

DATA MANAGEMENT PLAN

Data management to date

During the startup phase of the project, data and programs recovered from Lawrence University were documented and deposited in MINDS@UW, the institutional repository of the University of Wisconsin-Madison (where Burkert completed her doctoral work in 2016). The scans Daland produced of the printed code base were deposited in the same location.⁵ UW-Madison hosts the items under a non-exclusive license and has agreed to allow them to be saved to the Utah State University repository, as well, in order to maintain a single central access point to project data. In addition, Burkert has maintained detailed project logs since 2013 and keeps the project folders in Dropbox and Box, both of which provide cloud backup and sync to three computers (her personal desktop, personal laptop, and office machine).

Roles and Responsibilities

Hugie (Developer) will work with Betty Rozum (Data Services Coordinator, Merrill-Cazier Library, Utah State University) to ensure storage and preservation in accordance with current best practices. All project participants who edit the data and source code will be responsible for using shared file naming and versioning conventions, updating readme files, and saving incremental progress to shared folders in Box. Hugie will conduct monthly backups to a designated Box folder and to a local hard drive from the beginning of the grant period; he will make long-term storage deposits in Digital Commons and Digital Preservation Network at the end of the project. Utah State University and Merrill-Cazier Library will provide access to all preservation solutions (detailed below) at no cost and will be responsible for decisions about the data over long term.

Expected Data

The project is expected to generate up to 1GB of data, including:

- Flat .txt files representing multiple versions of the flat-file database, from the dirty recovered data through several stages of cleaning and parsing
- XML-tagged data representing all the entries in the flat-file database
- JSON-tagged data, again representing all the entries in the flat-file database
- Relational database (likely in MySQL or an open-source equivalent like MariaDB) populated with the XML- or JSON-tagged data
- Software code (in Python) for cleaning, parsing, and tagging data
- Website code (PHP, XML, JavaScript, HTML, XSLT)
- Documentation, including the database schema and README files, stored as plain-text .txt files
- The 2018 meeting report, peer-reviewed publications, and 2019 white paper, stored in PDF and .txt file formats

Period of Data Retention

USU will provide access to server space for the website and will maintain access to the Digital Commons repository for as long as reasonably possible.

Data Formats and Dissemination

The flat-file data will be maintained as .txt files. It will then be tagged in the preservation-friendly XML and JSON formats. This structured data will be imported into MySQL or its open-source counterpart, MariaDB, and end-user access to the database will be provided through a web interface. The server space will be provided by USU Central IT. Users who wish to download the raw data and documentation in its entirety can do so through Digital Commons, the USU institutional repository. The documentation will include the version(s) of software used to create the database on which the code runs. All code will be made openly available on GitHub. To the extent allowed, publications resulting from analysis of the data

⁵ <https://minds.wisconsin.edu/handle/1793/71768>

will be deposited in Digital Commons as well, either as post-prints or in the version of record. Our goal is to make all project outputs, as well as all procedural and contextual information, open access and open source.

Data Storage and Preservation of Access

All of the data and software code described above (.txt flat-file data, XML-tagged data, JSON-tagged data, database schema, programs, readme files, and additional documentation) will be bundled using BagIT and uploaded to USU Digital Commons repository for preservation. Digital Commons is backed up to Amazon S3. We also have arranged to deposit the files in Open SIUC, the institutional repository of Southern Illinois University, in accordance with our agreement with SIU Press, the publisher of the original reference books *The London Stage*. Long-term preservation will be provided through the Digital Preservation Network, which is one of the most robust digital preservation solutions available. The Merrill-Cazier Library at USU will provide cost-free access to all of these preservation systems.