

**Livingstone closeout**

Saturday, June 4, 2011 3:21 PM

**From:** "Doug Emery"**To:** "Dr. Adrian S. Wisnicki"

"Mike Toth"

Mike, Adrian,

Below I give a list of tasks to complete the assembly of the Livingstone data set. I will have this done by the end of July.

There are number of issues and questions to resolve that I mention in the task list. Note that I believe I don't yet have a copy of Keith's images. I have only Bill's color images. One of my main concerns is how many separate operations this will require to create distinct metadata for each type of image of each of seven documents. If we have seven image types for each document, that works out a series of 49 operations for each group.

I'm going to look into a way to handle all the documents together and retain their separate identities in the metadata. If it looks like this process will take many hours to complete, I won't be able to afford that or meet my other responsibilities. In that case I will need a way to simplify things so that we still have a quality result that comes in within our budget.

In the coming days, I will be working with the files and coming up with a list of the metadata I need.

I will need from you two answers to the questions below. Once I know where things stand, I will send emails to the scientists to get the information I need from them.

Adrian, if there is metadata about the individual documents that you would like included, please send that to me. In particular, a one or two sentence description of each document and a list of keywords would be good.

Doug

**# Tasks #**

**## Raw data (DNGs) ##**

- \* create checksums of raw files
- \* confirm integrity of files on secondary drive
- \* rename files
- \* reconfirm integrity of files
- \* create distribution drives of corrected originals

**## Delivery data ##**

- \* get 8-bit TIFFs
- \* rename BCB's color images
- \* get Roger's PCAs
- \* rename Roger's PCAs

**## Database ##**

- \* update database records with new file base names  
(Essentially the directory name changes from Keith's updates)
- \* write SQL script to do updates
- \* upload raw file names and metadata
- \* collect metadata for data set:
  - \* shot sets -- the correspondence between shot serial numbers and wavelengths (may be the same as Syriac)
  - \* light sources -- confirm same as Syriac
  - \* people -- need a list of project participants
  - \* licenses -- What is the Livingstone data license?
  - \* image set up information - see below
- \* Get processing descriptions
  - \* Color
  - \* PCA
  - \* Pack 8
  - \* Pseudo color
  - \* Sharpie
  - \* Color sharpie
  - \* New Livingstone techniques?

**#### Image setup information ####**

We have little to no image setup information. This is a real problem.  
Below is the information we normally collect.

- \* Subject to focal plane height
- \* Floor to left panel height
- \* Floor to left raking panel height
- \* Floor to right panel height
- \* Floor to right raking panel height
- \* Camera incidence angle
- \* Floor to subject height
- \* Lens make and model
- \* Lens focal length
- \* Illumination filters
- \* Acquisition filters
- \* Left panel to focal axis distance
- \* Left raking panel to focal axis distance
- \* Right panel to focal axis distance
- \* Right raking panel to focal axis distance
- \* Subject top azimuth
- \* Left panel azimuth
- \* Right panel azimuth
- \* X resolution pixels per inch

\* Y resolution pixels per inch

Some of these values we know or can get, like lens focal length, the filters used, the orientation of the lights w/r/t/ to the object, and the pixel resolutions. Others we simply do not have.

It seems we have no choice but to leave this information blank. Agreed?

**## Build metadata ##**

- \* Upload processed images
- \* Associate TIFF data with parent DNGs
- \* Build ImageDescription metadata
- \* Build XMP metadata

These steps are required for each image type for each of seven documents. For Syriac, one document, the number of image types was 18, which is rather large, with experimental images and two types of PCA (single channel and RGB) for each of five different creators of these images, each given their own Creator credit. It required 18 processing batches.

With Livingstone, there are seven DLC documents, each having its own metadata. Up to now each metadata combination of document + image type has required separately handling. This will mean, I'm guessing, at least 49 processing batches.

The processing may be complicated by the fact that the filenames have changed. I'll try to forestall this by verifying that the processed file names match the new corrected names. I doubt this will be a problem with Keith's images, but it will take some tweaking for Bill's and Roger's sets

**## Assemble data set ##**

- \* Update TIFF ImageDescription tags
- \* Create TIFF MD5s
- \* Add ReadMe and supporting documentation

--

Doug Emery  
Emery IT