Intro to Digital Research, LIB 201

Table of contents

1 Welcome to LIB 201: Introduction to Digital Studies

1.1 Meeting Time & Place

Monday, 6:00 - 8:45 Fant Library Seminar B (unless otherwise instructed)

1.2 Instructor information

Name: Hillary A. H. Richardson
Office Hours: by appointment at https://libreserves.muw.edu/appointments/hillary email: hhrichardson@muw.edu

1.3 Course Description

For this class, we will be using the corpus of letters in the Smith Papers to interrogate various conversations about race, gender, culture, economics, politics, and interpersonal relationships in rural, 20th century Mississippi. Using various digital humanities methods and tools, we will research the impact of these topics in the Smith Papers in a broader context, and we will display our findings in an online space that is visible to the public. In addition to sharing the findings of our research, we will make our research process visible, sharing raw data files, documenting the steps of our process, and detailing our collaborations.

1.4 Objectives and Goals

Students will create

- Define, digitize, assemble, and create metadata for a circumscribed dataset
- Design several small research projects that experiment with various digital tools and methods, creating a strategy for the project and documenting progress toward defined goals

- Demystify underpinnings of digital research projects, such as algorithms, formulas, and metadata
- Become familiar with various ethical, sociocultural, and technological issues associated with digital research and publishing.
- Synthesize the process of a digital project into a reflective paper that pinpoints the goals of the project, the challenges faced, and strategies used to complete those goals.

1.5 Texts

Links and PDFs for selected readings will be posted to the schedule prior to each meeting. There are no required textbooks or purchases for this class.

1.6 Grading

Assignments	% of Grade
Weekly Writings, check-ins, and discussion posts	30%
Digital Prepwork (metadata, transcriptions)	30%
Mini-projects	30%
Portfolio/Reflection Essay	10%

2 Schedule

2.1 Week 1 - Introduction to the course and digital scholarship

2.1.1 Reading

- Explore the course roadmap.
- Watch the Introduction to Digital Studies recorded lecture

    - Slides for this lecture are available here.

- Explore example digital scholarship projects:

    - Letters of Maynooth - 1916
    - The Digital Welty Lab
    - Goin' North: Stories of the Great Migration to Philadelphia
    - Colored Conventions Project
    - Salem Witch Trials Documentary Archive and Transcription Project -Lincoln Logarithms: Finding Meaning in Sermons

2.1.2 Deliverables

- Answer the Roadmap Questionnaire
- Complete the "Evaluating Digital Scholarship" check-in on Canvas

2.2 Module 1: Metadata and Project Management

2.2.1 Readings

- Drucker, Johanna. "Ontologies and Metadata Standards," *Introduction to Digital Humanities* 2014, pp 24-27. (Available for LIB 201 students in Canvas).
- Schöch, Christof, 2013. "Big? Smart? Clean? Messy? Data in the Humanities," *The Dragonfly's Gaze* (revised and published in *Journal of Digital Humanities* 2(3).) https://dragonfly.hypotheses.org/443

- Hanna, Laurel, 2021. "Pitts Staff Works to Counteract Racism in Library Catalog." *Candler News and Events*, Emory, Candler School of Theology, https://candler.emory.edu/news/releases/2021/01/pitts-staff-works-to-counteract-racism-in-library-catalog.html
- Ketchley, Sarah. "Planning & Managing Digital Projects," *DH in Practice Coursebook*. pp 46-55.(Available for LIB 201 students in Canvas).

2.2.2 Recommended further reading (optional)

- Erin Baucom, 2018. "An exploration into archival descriptions of LGBTQ Materials. *The American Archivist*, 81(1): 65-83. doi: https://doi.org/10.17723/0360-9081-81.1.65

2.2.3 Deliverables

- Complete Metadata module and the check-ins in each section

2.3 Module 2: Transcribing

2.3.1 Readings

- Woodford, Chris. 2021. "Optical Character Recognition," *Explain that Stuff*, https://www.explainthatstuff.com/how-ocr-works.html.
- "About the Smith Papers Collection." This document gives context for the letters and describes the main contributors. Read this before you begin transcribing.
- Transcribing from Start to Finish
- Transkribus Transcribing Conventions, 2021. Read Co-op, https://readcoop.eu/transkribus/howto/transkribus-transcription-conventions/.
- Smithsonian, "General Instructions for Transcription and Review," https://transcription.si.edu/instructions
- YouTube video of Hillary introducing Transkribus: https://youtu.be/szGPn0K_WWE
- YouTube video of Hillary transcribing a letter from beginning to end: https://youtu.be/-cDD9P0rnLw
- Interview with Bridget and Steve Pieschel about the contents of the Smith Papers letters: https://youtu.be/zPynkw9V-Q8

2.3.2 Installations

- Transkribus: https://readcoop.eu/transkribus/?sc=Transkribus (Installation Instructions here. *Be sure you have Java 8 installed before installing Transkribus.* This is in the instructions, but not in the first step. If you run into an issue, email your instructor!)

2.3.3 Deliverables:

- Research Discussion 1 - Research Topics in the Smith Papers
- Research Discussion 2 - Research topics, continued
- Data Prep Work Assignment:
    - Metadata
    - Transcriptions
    - Tags

2.4 Module 3: Network Analysis

- What is a network analysis, and how does it help me answer questions about a group of people?
- Structuring Data
- Spreadsheets!

2.4.1 Readings

- Rawson and Muñoz, "Against Cleaning.", 2016. *Curating Menus*
- Scott Weingart, 2011. "Demystifying Networks," *Journal of Digital Humanities* 1(1).

- **Optional readings** for more help with NA:
    - Thomas Padilla and Brandon Locke. "Introduction to Network Analysis."
    - Miriam Posner, "Network Analysis."
    - Scott Weingart, 2015, "Networks Demystified 9: Bimodal networks."
    - Palladio Tutorial and FAQs.
    - Miriam Posner, "Creating a Network Graph with Gephi."

2.4.2 Installations

No installations necessary for this module! We will use Google Sheets and Palladio

2.4.3 Deliverables

- Cleaned data tags
- Network Analysis Visualization
- 1-2 page Write-up
- Extra credit: "Fun with Spreadsheets"

2.5 Module 4: Maps and Timelines

- What does a map or a timeline help me answer?
- What kind of data does a map or a timeline use?
- Collaborating on a class project

2.5.1 Readings

- White, Phil. (2014). "Story map blog" https://blogs.library.duke.edu/data/2014/10/28/story-maps/
- Manuel Gimond, 2021, "Introduction to GIS," *Intro to GIS and Spatial Analysis.* https://mgimond.github.io/Spatial/introGIS.html
- **Explore** mapping and timeline projects:

  - "Torn Apart/Separados," http://xpmethod.columbia.edu/torn-apart/volume/2/index
  - "Native Land" https://native-land.ca/
  - Maps of the Starkville Civil Rights Project: https://starkvillecivilrights.msstate.edu/wordpress/the-place/
  - "How Wine Colonized the World," TimelineJS by VinePair, Inc.

2.5.2 Further Reading (optional)

- Lisa Charlotte Rost, 2021, "Which color scale to use when visualizing data" *Datawrapper*, https://blog.datawrapper.de/which-color-scale-to-use-in-data-vis/
- J. B. Harley, 1989, "Deconstructing the Map," *Cartographica*, 26(2), http://hdl.handle.net/2027/spo.47615
- Shannon Mattern, 2015, "Gaps in the Map: Why we're mapping everything, and why not everything can, or should, be mapped." https://wordsinspace.net/2015/09/18/gaps-in-the-map-why-were-mapping-everything-and-why-not-everything-can-or-should-be-mapped/
- Olivia Ildefonso, 2021, "Top Mapping Mistakes," CUNY Academic Commons, https://digitalfellows.commons.gc.cuny.edu/2021/05/12/top-mapping-mistakes/

2.5.3 Installations

- Create a free, public account with ArcGIS. Directions and links for doing so are here.
- Spreadsheet template for Timeline JS (by Knightlab). Step by step instructions are here, and a video introduction is here.

2.5.4 Deliverables

- Research Discussion 3 - Sharing Timeline Sources
- Collaborative Notes
- Map OR Timeline mini-project

2.6 Module 5: Text Analysis

- What are distant and close reading?
- How do different distant reading methods help me make sense of what I'm reading?

2.6.1 Readings

- Caulfield, Jack. "A quick guide to textual analysis," *Scribbr.com*, 2020.
- Sculley and Pasanek, "Meaning and mining: the impact of implicit assumptions in data mining for the humanities" *Literary and Linguistic Computing, 23*(4), 2008. *This link will take you to MUW Library databases. Sign in with your 950# if you are off campus!*

2.6.2 Explore

- Corpus Thomisticum
- The Willa Cather Archive
- Powerhouse Annotated: A digital edition
- Topic Modeling Martha Ballard's Diary

2.6.3 Installations

There are no installations for this module. We will use tools that are available through MUW Library and openly online:

Required Tools - Gale Digital Scholar Lab (No download required)

Optional tools - Lexos - web-based text-cleaning app - Voyant - web-based text analysis app - jsLDA: In-browser Topic Modeling
- dataBASIC - various web-based text analysis tools for different purposes

2.6.4 Deliverables

- Content Set Brainstorm
- Rough Draft of text analysis mini-project
- Text analysis mini-project

2.7 Module 6: Digital Exhibits

2.7.1 Readings

- Brannock, Jennifer. "Creating an Exhibit in Special Collections and Using it to Promote Collections and Educate Users," *Mississippi Libraries 73*(2), 2009. pp. 32-34.
- Ovadia, Steven. (2014). "Markdown for Librarians and Academics," *Behavioral and Social Sciences Librarian 33*, pp. 120-124.
- "Markdown Syntax" - A cheatsheet for markdown syntax

2.7.2 Installations

- Text Editor software (comes installed on most computers). Suggestions for text editors:
  - Visual Studio Code is the Microsoft Office equivalent of Notepad or TextEdit. VS Code is more user friendly, and works within your browser (by typing the period . in a github markdown page). For further direction see VS Code's Installation instructions for MAC or PC users.
  - PC users: Notepad comes already installed
  - Mac users: TextEdit comes already installed
  - If you are using an already-installed text editor, preview your markdown using something like https://dillinger.io/ or https://markdownlivepreview.com/

2.7.3 Deliverables

- Draft outline of Digital Exhibit
- Collaborative Notes Document
- Final narrative portfolio

2.8 Final Exam

In lieu of a final exam, you will turn in a Reflection Essay, due at the time of the university-scheduled exam time.

3 Assignments

Digital Prep Assignment - Due - Submit

Network Analysis Project - Due

Map/Timeline Project - Due

Text Analysis Project - Due

Narrative Project - Due

Reflection Assigment - Due

4 Digital Prep Rubric

4.1 Overview

Your work with individually assigned letters from the Smith Papers Collection will be the basis for this assigment. You will submit this work across 3 different submissions: - Metadata (worth 50 points) - Transcriptions (worth 100 points) - Tags (worth 50 points)

4.1.1 Metadata

From our Google Sheet, copy and paste the full metadata for the letters you were transcribing into a blank spreadsheet, and submit the .xlsx file to Canvas. Metadata will be checked for fullness (all the fields are filled in) and consistency (they are all formatted according to the guidelines).

Metadata that receives full credit will: - complete all metadata fields, following Smith Paper Guidelines formatting consistently - use Library of Congress subject headings that are relevant and appropriate - Consider and note harmful content in the letters within appropriate fields

4.1.2 Transcriptions

Export and upload the .txt files of your assigned transcriptions. The file names should follow our guidelines (mcj-dp018….txt), and the transcriptions should be complete and thorough, representing the letters as faithfully as possible (see Transcribing lesson 1!

Transcriptions that receive full credit will: - Maintain structural integrity of the original document through Transkribus' layout analysis tool - Transcribe the letters fully and as closely to the original letter as possible, keeping original spelling, markings (where appropriate), punctuation, etc.

4.1.3 Tags

Upload the metadata tags (e.g. person, place, organization, etc.) you created for each letter you transcribed. The exported tags will be in an .xls or .xslx format. They should have multiple tabs per sheet for each tag named (e.g. person, place, organization, etc. Each tag category has a separate tab), and there should be an "overview" tab for all tags.

Tags that receive full credit will: - tag all words as explained in our Tag Definitions list.

5 Spatial/Temporal Data Assignment Rubric

5.1 Overview

For this module, Maps and Timelines are ways for you to engage with the historical and contextual research of the collection. In groups, you will use either a map or timeline to show the progression of events or ideas, either over time or space, and to make an argument about these occurrences. For this assignment, you'll identify a topic within the letters of noted historical significance, research the broader impact of this topic, and put the conversations from the letters within this context.

Remember: **Your project is more than a list of things!** It's a visual argument about how/why they occurred.

5.2 Due

Date TBA

5.3 Rubric

Successful maps/timelines will be evaluated on the following criteria: Research, and the resulting timeline *or* map project.

5.3.1 Research:

- Defines the research question in both the headline of the timeline and the written accompaniment that goes beyond "a series of events" to explore the change over time/space
- Cites Smith Papers at least 5 times, using relevant examples and excerpts
- Cites at least 5 secondary sources, using relevant contextual information to ground research question

5.3.2 Timeline/Map project:

- Contains various media (e.g. images, sound, video, etc.) in a majority of entries
- Credits media (both within the Smith Papers Collection and that which is found online) fully
- Contextualizes events with secondary source information within the description text of appropriate entries.
    - Timeline only:
        * Uses at least 2 "groups" to put events into categories
    - Map only:
        * Uses additional shapefile data to enrich the individually created entries

6 Network Analysis Rubric

6.1 Overview

The Network Analysis project is a way for you to explore the relationships in the letters - who wrote to whom, how often, and to what degree. The idea is not necessarily to submit a clean, fully scrubbed network, but one that highlights a few relationships and the significance that may bring to the network as a whole.

6.2 Due

Date TBA

6.3 Submit

Follow Network Analysis lessons 3 - 4 for a detailed step-by-step for how to model your data for a network analysis. Submit a document with an image file (or several image files) of your network analysis in Palladio, and include an explanation (about 1-2 pages) of what's going on in the graph, which should give:

- a brief introduction to the collection of letters, or who/what the network represents, and
- a description what the nodes and edges represent within the greater network of the collection of letters.
- any insights you can provide on relationships or nodes to highlight (i.e. notes on who might have the highest degree centrality and why, or which node might serve as a bridge, etc.). If you need to zoom in on parts of the image to do this, include a zoomed-in screenshot of part of the network so that your explanation has a visual aid.

6.4 Rubric

Network Analyses that get full credit will be evaluated based on the explanation, the dataset, and the visualization itself.

6.4.1 The explanation:

- Adequately explains how the dataset was modeled and the methods of cleaning data, and links to the original data set
- Makes inferences about the patterns and anomalies in the visualization (e.g. who were the influential players, what the outlying subgroups mean, etc.)

6.4.2 The dataset:

- is as free from duplicates as possible (keeping "Against Cleaning" in mind!)
- Avoids copy and paste errors. Work done together maintains the "tidiness" of the dataset.

6.4.3 The Visualization:

- Contains an images of the network that are clearly visible and demonstrate the analysis from the paragraph
- Uses edges and node sizing to show impact of different people

7 Text Analysis Rubric

7.1 Overview

The Text Analysis assignment is a way for you to engage with the content (or, the text!) of the collection. The point of this assignment is not to *prove* anything through text analysis tools, but to ask questions and demonstrate the process you undertook to address those questions using different tools. Think of your text anlysis assignment as a scientific experiment that uses words as its test subject!

7.2 Due

7.3 Rubric

Text Analyses that get full credit will be evaluated based on 4 sections of the assignment: the research question, the methods, the results, and further research.

7.3.1 The Research Question:

- Clearly defines the purpose of the text analysis in the form of a research question
- Cites at least 2 secondary sources to establish a contextual foundation for the question
- Cites at least 3 primary sources, using them as evidence to reinforce claims made as a result of the text analysis

7.3.2 The Methods:

- Clearly define the parameters for inclusion and exclusion of the content set
- Explain the choice of 2 text analysis methods with sound justifications
- Provide detailed cleaning methods for the content set

7.4 The Results:

- Demonstrate the experiment's results with accessibly visual images from text analysis tools
- Provide insight and interpretation into the results of each analysis tool

7.5 Further Research:

- Clearly defines the limits of the study and how they affected the current results
- Outlines a concrete path for a future study as a result of the current study's limitations

8 Narrative Portfolio Rubric

8.1 Overview

You will create a digital exhibit that puts the research we've collected and the artifacts we've created in class into one project that tells a story about the Smith Papers Collection. You are synthesizing the items in the collection with what we've learned so far, and creating a digital story about it.

8.2 Due:

Date TBA

8.3 Rubric

Successful narrative projects will be evaluated on the following criteria: - Shows improvements on mini-projects from original submission - Uses narrative elements (text, links, and images) to weave together the mini-projects - Uses secondary sources to strengthen the validity of the argument within the narrative and mini-projects - Shares evidence of the digital scholarship (e.g. stop words lists, raw/clean data, explanation of process, etc.) as often and as necessary as possible - Provides ample text and explanation for all required sections - Uses markdown (and/or HTML if applicable) language to create a structured and dynamic digital exhibit

9 Final Reflection Rubric

9.1 Assignment objectives:

Throughout this class, we have focused on how to use different kinds of digital research to interact with archival documents. We have been balancing the line between automating processes with computers, and using computers to do processes manually. Think about how these processes you've employed throughout the semester have done this – transcribing texts with an algorithm AND manually, reading letters individually and having a computer read them all, hypothesizing an output of a text analysis mechanism, sorting and classifying data into different groups, finding and creating knowledge and sharing it in an online folder, etc. You have used the computer to process large amounts of data, and for what?

In your reflection essay, think about your experiences in each module of the class, and discuss what is gained through these processes, and possibly, what is lost. How do these processes help your research? How do they impede it?

Finally, what do all of these processes and outputs mean in your interactions with the Smith Papers Collection? What did these digital and manual processes reveal to you about these 100-year-old conversations?

9.2 To sum up:

- Discuss the skills, processes, and tools you had to incorporate for each mini-project
- Discuss the effect these processes had on your projects (feel free to reference the course readings to help make this argument), and what these digital artifacts allow (or disallow) you to do
- Discuss the impact these digital processes had on your engagement with the archival collection

9.3 Paper requirements:

- 3-4 pages (1200-1500 words)

- Include 2-5 links to (or images of) artifacts you reference (e.g. visualizations from mini-projects, cleaning configurations, transcription errors, etc.)
- Submit as a word document to Canvas

9.4 Due:

In lieu of final exam (due date will correspond with university's exam schedule) You may share these with me in draft form for comments and feedback before the due date.