

## **Attachment G. Freedom on the Move – Data Management Plan**

*Expected Data.* The project is expected to produce the Freedom on the Move (FOTM) database and its metadata, source codes of data entry applications and scripts used to automate data processing and management. Fields from the FOTM database mainly come from ads about runaway slaves in the United States pre-1865 scanned from original newspapers and stored in PDF form by researchers participating in this crowdsourcing project and by subscription-based private repositories. The FOTM database contains transcript (text) of the ads, the links to the PDF version of the ad, demographic and physical characteristics of caught and runaway slaves, their children, and their owners; and other information such as newspaper edition and geographic location of the runaway, owner, and newspaper, among others. By including in the database, the actual texts along with the researcher-generated codes derived from the texts, researchers have the option to apply computational analysis (“text mining” tools), or quantitative research techniques, or both to analyze the data.

Static and dynamic versions of the FOTM database will be produced and made publicly accessible for browsing, local analysis, and download for research and analysis with no restrictions imposed on its use.

We will also share the source codes of the web-based user interface so it could be used as a model for other crowd-sourced archival digitization analysis projects.

*Period of data retention.* It is expected that the FOTM database produced by this project will be dynamic and continue to grow as newfound ads are added, transcribed, and processed. As soon as new records are added to the database, the dynamic version, which is the most up-to-date, will be made available immediately to the public. The static version will be released on a regularly scheduled basis, once a year.

*Data formats and dissemination.* We will provide three mechanisms of sharing the FOTM database. The static version of the database will be hosted at the CISER Data Archive. The dynamic version will be made available via the download form on the FOTM website. An SQL script will also be provided so that users can execute the script in their own instances of PostgreSQL (an open source data entry program) and produce a replica of the FOTM relational database including all tables, keys, constraints, and data.

The FOTM database will be made available in CSV format so it can be used in various analytical software packages. The SQL script will be made available in .sql format, which is viewable in any text editor. To enable project and data searching, discovery, versioning, sharing, and access, the Data Documentation Initiative (DDI) metadata standard will be used. DDI allows for the discoverability and access of all metadata pertaining to this project across the data life cycle (i.e., from data conceptualization to collection, processing, distribution, discovery, analysis, repurposing, and archiving). The metadata will reside at CISER’s Data Archive indefinitely. While no data archive can be guaranteed to exist forever, CISER is a university-supported data archive now in its 31<sup>st</sup> year of existence with good support at the highest university levels.

The source codes of the data entry applications or web-based user interface and SQL scripts will be made publicly available at Github (github.com), an open repository for collaboration, review and management of codes. The release of this source code will be concurrent with the release of the FOTM database.

*Data storage and preservation of access.* The metadata and static version of the database will reside at the CISER Data Archive and the dynamic version's FOTM database will be hosted at the CISER Database Production Server. CISER is committed to providing researchers access to the database past the end of the project. Cornell University Library will continue to host the FOTM website and its interface for crowd-sourcing and downloading of dynamic version of the database. Github will continue to be used for sharing of the source codes, the SQL scripts for as long as their use policy does not change. In the event of a change, these codes will be hosted on the CISER Data Archive.