

Crowdsourced Digitization

Due: Tuesday, October 19, 5pm. Points, per the syllabus: 1.

Last week we looked at digital datasets. For this week, we will actually contribute to these.

There are many crowdsourced history projects out there. The purpose of this is to get you some more experience with a different type of paper dataset.

1. Choose a crowdsourced historical data project.

I recommend choosing one from the [Zooniverse history projects](#), but you can pick your own as long as it targets a paper database of the sort we have been discussing.

2. Spend 30-60 minutes contributing to that project. (Build in time before this to create an account, read training materials, etc.)
3. Write a short (500 words) response in the assignments tab of the course about the experience.
4. What did you transcribe? What kind of data? Who seems to have sponsored it?
5. What is the value of the data that you transcribed? Is it the same as the sponsor would want?
6. How difficult was it (technically, emotionally, whatever) to do the transcription?
7. What sort of things that existed on the paper did you have to omit?
8. What sort of the things were not on the paper that you would have expected to be?
- 9.

What does it indicate about what kind of data might be “good for digitization” that you wouldn’t have necessarily known before?

Instructor_Name: Benjamin Schmidt Institution: New York University Course_Number: HIST-UA 283
 Course_Title: “The History of Big Data” Email: [bs145@nyu.edu] Office_Hours: Zoom or outdoors, by
 appointment in my e-mail signature. Semester: Fall 2021 Website: “http://benschmidt.org/bigdata21”
 web_theme: journal Code: bigdata21 exclude_dirs:

- tests
- Lectures
- Code days: Tues/Thurs calendar: Fall 2021 NYU link-citations: true background-image:
- ‘BigNumbers.jpg’
- ‘high.jpg’
- “IBM_card_storage.NARA.jpg”
- “Placcius1.png”
- “TimeSheet.jpg”
- “Panorama_of_New_York_City.jpg” course_type: ‘lecture’

...