

The David Livingstone Spectral Imaging Project

Data Management and Sharing Objectives

(Internal Document, Nov 2010)

The Nyangwe Field Diary will be provided for public access on the internet as a complete package of images with documentation and full metadata. The resulting archival data set will be based on the archive and metadata model used for the Archimedes Palimpsest and other spectral imaging projects.

The pilot project requires two key deliverables:

1. Standard tagged TIFF image products of the Nyangwe Field Diary for selected imaging modalities, including visible, UV, IR, Pseudocolor, accurate color and raking light (where available). The various images of each leaf will be registered as much as possible. If needed, a compressed subset of the images may also be provided for increased ease of web access.

2. Marked up digital XML transcriptions and apparatus that conform to the TEI P5 guidelines.

The full set of data and metadata will be stored in a proven archive format used on other spectral imaging projects, including the Archimedes and Syriac Medical Palimpsest projects. This archive model is designed to be self-documenting for human and automated access, and provides data and metadata in a regular and predictable structure. The archive design is based on principles set out by the Open Archival Information System (OAIS). The data and metadata to be archived for the images and transcriptions will follow the Archimedes Palimpsest Metadata Standard (available at http://www.archimedespalimpsest.org/programmanage_documents.html). This standard is based on accepted, international consensus standards, including Dublin Core Metadata Elements. The metadata is included in the TIFF headers to ensure preservation with the archival images.

The image and transcription products are intended to stand alone, as well as link with each other for specific applications and studies. Identification metadata are amongst those that will follow the Dublin Core Metadata Standard. This metadata is intended to meet the needs expressed by various users, and to provide an open platform for work by those users who have expressed interest in building on the distributed dataset. The digital product will include XML-encoded mappings to locations on the images corresponding to the lines of text of the transcriptions. These mappings will help readers and support software tools to relate passages on the images to the transcriptions and vice versa.

Metadata has not been limited to identification of the images, but also information about the imaging parameters for each of the multiple images in a spectral image cube. Our

team also records structural and descriptive metadata, along with camera and lighting specifications, the algorithms used to process images, and software information. After imaging and processing is completed, the imaging logs will be collated with other required metadata to generate complete metadata records for raw and processed images, and to assemble the final data archive.

Specifically, the archive will include:

1. A complete package of raw spectral images with documentation and full metadata that meets international library and archival standards (the data set will be based on the archive and metadata model used for the Archimedes Palimpsest);
2. The images stored as “flat files” that are not dependent on any specific graphical user interface (GUI) so as to ensure long-term viability;
3. Pseudocolor processed images to highlight the overtext (Livingstone’s handwriting);
4. All other intermediate or final processed images produced to enhance the overtext, including images tailored to meet users’ needs or highlight document-specific areas of interest.

Our policy of sharing data with others focuses on free and open access, with all data available to the public under the Creative Commons Attribution-Noncommercial 3.0 Unported License (see <http://creativecommons.org/licenses/by-nc/3.0>). We have obtained permission from the copyright holders of Livingstone literary works to publish all our project results under this license. Furthermore, the transcription standard for the Livingstone Data Product will provide an end product that meets globally accepted standards. Scholars will use a range of tools and work processes that provide transcriptions compatible with the standards defined for the Data Product.

Back-up data hosting will be provided by the UCLA Digital Library Program. All digital files created by and for the UCLA Digital Library program are stored on the UCLA Library’s secure file system and backed up daily. The Library’s digital library metadata collection system includes administrative and technical metadata for all master files and their derivatives. Master files are maintained on the file system both to facilitate access by collection curators, and to allow for their maintenance.

In addition to local storage and backup services, the UCLA Digital Library works with the University of California Curation Center to provide a full set of services to support digital preservation of content. These services include the assignment of globally unique identifiers to all digital content; redundant and geographically diverse digital storage; verification of bit-level integrity of content files; inventory, indexing and search services; and transformation services (forward migration of file formats when needed). These services together provide the UCLA Digital Library with the ability to provide various levels of preservation services, from simple bit-level preservation (i.e. what you put in

you can get back), to a full digital preservation service that includes replication, integrity services, and forward migration of file formats.

The relationship with the UC Curation Center provides the UCLA Digital Library with an important degree of confidence that our digital assets will be preserved for all future generations of scholars.

UC Curation Center References:

Home page: <http://www.cdlib.org/services/uc3/>

Digital Preservation Repository: <http://www.cdlib.org/services/uc3/dpr.html>

Products and Services: <http://www.cdlib.org/services/uc3/products.html>

Curation Micro-Services: <http://www.cdlib.org/services/uc3/curation/index.html>

Finally, the product and all supporting images will also be disseminated on hard drives to designated institutions for hosting on their internal networks and external Web sites and archiving.