DATA MANAGEMENT PLAN

Expected data:
The "Integrating digital humanities into the web of scholarship with SHARE: an exploration of requirements" project will produce a number of outputs and digital assets. The first phase of the project involves gathering DH requirements to enhance the SHARE data set; this will be conducted via an online survey and in-person focus groups. The survey will produce a survey instrument, raw results, analysis scripts, and the analyzed data. The focus groups will produce notes, code book, coded answers, and the focus group script and question set. The next phase of the project involves a workshop to explore the alignment of the SHARE Curation Associates program and the CLIR Curation Fellowship. The expected outputs of this phase include notes from the workshop, presentations, and a final report. The final phase of the project will results in the development of technical prototypes, including software code, scripts, and wireframes. Finally, the results of the technical evaluation of these prototypes will be a result of this project.

Period of data retention: The project team will make the resulting assets (redacting any personally identifiable information) available to the public as each phase of the project comes to a completion. The project team will use the Open Science Framework to manage the project workflow, collect documentation, and make the assets public upon completion. These assets will also be archived and curated in the Washington University in St. Louis institutional repository, Open Scholarship, where they will be retained and curated for a minimum of 10 years.

Data formats & dissemination: Whenever possible, open data formats or formats that do not use closed proprietary specifications will be adopted as asset accessibility and archiving standards for the project. For example, all text/data will be encoded using Unicode to prevent data loss. Uncompressed TIF (or comparable) will be used for all images. Archival copies and originals of the data will be maintained according to the WU Libraries archiving policies outlined below.

Metadata will be created and saved throughout the lifecycle of the research project and will be in line with the commonly accepted scholarly standards. All collections will have a DublinCore metadata record created. This record includes elements such as author, license, abstract, publication date, funder, and others. As appropriate, the Data Documentation Initiative (DDI) metadata standard will be used to describe data granules (individual files rather than collections) and survey instruments. For continued preservation of the data and materials, metadata elements consistent with the PREMIS data dictionary and data model will be implemented.

A "read-me" file will also be created that includes an asset inventory, general rights for reuse, contact information, a recommended citation, and a synopsis of the project.

To improve dissemination of the research outputs a digital object identifier (DOI) will be assigned to the collection or at necessary granularities of the project. Additionally, metadata records are propagated to metadata harvesters, such as SHARE through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

The dissemination information package (DIP) will include access copies of the assets, any required analysis code or software, the data dictionary/code book, the DublinCore metadata record(s), the discipline metadata record(s) and the "read-me" file.

Data storage & preservation of access: As previously stated, day-to-day project management and storage will take place on the Open Science Framework (OSF). The OSF will allow the project team to integrate with existing systems, such as box for large file management. Given its design, it also facilitates the creation and storage of various asset types, including text, tabular, images, and more. Once a phase of the project is complete, the OSF project space or component will be made publically available.

Digital assets archiving and preservation is supported through an approach that starts with adequate documentation of data using metadata and other formats appropriate for long-term preservation. Unique digital assets from this project will be redundantly deposited in the Open Scholarship repository, and archived according to the Open Archival Information System (OAIS) framework.

The archival information package (AIP) will contain all the materials previously mentioned in the DIP, a copy of the untreated, original submission information package, the PREMIS compliant metadata, a checksum manifest, and a curation treatment actions file.