

DATA MANAGEMENT PLAN

Six Degrees operates under the LOCKSS principle--Lots of Copies Keeps Stuff Safe. Chris Warren is the owner/registrant of *Six Degrees*' shared accounts and emails, with ultimate control over the Google Drive folder and Github repositories. The postdoctoral fellow is responsible for active data management.

Operational Data

All of our operational data--including text documents, spreadsheets, images, and videos associated with the project--is stored online in a shared Google Drive, which all current and former collaborators on the project may access. In addition to Google Drive's versioning system, this data is manually backed up on a monthly basis and these versioned backups are stored on four hard drives maintained in private offices and residences in three separate locations--two within the city of Pittsburgh, PA, and one near Washington, D.C. All hard drives stored in residences are password protected for security purposes.

Datasets

Since all of the subjects of the project are long dead, no privacy issues inhibit data sharing. While copyright prevents us from sharing the textual corpora that are the basis for the project's statistically-inferred data, the datasets themselves are stored online in our main project website. Registered users may already freely download the datasets in .CSV format, under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Unported License.

All our users agree to Terms of Service that allow their contributions to our datasets to be downloaded along with our statistically inferred data. These Terms of Service also make clear that usernames will be publicly visible in order to give users credit for their contributions. Special user classes can further download user-contribution-related metadata--such as timestamps--in the same .CSV files as the rest of the datasets.

The datasets are manually backed up on a weekly basis and these versioned backups are stored as .CSV files on the same four hard drives, in three separate locations, as our operational data. These datasets are also backed up to an account in Box Cloud Storage--a commercial service licensed by Carnegie Mellon University for campus file storage--which the project co-PIs and postdoctoral fellow can all access. The Folger Shakespeare Library, in Washington, D.C., also wishes to host versioned copies of our dataset in their current digital media repository, which will ensure the datasets' long-term preservation. Further programming is necessary to automate the database backups and establish a remote backup with the Folger Shakespeare Library. We consider this a medium-level project priority.

Software Code

All R code for statistical aspects of the project are available online through a public GitHub repository (github.com/sdfb/sdfb_network). JavaScript code for a preliminary version of our website interface--containing most of our network visualization features and using simple, humanist-friendly Google Sheets as the database framework--is available online through a public GitHub repository (github.com/sdfb/sdfb_spring2014IS). The JavaScript and Ruby on Rails code for our current website interface is stored online in a private GitHub repository and will be released to the public at the end of the grant period, a delay which is necessary to ensure the stability and security of the user contribution component of our interface. Prior to the end of the grant, the code may be released to known researchers on request. In addition to GitHub's versioning system, this code is also manually backed up on a weekly basis. Versioned backups of all code - R, JavaScript, and Ruby on Rails - are stored on the same four hard drives, maintained in three separate locations, as our operational data.

Articles and Scholarly Publications

To ensure public access to publicly-funded research, the full text of all articles and similar scholarly publications will be made available via the project website in .PDF format as well as deposited in [arXiv.org](https://arxiv.org), which has been in existence since 1991, and is more permanent than many commercially-published journals. These articles will also be archived through the Modern Language Association's Common Open Repository Exchange (CORE) (commons.mla.org/core/) and Carnegie Mellon University's institutional repository (repository.cmu.edu). These steps will be taken as soon as possible after publication in order to ensure that, even if the project's website folds, there will be permanent public access to the papers.

The full text of all our scholarly publications are also are stored online in our shared Google Drive and the same four hard drives, maintained in three separate locations, as our operational data.