

5 Data Management Plan

5.1 Roles and responsibilities

The execution of this data management plan is the responsibility of the project PIs, David Bamman and Taylor Berg-Kirkpatrick. After transference of the data to the Merritt repository at the close of this grant, the University of California Curation Center will hold responsibility for management of the data.

5.2 Expected data

There are three types of data produced under the scope of this grant:

- Manually created training data for page-level classifications
- Manually created training data for intra-page classifications (e.g., paratext regions, including footnotes)
- Manually annotated data for regions of lacunae in documents.
- Code for performing page-level and intra-page level classification, lacuna reconstruction, and compositor attribution

5.3 Period of data retention

All data (annotations and code) will be immediately posted to Github on the publication of research papers that make use of them.

5.4 Data formats and dissemination

All annotations produced in this project will be published as text files in open formats and freely disseminated with a Creative Commons Attribution-4.0 license.

5.5 Data storage and preservation of access

Manually created annotations and code produced under this grant will be sustained in the long term in two locations: on Github (an online version control system), where it will be freely downloadable by others and can be continually updated and improved by other members of the community.

At the end of the grant, we will take a snapshot of all annotations and code and commit them to the Merritt repository of the University of California Curation Center for long-term preservation, including issuing digital object identifiers (DOI) for them.