

Attachment 9. Freedom on the Move: Data Management Plan

Expected Data. The project will produce the Freedom on the Move (FOTM) database, its metadata, source codes for data entry applications and scripts used to automate data processing and management. The FOTM relational database will hold all surviving runaway slave advertisements placed in North American newspapers before the end of slavery, gathered 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; as well as additional information such as newspaper edition and geographic location of the runaway and owners. By including the actual texts along with the researcher-generated codes derived from the texts in the database, researchers have the option to apply computational analysis ("text mining" tools), 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 downloading for research and analysis with no restrictions imposed on its use.

Source codes of the web-based user interface, so it could be used as a model for other crowd-sourced archival digitization analysis projects, will also be available for sharing.

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 always be available to the public as long as the FOTM website remains online. A dynamic and ever-growing FOTM database is the optimum solution and our preferred strategy. However, to protect prior investments in FOTM and data generated to-date, a static version of the database will be made available at Cornell University as part of CISER's Data Archive. CISER is a world-renowned social science data archive and is directed by FOTM Co-PI William Block.

Data formats and dissemination. To ensure FOTM access to members of the public as well as professional scholars, we will provide multiple mechanisms of sharing and accessing the FOTM database. First, a publicly-available static version will be periodically updated and hosted at the CISER Data Archive. Second, as long as FOTM is active, the dynamic version will be accessible on the FOTM website and available to anyone with an Internet connection. Third, and intended for scholars who possess or have access to technical skills, 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. Last, for scholars and others wishing to analyze FOTM data in various analytical software packages, the database will also be made available in CSV format and the SQL script will be made available in .sql format (which is viewable in any text editor).

Finally, 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 35th 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.