

Data Management Plan

DATA DESCRIPTION

Two distinct classes of research products require sharing and archiving: software and collected artifacts from our interviews with poetry scholars. This second class of products will include notes from interviews and screenshots from sessions with our software.

ACCESS AND DISSEMINATION

Software developed for this project will be written in languages such as Processing, Processing.js, C, C++, and Matlab. Both the code and executables will be released publically. All distributed software will be provided under an appropriate open source license.

Centralized access to software will be managed by the Co-PIs of the project. The primary repository of software will be in the University of Utah's Scientific Computing and Imaging (SCI) Institute. A project web site, hosted by the SCI Institute, will provide free public access. The site www.poemage.org will redirect to this site. We will provide documentation for the software executables as well as for the source code.

We will adhere to NSF guidelines for the dissemination and sharing of the software. We will promptly prepare and submit for publication and/or conference presentation all significant findings from the work that we conduct, with authorship that accurately reflects the contributions of those involved. In addition, we will share with other researchers within a reasonable time the data that we have collected in the course of NEH work, as long as the sharing of materials is not excessively costly.

All distributed software will be provided under an appropriate open source license. The publicly available project web page will serve as a common portal for all publicly released data stemming from this work. No fees will be charged for access to this data. No significant ethical or privacy issues arise in releasing this data.

Access to the collected artifacts from interviews will be managed by the Co-PIs of the project, and used solely within the project group for refining the software. We have applied for IRB approval for the participant support indicated on this project. It is anticipated that this will be found exempt according to IRB protocols.

Data will be made available for sharing to qualified parties by the Co-PIs, so long as such a request does not compromise intellectual property interests, interfere with publication, invade subject privacy, betray confidentiality, or precede data curation. Data that are shared will include standards and notations needed to interpret the data, following commonly accepted practices in the field. In the event that discoveries or inventions are made in direct connection with this data, access will be granted once appropriate invention disclosures and/or provisional patent filings are made. Key data relevant to the discovery will be preserved until all issues of intellectual property are resolved.

Copyrighted publications have reuse and redistribution terms established by the publisher or copyright holder. We will make publications stemming from this work as openly available as is permitted under these terms.

ARCHIVING AND PRESERVATION OF ACCESS

The software and data will be maintained and archived at the SCI Institute indefinitely. The SCI Institute maintains three separate backup systems managed by a dedicated backup system. The backup server controls three tape libraries and is connected to SCI infrastructure with 2x10GB network connections. A three drive LTO-3 tape library for archiving data for off site storage with slots to hold a 50TB(compressed) single load tape rotation, all directories are fully archived once a quarter. A six drive LTO-4 for on site data backup with a 417TB(compressed) single load tape rotation, all directories are backed up fully once a quarter with nightly incrementals and by weekly synthetic differentials. A single drive LTO-4 tape library with 12TB(compressed) single load tape rotation capacity backups infrastructure systems such as DNS, DHCP, mail, and accounting data.