

## **Data Management Plan**

### **1. Roles and responsibilities**

System architect Dale MacDonald will have overall responsibility for data management over the course of the research project and will monitor compliance with the plan. He will liaise with staff at each collaborating institution to ensure that the appropriate data is preserved and accessible. Mobile application infrastructural material and content will be stored indefinitely on servers managed by one or more of the collaborating institutions, which will provide server redundancy, 24x7 monitoring, and regular server backups.

### **2. Expected data**

Five types of data will be generated by this project: web-based mobile application infrastructure, descriptive data pertaining to physical artifacts, mobile application content, training and curriculum materials, and evaluation instruments and data. The web-based mobile application infrastructure includes dynamic, PHP-driven Drupal sites to input, edit, and display user- and project-generated content. Application structure and the “look-and-feel” are stored in a MySQL database (typically) local to the Drupal server. Descriptive data pertaining to physical artifacts includes 2-dimensional digital imagery of the artifacts and relational databases of descriptions of the history, content and techniques of the artifacts, including textual, audio, photo, and video data. Mobile application content includes a MySQL database of introductory, instructional and explanatory texts. Training and curriculum materials will include project documentation on the mobile application infrastructure and its use as well as course guides in HTML, Wiki, Word document and/or PDF format. Server logs and issue tracking documents in text format will be maintained for troubleshooting and interface analysis.

### **3. Period of data retention**

Source code, wiki-based documentation, and issue tracking documents (bug reports) will be indefinitely housed in a publicly available, version-controlled repository. Data and metadata about the cultural artifacts will be stored on collaborating institutions' servers indefinitely.

### **4. Data formats and dissemination**

Data pertaining to the cultural artifacts will be placed into repositories using the Qualified Dublin Core metadata schema. A copy of the digital master will be stored on collaborating institution-managed servers and derivative formats for online access will be generated (jpegs for images, mpeg-4s for moving images, etc.). As standard practice, the collaborating institutions will submit metadata to a federated database, and submit web links to search engines and directories. The app and its content will be available through any device with browser access to the web. Media format choices were based on widespread browser compatibility. Project-related publications will be made accessible through the collaborating institutions' public portals as publication copyright allows. These repositories will provide: a permanent URL, secure replicated storage (multiple copies of the data, including onsite and offsite storage), accurate metadata, a globally accessible repository, and the option for contextual linking between data and published research results.

### **5. Data storage and preservation of access**

All collaborating institutions are committed to preserving the Quilt data, websites, and tools indefinitely and updating file formats in a timely manner to be currently accessible as technology changes. Letters from the Digital Studio for Public Arts and Humanities at the University of Iowa and from the School of Media Studies at The New School state the commitment by each institution to support the data storage and preservation of access to applications. Issues of accessibility will be addressed from the beginning in the design of web interfaces. Each run industry standard servers in a climate controlled, physically secured rooms with replicated storage

(multiple copies of the data, including onsite and offsite storage). Offsite storage will additionally be accomplished by placing all source code and issue tracking data on one or more commercial repositories such as GitHub or BitBucket and non-textual content uploaded commercial services (e.g., video to YouTube, audio to SoundCloud, images to Flickr).

- **Digital repository:** The primary candidate for storage of both datasets and digitized cultural objects is MATRIX's open source digital repository application KORA (developed with funding from the NSF Digital Library Initiative, IMLS, and NEH). KORA's architecture is unique in that it can accommodate any set of metadata schema in individualized digital datasets. Project staff can easily create metadata elements using a simple point-and-click interface, select the type of form control for each element (e.g., required formats for date, URL, audio file upload, etc.), and then determine whether the element is required for each record and if it should appear in database search returns and advanced search feature. KORA automatically generates storage structures, ingestion forms, and validation requirements for each metadata scheme.
- **Interoperability:** This project will comply with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) standard for sharing metadata so metadata about digital objects may be harvested and aggregated for use by other repositories. This work will build upon KORA's already-implemented MySQL and XML platform-agnostic data design in a manner that allows extensible interoperability between repositories.
- **Digital Preservation Strategies:** Each collaborating institution will specify and develop digital preservation strategies appropriate to their local needs and infrastructure. At a minimum each strategy will encompass a secondary preservation storage tier with regular checking of fixity and readability along with a format migration plan to maintain presentability as technologies change.