

6. Data Management Plan

Expected Data

A variety of data types are expected to be generated as a result of this grant. The primary data to be generated by this project will be the programming code of the interface plugin itself—a BuddyPress-compatible WordPress plugin (PHP) with accompanying JavaScript and CSS, comprising approximately 10,000 lines of code. All project staff will check the code in and out of a Git repository, hosted centrally through the MLA's private GitHub account. Documentation of this code will be produced alongside the code itself in PHPDoc format, with accompanying Markdown-formatted configuration and installation instructions. Some further data may also be collected to assess the usability of the user interface plugin, using bibliometric, qualitative, and click analytic methods. Minimal data generation is expected using these methods. Both the collected and the analyzed data will be anonymized or aggregated to the extent necessary to obscure any individual participant's identity before they are shared. Prior to sharing, collected and analyzed data will be kept in private spaces on the MLA and CUL networks, backed up securely, and accessible only to project staff.

Period of Data Retention

In accordance with the Columbia University Retention and Access to Research Data policy, the implementation project staff agree to retain all project data for a minimum of three years after the period of the grant. All project data to be made publicly accessible will be deposited in public repositories (such as *CORE*) within 30 days of the submission of the final report to the NEH. At the project outset, no embargo periods are expected.

Data Formats and Distribution

Final project data connected to the *Humanities CORE* software along with its associated documentation will be fully released upon project completion through open repository and open-source software distribution platforms. Where permissible, anonymized and analyzed assessment data will also be published openly in connection with the final project deliverables. Data reuse conditions will be made explicit upon publication, with a preference wherever possible for maximum potential reuse. This includes data in all expected formats: source code; documentation in text, word-processing, and printable formats as necessary; and documentation in all other formats aggregated over the course of the project period. Preference will be given at all times to platforms and formats that favor open distribution and technology free from proprietary limitations, in pursuit of the goal of data freedom upon project completion (within 30 days of the submission of the final report to NEH). No embargo periods are expected at the project outset, and it is not anticipated that any of the project data will be subject to confidentiality concerns. Access to project data will be facilitated by the policies of the open repositories that will be used to publish them. No access arbitration will be necessary for project data.

Data Storage and Preservation of Access

The investigators will use *CORE* as the primary preservation and access platform for the final project data. Domain-based repositories may also be employed where identified as trustworthy and beneficial to project data during the project period. Deposit in *CORE* provides a persistent URL, secure replicated storage (multiple copies of the data, including onsite and offsite storage with verified checksum procedures), accurate metadata, a globally accessible repository and the option for contextual linking between data and published research results. Any file type may be deposited in *CORE*. Files are available for public consumption and reuse, and for machine extraction.

Roles and Responsibilities

Throughout the Implementation grant period, access to the project data will be limited to the named project staff and the CUL system administrators. Responsibility for data management will remain with CDRS. Adherence to this Data Management Plan (DMP) will be the responsibility of the designated Project Manager, who will ensure the following with respect to software developed: (1) that project staff are following the agreed-upon practices for code versioning and development, (2) that the data archives are comprehensive throughout the period of the start-up grant, and (3) that the source code for the final production code is published to an openly accessible data repository. The project manager will likewise ensure appropriate practices with respect to data produced during the assessment portions of the project. If the Project Manager should need to be replaced midway through the period of the start-up grant, it will be the joint responsibility of the Principal Investigators (PIs) to periodically audit the data practices of the project staff until a new Project Manager is named.