

## **7. Data Management Plan**

### **Roles and Responsibilities**

Throughout the Start-Up II grant period, access to the project data will be limited to the named project staff, the staff of CBOX, the staff of Cast Iron Coding, the staff of the Center for Digital Research and Scholarship (CDRS), and the LITO 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 (1) that all 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. 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.

### **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. All project staff will check the primary data (programming code) in and out of a Git repository, hosted centrally through a private GitHub account supported by the Libraries Information Technology Office (LITO) of the Columbia University Libraries/Information Services. Documentation of this code will be produced alongside the code itself. Some qualitative data may also be generated with regard to assessing the usability of the user interface plugin. These data will be anonymized or aggregated to the extent necessary to obscure any individual participant's identity before they are shared.

### **Period of Data Retention**

In accordance with the Columbia University *Retention and Access to Research Data* policy, the start-up 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 Columbia University's Academic Commons) 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**

Project data will be fully released upon project completion through open repository and open-source software distribution platforms. This includes data in all expected formats: source code; documentation in text, word-processing, and printable formats as necessary; XML data; and documentation in all other formats aggregated over the course of the project period. Preference will be given at all times to platforms 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). Only XML schema with adequate documentation will be

used for encoding project data. 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 Academic Commons, the Columbia University digital research repository, 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.

### **About Academic Commons**

Depositing research or scholarly works in Academic Commons provides depositors with a permanent URL, secure replicated storage (multiple copies of the data, including onsite and offsite storage), accurate metadata, monthly usage reports, a globally accessible repository, and the option for contextual linking between data and published research results. Academic Commons employs the Metadata Object Description Schema (MODS), version 3.4, an XML-based metadata schema developed by the Library of Congress. This schema is appropriate to both traditional publications, as it is Columbia University Libraries/Information Systems' preferred descriptive metadata schema, as well as data, as it maps clearly and easily to those elements noted in the DataCite metadata schema, version 3.0. Files deposited in Academic Commons are stored in the Libraries' long-term digital preservation storage system, which includes at least two copies: one stored on-campus in Morningside Heights in Manhattan and another stored at the University disaster recovery site in Syracuse, New York. The system verifies with an MD5 checksum that these copies are identical on write.