

Data Management Plan

Types of Data

The data produced in the first prototype will be in the form of Markdown and derivatives (Word, PDF, HTML). In addition we will produce spreadsheets documenting the cost of production of our workflow. The second component will produce an unpredictable set of data depending on the exigencies of the Scalar prototype we will use to disassemble into the repository. This may include images, video, text, sound, maps, etc. In addition to the individual media-specific digital assets, we will count the database itself, a snapshot of the site, and a tarball of all assets as discrete data units.

Data Standards and Capture

For our first prototype, Markdown files encoded in UTF-8 will be mostly produced by authors after receiving adequate training, using text editors recommended by us. We will then process these .md files using Pandoc in order to generate all other derivatives. Eventually these will go into github for production of this part of the site. In the second prototype we will set standards for media types that conform to the Library of Congress "Recommended Format Specifications," in combination with the data standards of CDRS and Academic Commons.

Metadata

In both prototypes, we will charge authors with the task providing certain metadata and will adhere to the metadata standards of Academic Commons. In the case of the second prototype, which may generate additional metadata, we will encourage the use of Dublin Core. This metadata in turn will become an added discrete data unit for us.

Legal Policy

Our authors will retain copyright, inasmuch as they agree to make their research available to the public, once it is in a state that pleases all parties, using an appropriate Creative Commons license. Specifically, we will encourage them to use a CC BY 3.0 US license. Our own content will be available to the public with that license. In the case of the material depositing in Academic Commons as a result of our collaborations, that content will be available under the existing provisions of the repository.

Period of Data Retention

One of the main goals of our project is to prolong the life of digital scholarship affordably and realistically, whether textual or multimodal, to the greatest extent possible. To that aim, we are building systems that ultimately align with the university repository as the best guarantors of longevity available to academics at the moment.

Access Sharing and Reuse

All data generated in this project, aside from the spreadsheets detailing private salary information, will be released to the public in the form of a white paper, a series of research essays published using our minimal computing workflow, and the Scalar prototype with its disassembled components and other discrete data units necessary for its reconstruction.