DATA MANAGEMENT PLAN

Data Generated through Project

Data Type	Sharing Timeline	Conditions
Raw image data and video files (e.g., digital photographs, vector line drawings, 3–D laser scans)	On completion of Start – Up phase and as assets are added thereafter	Free and open public access via project website and, ultimately, user interfaces*
Descriptive meta–data in database	On completion of Start – Up phase and as data is added thereafter	Free and open public access via project website and, ultimately, user interfaces
Digital assets and interpretive materials (e.g., academic papers, interactive maps, architectural reconstructions)	On completion of Start – Up phase and as assets are added thereafter	Free and open public access via project website and, ultimately, user interfaces*
Analytic and methodological data (e.g., iconographic seriation, chronological analyses, methodological outlines)	On completion of Start – Up phase and as added thereafter	Free and open public access via project website and, ultimately, user interfaces
Open–source computer code associated with database, tool, interface, and server–side components development	Following Start–Up phase testing and as developed thereafter	Free and open public access via project website and open–source code repository (e.g., GitHub)
Assessment and evaluation data generated during the testing phase and through feedback	Aggregated data will be shared via the white paper and final report	No personally identifiable information will be shared
White Paper	On completion of Start–Up phase	Free and open public access via project website
Final report to NEH	On completion of Start–Up phase	Dissemination to be responsibility of NEH

*   *Where applicable, copyrighted data will be dynamically excluded from publicly available collections and documents and made available at the discretion of the creator.*

Period of Data Retention
We anticipate making all data generated by project staff available immediately on completion of the start–up project. As data are added in subsequent project phases, they will become immediately available, except in those cases where copyright restrictions may apply (e.g., user–generated data or materials).

Data Formats and Dissemination
Access to all raw data, imagery, descriptive meta–data, interpretive materials, and digital assets related to our project will be available for download through the project website as they are added and ultimately accessible through the interfaces linked to the centralized database. We will follow the recommended best practices of the Text Encoding Initiative (TEI) and the Digital Antiquity/Archaeology Data Service Guides to Good Practice, adhering to established protocols for long–term storage of digital data files. Data files will employ accepted archiving standards and non–proprietary formats to ensure accessibility, such as the use of archival PDFs (pdf/a), JPEG/PNG image files, AVI media files, and geoTiff imagery. All project materials will be licensed under a Creative Commons Attribution–NonCommercial 3.0 Unported License, thus permitting free, noncommercial access, use, and academic exchange of contents. We reserve the right to control the commercial use of the database and interfaces we develop, primarily to ensure that any commercial use benefits the long–term viability of the project. Structured meta–data will employ Data Documentation Initiative (DDI) standard XML formatting, offering flexibility in display and preservation–ready, machine–actionable quality. Database materials will also be tagged with actionable, persistent, globally unique ARK (Archival Resource Key) data identifiers, ensuring stable long term

23

access to distinct versions of datasets or their individual components. ARK identifier schemes allow users to query both descriptive and archival meta–data and to recognize relationships between identifiers.

The use, distribution, and modification of any computational tools, techniques, or methodologies developed for the project will be open. Source codes and software will be made publicly available through the project website and through public repositories such as GitHub (https://github.com).

Data Storage, Management, and Maintenance
Data will be archived and disseminated through a dedicated website that project staff will create and maintain, to be housed through the institutional repository hosted by the Florida State University's Department of Art History and Information Technology Services (ITS) and managed through WebDAV protocol. This storage is provided through an enterprise class network and includes automatic mainframe and cloud–based backup. FSU ITS also provides data stewards and custodians, as well as database and security administrators to manage archives following ultimate project completion. Project co–directors will continue post–grant curation of the database and website.

Project data will be stored in a NoSQL database for long–term archival storage. NoSQL databases provide the ability to grow and change through time with minimal system interruption. For instance, if a better algorithm replaced the initial set of icon matching algorithms, the addition of required data would be less disruptive of the system, thus providing important flexibility to application stacks. Since the data model is not rigidly defined, individual records can be different without disturbing the system as a whole. By spreading the backend database among several physical and geographically remote servers, NoSQL systems also offer extraordinary resiliency and fault tolerance, providing uninterrupted uptime and continuous user access to cloud data with no single point of failure, a crucial consideration for interactive, digital publications where the user simply expects the cloud data to be available when they are connected. Such systems also support off–line access to the database through synchronization. During the grant we will evaluate the NoSQL technologies that are appropriate for our mobile interface.

Database architecture and content management technology will utilize an open–source application such as MongoDB, PostgreSQL–based HadoopDB, or CouchDB that makes use of NoSQL technologies. Another possibility under evaluation is the Apache Cassandra database with geoaware Solr as a data and document store, which permit location–aware searching and carry the added benefit of a query language (CQL) that is similar to the SQL used in relational databases. These technologies combine maximal fault tolerance, the speed and flexibility of key–based storage, and the ability to move data and specialized data structures between client apps and our distributed, centralized database. Data would be stored as JSON "documents." The latter point is critical for the development of non–web–based applications. The functionality we envision will be best supported by a system that can update and sustain a database built into the application, providing access to enhanced information when online, but allowing for continued use when off–line. Traditional relational databases are less well suited to the powerful opportunities these new technologies present. Available Application Programming Interfaces (including C, Objective C, Java, PHP, and Ruby bindings) support both web server and other client–server models. These open–source systems and software will ensure accessibility and allow for the future expansion of core database assets during and beyond the duration of the project. Our open database will allow for the creation of other novel applications by developers and users other than ourselves. These data management technologies also enable an efficient workflow, as content contributors and editors can have direct access to editable areas of the site through an administrative backend. Thus, elements can easily be added through time.

The long–term development of the digital product demands the highest standards of scholarship, editing, and software design. We plan to partner with an academic press to ensure peak production values and consistency. We also plan to incorporate the participation of the Florida State University Facility for Arts Research, which is dedicated to the integration of traditional print and electronic media. Since *The Mesoamerican Corpus of Formative Period Art and Writing* will be a digital publication in conjunction with traditional media, it will be cataloged as a resource via such research hubs as the Foundation for the Advancement of Mesoamerican Studies, Inc. (www.famsi.org) and Mesoweb (www.mesoweb.org).

24