

Data Management Plan

This plan is provided according to requirements set forth in the NEH Digital Humanities Start-up Grants. It explains how the project team will manage and disseminate original student research described in the narrative portion of this proposal.

Data Description: This study will produce original student translations of primary source documents which were chosen by the *Tisch Library Special Collections Department* in collaboration with the principle investigators. These translations will be encoded in XML files according to [Text Encoding Initiative](#) (TEI). The translations will also be linked to bibliographic records for the primary source documents. These bibliographic records also will be encoded in XML. Both student translations and the associated records will be generated and maintained in an online digital toolkit.

Data Storage and Preservation of Access: The principal investigators along with librarians from Tisch Library, Tufts University will have overall responsibility for data management over the course of the research project and will monitor compliance with the plan. We will also investigate ways to take advantage of the chain of authority supported by the digital tool-kit to facilitate the process of archiving the resources. The principal investigators and librarians will work to identify the best way to archive the data at the end of the active data generation life-cycle.

Access and Sharing: The project team will make the research data from this project available to the broader humanities community throughout the project's information lifecycle by the use of appropriate digital platform tools. These tools will be online, hosted by a variety of servers, including the *Perseus Project*, and will generally be open-source.

Tufts University maintains a distributed information technology environment, with central as well as local aspects of overall planning and control. Tufts' information security program is structured in a similar manner. Operationally, Tufts central IT organization (UIT) and each local IT group maintain standards of quality and professionalism regarding operational processes and procedures that enable effective operational security.

All the project data will be stored using Tufts University Information Technology resources. Research data for this study will be stored on Tufts University's centrally provided "shared research storage facility", which is provisioned on Tufts' NetApp enterprise networked storage. Provisioning occurs through self-service at <http://researchstorage.uit.tufts.edu>. Data security and confidentiality are protected by using Microsoft Active Directory authentication, and the storage is backed up to LTO-4 tape on a daily and weekly basis and stored offsite at Iron Mountain facilities. As storage needs for the research study increase over time, additional storage can be easily provisioned on the same appliance. UIT issues storage space in 50GB increments, up to several TB. All devices and users are subject to the [Tufts Acceptable Use Policy](#)

Selection and Retention: The principal investigators will work with the project's librarians and Tufts Digital Collections and Archives (DCA) to identify the best way to determine the data to be selected for retention and archiving.

Metadata: Substantive metadata will be provided in compliance with the most relevant standard for the digital humanities -- TEI. This XML standard provides for the tagging of content, which facilitates preservation and enables flexibility in display. These types of metadata will be produced during the active phase and ultimately archived:

- *Bibliographic Metadata Record.* A summary TEI and MARC-based record will be created for inclusion in the online discovery tool created for the project, Tisch Library's OPAC and DCA's discovery platform. This record will be indexed with terms from Library of Congress authorities and subjects to enhance data discovery. This bibliographic-level record will provide access to the research and translations created by the student.
- *Data Citation with Digital Object Identifier (DOI).* A standard citation will be provided to facilitate attribution. The DOI provides permanent identification for the data and ensures that they will always be found at the URL specified.
- *Technical Documentation.* The files described above will serve as the foundation for the technical documentation or codebook that DCA in collaboration with the Tisch Library Technical Services Department will prepare and deliver.

Ethics and Privacy: The data to be acquired in the proposal project do not involve human subjects. The data acquired, retained and preserved as part of the proposed research will be governed by [Tufts University's Policy on Rights and Responsibilities with Respect to Intellectual Property](#).

Format: *Access:* During the active phase the project team will work with librarians in establishing the most appropriate manner to make the data files available to the broader scholarly community. *Preservation:* Data will be stored in accordance with prevailing standards and practice. Documentation is preserved using XML and PDF.