

LINK TO GITHUB (private, but added Elizabeth as a collaborator):

<https://github.com/asnowmenjig/IS590-Open-Data-Mashups>

REPOSITORY STRUCTURE:

- **"dataset-inventory"** - everything substantive is in here; everything else is previous assignment submissions or other content from class
- **/RETIRED** = 2 datasets I've chosen not to use for my final dataset
- **/eebo|estc|open-syllabus|ota|project-gutenberg-2** = contain files relevant to processing each dataset and moving towards forming the final dataset
- **/dataset-documentation-files** = contains documentation that I submitted for the WK10 assignment and additional notes to myself (cleaning instructions, percent breakdowns of records I'll be sampling from each dataset, data processing timeline, etc.)

DELIVERABLES PROGRESS:

- **Final dataset file** = TBD (to be drafted)
- **Intermediate datasets** = For each dataset, the records that I'll use to form the final dataset are clean and ready for extraction. Each dataset also has its own versioning folder with different versions at fairly consistent stages
- **Dataset documentation files** = At whatever stage of completion they were at when I completed the assignment; I'll go back and finish them when I'm completely done working with the datasets
- **Code** = Need to add comments to code used for processing datasets
- **Notes on hand-editing** = Will add to dataset documentation files where relevant
- **Jupyter notebook** = I have an idea of how I want it to work, just have to actually write the code and create some dummy files for testing
- **GitHub repo** = Chugging along; pretty sure it contains everything except the thousands of RDFs?
- **Résumé entry** = TBD
- **Summary slide** = TBD
- **Assessment report** = TBD
- **Presentation slide deck** = TBD
- **Presentation recording** = TBD