build heavily on data-as-evidence, and the obligations that the social construction of the digital archive creates on us as critical readers of digital primary sources.

2. Algorithmic transformations

Tools and methods developed over the last few decades can be of tremendous value for humanists. We'll spend a week learning the theory and basics of these fields, including network analysis, digital mapping, text analysis and categorization, and analysis of tabular data.

3. Digital Publishing

All publishing today is digital. Some digital projects express their final form on paper, some in printed 3D models, and some on phone and computer screens. This section will focus on the debates and possibilities surrounding telling historical story through these various digital media, with a particular focus on building web exhibitions for public history projects and online components for more traditional scholarship.

Note on scheduling

We'll be working together to make this course as useful to your professional development as possible. That means you should feel free to communicate any changes you'd like to see to me; it also means that the syllabus is subject to be updated, with due notice, at any time. The latest version of the syllabus will be the one on the

course website; in the event of any conflicts about what to read, first priority goes to any e-mails from myself; second to the latest version on the website; and only third priority to what's listed on the paper copy of the syllabus.

Course Goals

In three course stages, we're going to get our hands messy. To understand the possibilities and the limitations of digital work, there is no substitute for engaging with it directly. That said, you're not expected to come into this class a master programmer, and you shouldn't expect to leave it as one. This is an introductory course to a large and growing field.

Rather than try to completely master any single technique (which other courses at Northeastern will help you do), we'll be trying to get a broad introduction to a variety of tools let us assess and create some of the work already out there. The worksets are designed to get you started in a number of fields so that you know where to start when the time comes to make a map for a dissertation chapter, to search a large collection of texts for a particular construction, or to make a network diagram for a conference talk.

In this course, you will:

1. Describe the sort of work being done under the banner of digital history today, and be able to participate in some of the debates in the field.

2. Know the practical and social mechanics of digitizing cultural artifacts.

3. Cultivate a base of experience that will help you to engage in digital research projects for your own work with texts, geographical systems, or networks. This experience will hopefully be both knowledge–how do you make a publishable digital map?–and practical wisdom–how do you select the right software platform or tool for a research project?

4. Build and curate online displays that find new modalities for sharing historical artifacts and knowledge.

5. Create and/or further develop a professional online identity.

Requirements

Readings and attendance

You know the basics: you should complete the required readings and attend class prepared with questions and criticisms. This is a large class for a seminar; it is more likely than usual that you will need to be prepared before class to make useful interventions in it.

Computers

Most of the assignments for this course will need to be done on a computer; a number of the worksets will ask you to install software for the lab section at the end of class. If your computer is nonexistent, very old, or is in any other way unable to fulfill some particular assignment, you may need to complete certain assignments in a computer lab and arrange to share with a colleague during class. I have an old laptop and some desktop site I may also be able to make available. Everyone, including me, will encounter some technological problems in the class; the important thing is that you take them on in advance.

Blogging

You are expected to contribute almost every week to a blog that will syndicate to the rest of the course. Over the course of the semester, you should have at least 12 posts which cumulatively reach about 4,000-5,000 words. These are not expected to be completely polished pieces of writing, but they should show your engagement with the texts and your peers, and create a ground you can build on in later work and discussion. The word limit is not particularly high; try to stick to the genre, avoid all throat clearing and generalizations; either get right to the point, or keep your digressions interesting. Write for the audience of the class and anyone like you who might stumble across the pages.

You should read your peers' blog posts for at least the weeks that you post.

Posts should available by 1pm the day of class.

Topics would typically include things like:

1. Reactions to the reading: questions you want your peers to answer, things you don't understand, or angry denunciations of what you think the writers got wrong.
2. Reflections on connections between the readings and issues you've encountered in digital history in other courses or online.
3. Responses to questions posed by your colleagues.

I also **highly** encourage you to the think about the image analysis overlaps as part of your posting regimine. Over the course of the semester I'll periodically ask for some short reflections on various questions

Required posts

Most of your posts will be on topics of your choosing. Others, though, may be required as listed in the syllabus.

Blog Privacy

You may have good reasons not to want your name associated with your blog posts or Internet presence–if so, we can make your posts private or (preferably) pseudonymous. But remember, conversely, that building up a strong professional online presence can be enormously beneficial. Student blog posts from this class have, in the past, ended up on the syllabi for graduate courses at other universities. There is much to gain, as more senior graduate students will tell you, from developing a public professional identity beginning early.

Worksets

Each week we will begin a *practicum*, learning to use some set of tools. At the end of class you'll receive a short list of tasks to accomplish. By the end of the course, you will know how to make a map, mine a text, create a network diagram, and set up an online exhibition.

Completing each workset is required, but the individual tasks will be handled on a pass/fail basis. If you do every workset, you get full credit for this portion of the course. I encourage you to talk to other students while completing the worksets (much frustration can be avoided not doing it alone), but unless otherwise indicated you must do the tasks yourself, even if someone else tells you how to do it.

You should **give evidence of having completed the assignment** to me before the start of the next class. This may be screenshots, some brief textual reflections, or a description of a website you visited. Use common sense here. Late or incomplete worksets can receive partial credit.

If you want to go above and beyond the basic assignment–adding a colored layer to the maps we build, say, you can post to the blog. Certain worksets will tell you to create a blog post–in those cases, you'll get credit for completing practicum as well as towards your blog post quota.

Final Projects

As we take on classroom exercises in the second third of the course, you should think about which one(s) you want to expand into a longer form. We'll also try to get a trip to the archives in so that you can practice some digital curation on your own.

Projects may be collaborative. In general, public history students should be predisposed towards a collaborative project, and world history or comparable students towards an individual one. Collaborative projects should have a sensible division of labor, and include individual statements of the work done.

You will submit a proposal for your project by in early November: projects are due a week before the semester ends.

Grading

1. Class participation/attendance (and blog comments): 25%
2. Blogging assignments: 20%
3. Practicums: 25%
4. Projects: 30%

Required Texts

Most of the texts for this class are available online; digital humanists are good that way. Some are also available for physical purchase.

- Tufte *Envisioning Information*.

Schedule

Unit 1: Digital sources

In the first unit, we'll be exploring how digitization changes the sort of sources–primary and secondary–that historians work with. What are the biases and inherent assumptions in digital scholarship? What gets digitized, and what doesn't? What sort of answers are computational works of scholarship bringing to historical practice?

Week 1 (Sep 11): Introductions

- Cohen et al. "Interchange."

Practicum: Building a website and blog for this course.

Week 2 (Sep 18): Digitization as a social condition

- Benjamin "The Work of Art in the Age of Its Mechanical Reproducibility (Third Version, 1939)."
- McLuhan *Understanding Media*., Introduction and chapters 1-2.

Digital Humanities as a social obligation

- Liu "The Meaning of the Digital Humanities."
- Cecire "When Digital Humanities Was in Vogue."

Practicum: What's been digitized? And Manipulating Images.

Week 3 (Sep 25): Digitization as a practical endeavor

- Putnam "The Transnational and the Text-Searchable."
- On Google Books

    - Torching the Modern-Day Library of Alexandria, James Somers, *The Atlantic*, 2017-04-20.
    - What Happened to Google's Efforts to scan Millions of University Library Books? Jennifer Howard, Edsurge

- Bill Turkel's blog posts on digitizing text. Just browse these so you get an idea what's involved.
- **Start with a backup and versioning strategy**.
- **Make everything digital**.
- **Doing OCR Using Command Line Tools in Linux**.
- **Working with PDFs Using Command Line Tools in Linux**.
- Melissa Terras, "Digitization's Most wanted": http://melissaterras.blogspot.com/2014/05/digitisations-most-wanted.html
- Trevor Owens, 4-part series on crowdsourcing

Practicum: Digitizing a a historical object.

Week 4 (Oct 2): Data as a source – the slavery debates

- Gibbs and Owens "Hermeneutics of Data and Historical Writing."
- Haskell, Review of *Time on the Cross*
- Ruggles "The Transformation of American Family Structure."
- Hilt "Economic History, Historical Analysis, and the "New History of Capitalism"."
- Gitelman *"Raw Data" Is an Oxymoron*., Introduction

Practicum: Tapping into the world of social science historical data from IPUMS, OECD, ICPSR, etc. Practicum w/ blog post: Crowdsourcing participation.

(No class October 9: Columbus day)

Unit 2: Historical Computing

For years, much of what's now called the"digital humanities" was called, instead, "humanities computing." The term tended to denote a more circumscribed set of practices than all the digital publishing, public history, and new media studies that are now part of digital humanities; it was, specifically, about the the possibility of digital techniques to transform the ways we do research. This unit aims to get your hands dirty with some of the research techniques you might be able to use taking full advantage of your computation.

Week 5 (October 16): Texts

- Stephen Ramsay, "The Hermeneutics of Screwing Around"
- Witmore "Text."
- Funk and Mullen "The Spine of American Law."
- Goldstone and Underwood "The Quiet Transformations of Literary Studies."
- Tressie McMillan Cottom, "More Scale, More Questions: Observations from Sociology"

Look at two websites, and skim the associated papers if you like. * books.google.com/ngrams (Michel et al. "Quantitative Analysis of Culture Using Millions of Digitized Books.") * https://nlp.stanford.edu/projects/histwords/ William Hamilton, Jure Leskovec, and Dan Jurafsky. ACL 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.

Practicum:

- Voyant Tools (no installation needed, online)
- Install R-Studio on your computer.

Week 6 (October 23): Images

Some image-heavy DH projects:

- Kate Bagnall and Tim Sherrat,"Invisible Australians: Living under the White Australia Policy"
- Laura Wexler, Lauren Tilton, Taylor Arnold, et al.: Photogrammar: http://photogrammar.yale.edu/
- http://charliebyers.org/AFID/index.php?matchImage=2484&sourcesPage=1&facesPage=1&linkAct=matchImage#

A few articles about things that people are doing with images analysis nowadays (list to be extended):

- https://www.economist.com/news/science-and-technology/21728614-machines-read-faces-are-coming-advances-ai-are-used-spot-signs
- https://www.nytimes.com/2017/08/14/arts/design/google-how-ai-creates-new-music-and-new-artists-project-magenta.html
- Shiry Ginosar, Kate Rakelly, et al., A Century of Portraits: A Visual Historical Record of American High School Yearbooks. (2015)
- http://www.npr.org/2016/10/23/499042369/police-facial-recognition-databases-log-about-half-of-americans
- Either this dense paper

(For background, you could read this generally useful, slightly hyperventilating introduction to neural networks from the NY Times)

Week 7 (October 30): Maps

Guest: Cameron Blevins?

- Richard White, What is Spatial History
- Knowles "Placing History.", Chapter 1 and Dust Bowl chapter.
- Blevins "Space, Nation, and the Triumph of Region."
- Browse through the Orbis Project, Stanford.

Practical:

- Install QGIS from here (this can be complicated, particularly on OS X: leave an hour, at least).
- Review Map Projections at jasondavies.com.

Week 8 (November 6): Networks

Image lab stock-taking. Maybe more image readings? We'll see.

- Weingart "Demystifying Networks, Parts I & II."
- Winterer "Where Is America in the Republic of Letters."
- Shin-Kap Han,"The Other Ride of Paul Revere"

Software: Install *Gephi*.

Unit 3: Creating Digital Scholarship.

The sort of work historians create and share matters as much as the sort of work they do.

You could use the techniques from unit 2 and produce a wholly conventional work of scholarship; and you could create a groundbreaking multimedia installation without using any algorithms or even programming. This unit focuses on the opportunities for scholarly communication afforded by the web and other digital media.

Week 9 (November 13): Digital Collections and exhibitions.

- Wyman et al. "Digital Storytelling in Museums."
- Sharon Leon: Layers and Links: Writing Public History in a Digital Environment

We'll be doing presentations on born-digital exhibitions.

There are a wide variety of professional digital collections and exhibitions. Rather than have each of you explore all of them, find one and explore it at length, bringing several URLs to class to discuss as successes or failures of design, narration, and public engagement.

Some possible online archives/exhibits to present on will be listed in the "Handouts" section of the website.

Practicum

- Workset: Building with Omeka.

Week 10 (November 20): Visualizing Data

Project proposals due Friday 11/11

- Theibault "Visualizations and Historical Arguments."
- Tufte *Envisioning Information*.
- Drucker "Humanities Approaches to Graphical Display."
- Klein "The Image of Absence."
- Catherine D'Ignazio and Lauren F. Klein, "Feminist Data Visualization" (2016)

Practicum: visualizing in R: the "grammar of graphics"

Week 11 (November 27th): Stories in New Media

- "Snow Fall."
- MacAskill and MacAskill "NSA Files Decoded."
- Something TBD, hopefully something released this fall; suggestions welcome.

Practicum: Workshop on Public History project

Week 12 (December 4): Publishing and sharing research

- Ayers "The Valley of the Shadow."
- Thomas and Ayers "The Differences Slavery Made."
- Glanz "Writing a Digital History Journal Article from Scratch."
- Dan Cohen, "The Social Contract of Scholarly Publishing," Gold *Debates in the Digital Humanities*.

Practicum: none

Acknowledgements

This syllabus draws on elements of classes taught by Sharon Leon, Ryan Cordell, Dan Cohen, Adam Crymble, Amanda French, Alan Liu, and many others.

Andrew Goldstone designed the pdf template used for the syllabus. The website uses Yihui Xie's rmarkdown code and styles.

Full Citations

Ayers, Edward. "The Valley of the Shadow: Two Communities in the American Civil War," 1995. http://valley.lib.virginia.edu/.

Benjamin, Walter. "The Work of Art in the Age of Its Mechanical Reproducibility (Third Version, 1939)." In *Selected Writings*, edited by Howard Eiland and Michael W Jennings. Vol. 4. Cambridge, Mass.: Belknap Press

of Harvard Univ. Press, 2002.

Blevins, Cameron. "Space, Nation, and the Triumph of Region: A View of the World from Houston." *Journal of American History* 101, no. 1 (May 1, 2014): 122–147. doi:10.1093/jahist/jau184.

Cecire, Natalia. "When Digital Humanities Was in Vogue." *Journal of Digital Humanities* 1, no. 1 (February 9, 2012). http://journalofdigitalhumanities.org/1-1/when-digital-humanities-was-in-vogue-by-natalia-cec

Cohen, Daniel J., Michael Frisch, Patrick Gallagher, Steven Mintz, Kirsten Sword, Amy Murrell Taylor, William G. Thomas, and William J. Turkel. "Interchange: The Promise of Digital History." *The Journal of American History* 95, no. 2 (August 1, 2008): 452–491. doi:10.2307/25095630.

Drucker, Johanna. "Humanities Approaches to Graphical Display" 5, no. 1 (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Funk, Kellen, and Lincoln Mullen. "The Spine of American Law: Digital Text Analysis and U.S. Legal Practice." *American Historical Review* 123, no. 1 (2018).

Gibbs, Frederick W., and Trevor Owens. "Hermeneutics of Data and Historical Writing." In *Writing History in the Digital Age*, 2011. http://writinghistory.trincoll.edu/data/hermeneutics-of-data-and-historical-writing

Gitelman, Lisa. *"Raw Data" Is an Oxymoron*. Infrastructures Series. Cambridge, Massachusetts: The MIT Press, 2013.

Glanz, James. "Writing a Digital History Journal Article from Scratch: An Account'." *Digital History Project* (November 2007). http://digitalhistory.unl.edu/essays/thomasessay.php.

Gold, Matthew K. *Debates in the Digital Humanities*. U of Minnesota Press, 2012.

Goldstone, Andrew, and Ted Underwood. "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us." *New Literary History* 45, no. 3 (2014): 359–384. doi:10.1353/nlh.2014.0025.

Hilt, Eric. "Economic History, Historical Analysis, and the "New History of Capitalism'." *The Journal of Economic History* 77, no. 2 (May 2017): 511–536. doi:10.1017/S002205071700016X.

Klein, Lauren F. "The Image of Absence: Archival Silence, Data Visualization, and James Hemings." *American Literature* 85, no. 4 (November 1, 2013): 661–688. doi:10.1215/00029831-2367310.

Knowles, Anne Kelly. "Placing History." ESRI Press, 2008.

Liu, Alan. "The Meaning of the Digital Humanities." *PMLA* 128, no. 2 (February 2013): 409–423. doi:10.1632/pmla.2013.128.2.409.

MacAskill, Ewen, and Ewen MacAskill. "NSA Files Decoded: Edward Snowden's Surveillance Revelations Explained. The Guardian," October 1, 2013. http://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded.

McLuhan, Marshall. *Understanding Media: The Extensions of Man*. McGraw-Hill, 1964.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Joseph P Pickett, Dale Hoiberg, et al. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science (New York, N.Y.)* 331, no. 6014 (14 2011): 176–182. doi:10.1126/science.1199644.

Putnam, Lara. "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast."

The American Historical Review 121, no. 2 (March 1, 2016): 377–402. doi:10.1093/ahr/121.2.377.

Ruggles, Steven. "The Transformation of American Family Structure." *The American Historical Review* 99, no. 1 (January 1994): 103. doi:10.2307/2166164.

"Snow Fall: The Avalanche at Tunnel Creek," 2012. http://www.nytimes.com/projects/2012/snow-fall/.

Theibault, John. "Visualizations and Historical Arguments." In *Writing History in the Digital Age*, by Jack Dougherty, 2012. http://writinghistory.trincoll.edu/evidence/theibault-2012-spring/.

Thomas, William, and Edward Ayers. "The Differences Slavery Made: A Close Analysis of Two American Communities." *The American Historical Review* 108, no. 5 (November 2003): 1299–1309. http://www2.vcdh.virginia.edu/AHR/.

Tufte, Edward R. *Envisioning Information*. Cheshire, Conn. (P.O. Box 430, Cheshire 06410): Graphics Press, 1990.

Weingart, Scott B. "Demystifying Networks, Parts I & II." *Journal of Digital Humanities* 1, no. 1 (February 15, 2012). http://journalofdigitalhumanities.org/1-1/demystifying-networks-by-scott-weingart/.

Winterer, Caroline. "Where Is America in the Republic of Letters." *Modern Intellectual History* 9, no. 03 (2012): 597–623. doi:10.1017/S1479244312000212.

Witmore, Michael. "Text: A Massively Addressable Object," November 31, 2010. http://winedarksea.org/?p=926.

Wyman, Bruce, Scott Smith, Daniel Meyers, and Michael Godfrey. "Digital Storytelling in Museums: Observations and Best Practices." *Curator: The Museum Journal* 54, no. 4 (September 1, 2011): 461–468. doi:10.1111/j.2151-6952.2011.00110.x.