

## Data Management Report

Ashanka Kumari

November 23, 2013

Between 2005 and 2010, Livingstone Online (<http://www.livingstoneonline.ucl.ac.uk/>) partnered with some 15-20 archives around the world to collect TIFF images of David Livingstone manuscript holdings. The archives supplied these images on CD-Rs and, in most cases, used custom directory and name schemes for the files. During this time, Livingstone Online's photographer also captured images of some Livingstone manuscripts.

Once captured and delivered to Livingstone Online, the project staff -- most likely in this order -- converted the images to JPEGs, separated the images into discrete objects, collected the objects into directories with names linked to the Clendennen and Cunningham Catalogue of Livingstone Documents (1979, 1985), and renamed the images with a simple file naming scheme (1, 2, 3, etc). The team left the original TIFF images on the CD-Rs and did not otherwise organize them. In a few cases, probably because of accident or oversight, the original CD-Rs have disappeared so that the JPEG derivatives are all that remain of the original TIFFs.

As part of the Livingstone Online Enrichment and Access Project (LEAP, 2013-16), the project team, now co-directed by UNL assistant professor Adrian S. Wisnicki, is returning (where possible) to the original TIFF images in order to create a core data archive that will contain the images in a regular and predictable directory structure and will provide appropriate metadata, which is also not directly available for the TIFF images from the 2005-10 phase of the project. This archive will, therefore, digitally preserve these images in the highest possible quality and facilitate discovery, access, and interoperability, particularly as users will be able to download the images as "flat files."

As a result, the first step to achieving these objectives lies in converting the existing TIFF images (and in some cases, JPEGs) from the non-systematic organizational scheme of the original CD-Rs (now transferred to a hard drive) to a directory structure that is more user-friendly and manageable.

### **The following report details my work with reorganizing these directories:**

As Dr. Wisnicki's research assistant, I was given the task of assisting him with this directory organization project. We started with the large directory of existing image files. The 10,000+ image files were located on a portable hard drive provided to me by Dr. Adrian

Wisnicki. This hard drive is where I did my work. In the beginning, we started with around 20 directories from various institutions with mixed names. Clearly, there was not a distinct labeling system in place for the directories. From here, my work began by renaming all of the main directories using the C&C labeling system used by the website and some of the institutions. All of the work I did for this project was continuously updated in excel spreadsheets.

After renaming the directories, I went in and created TIFF and JPEG subdirectories in each institutional directory to organize the image files. This reorganization allowed us to group the images by type more easily, which was the next step. We organized images into groups containing letters, photos, and miscellaneous files in each directory.

My remaining work focused on the letters, which are the vast majority of the files in this image data directory. I began by going image by image and comparing them to the data on the Livingstone Online website in order to determine which images went together to form a letter and what the C&C number was for that particular letter. After determining the correct C&C number, I created a subdirectory named with that C&C number and placed the coordinating images into the new subdirectory. Each subdirectory was completed by adding additional TIFF and/or JPEG subdirectories and placing the image files accordingly to help with ease of access. For example, each directory now looks like this: Institution C&C Name → List of Letter C&C Numbers → (each containing corresponding) TIFFs or JPEGs subdirectories → Corresponding Letter Images.

At this point, each existing set of letter image files can be accessed quickly by locating choosing the institutional directory and the corresponding C&C letter subdirectory; however, the Livingstone Online staff still has work to do. The staff needs to compare the core TIFF/JPEG image C&C directories to Livingstone Online and determine which TIFF images are missing, create a master directory that lists all of the C&C numbered subdirectories, compare this master directory to the foliation code and records spreadsheets to identify any objects that may be missing, review TIFFs for all of the C&C numbered objects in the master directory to ensure all of these objects have all the necessary images, select master images for each object, and determine what to do about images that contain more than one manuscript page.