**WEEK 8 (Oct 13 – Oct 19)**

Data management & project management DUE

* “Describe the locations and platforms your data will be stored under” can be in diagram form
* Examples: <https://dmptool.org>

ESTC dataset

* STAGE 2 - data extraction
* STAGE 3 - cleaning and count consolidation

**WEEK 9 (Oct 20 – Oct 26)**

Gutenberg dataset

* STAGE 2 - data extraction
* STAGE 3 - cleaning and count consolidation

Write standardization schema

**WEEK 10 (Oct 27 – Nov 2)**

Data documentation/cleaning assessment DUE

* “start making documentation you’ll need for your final deliverable and assess cleaning needs for each dataset”
* Start skeletons of documentation files
* Discuss cleaning plan
* Assessing cleaning needs = retrospective or prospective (depending on project needs)

Execute STAGE 4 (reformatting and standardizing datasets)

* Each dataset

**WEEK 11 (Nov 3 – Nov 9)**

Final dataset design DUE

* Discuss structure and design of your final data file
  + minimum viable product:
    - - 10-20% sample of 3 datasets
    - - jupyter notebook that extracts data from all three, selects & transforms as necessary, spits out final data file for analysis

Compare datasets and identify significant connections

* Mash together? Side-by-side comparison? Think about methodology

**WEEK 12 (Nov 10 – Nov 16)**

Midpoint check-in DUE

* Show what you have done so far
* Start tidying folder structure

**WEEK 13 (Nov 17 – Nov 23)**

Reproducible Jupyter notebook (draft) DUE

* Outline draft of jupyter notebook
  + Notebook should extract data from datasets, select & transform as necessary, spit out final data file for analysis
* Meet one-on-one to discuss plans, what should be included
* Should not encompass all of your code

**WEEK 14 – THANKSGIVING (Nov 24 – Nov 30)**

**WEEK 15 (Dec 1 – Dec 7)**

Conference talk proposal DUE

* Could reuse parts of recruiter pitch
* Hardest part = speaker bio