Manuscript Digitization: Overall Procedural Outline

The Hill Museum & Manuscript Library (HMML) has developed an efficient, cost-effective system to digitize manuscripts and rare printed books. HMML began its digital initiatives in 2003 with a single project operating in Lebanon. As of this writing, HMML maintains over a dozen separate manuscript preservation projects and intends to expand its activities.

During this time, various "how-to" documents were created to provide training reference materials for people working on these projects. This document is an attempt to provide an overview of the entire digitization process with references to the documents already written that address specific aspects of the process.

These documents are available on the World Wide Web at the URL:

http://www.vhmml.us/Resource/Downloads

HMML Digitization Process Overview:

Manuscript preservation is a photographic process; the technician photographs the pages of the manuscripts using a digital single-lens-reflex (dSLR) camera. The image files go directly into a personal computer rather than to the camera's internal memory card. Once an entire manuscript is photographed, the technician creates another folder with the proper name and moves on to the next one.

At some point in the process, the manuscript image folders are copied to an external hard disk drive attached to the PC computer. This hard disk drive is then moved from the "photography" computer to the second computer, where images can be sorted, renamed, rotated into proper orientation, etc. If a second technician can work on this PC, the camera operator can continue to photograph manuscripts without delays. In most cases, two external hard disk drives are provided to allow the two computers to swap the drives as needed.

In some cases, once the image files are in order, DVD discs are recorded for the holding library. In other cases, images are copied to external hard drives kept by the holding library. Another set of completed images is then copied to an external hard drive to be shipped to HMML when they are appropriately filled up.

The processes outlined here are generalized; the specific documents referenced will have the detailed information needed. In broad strokes, here are the important points:

Make Sure the Preliminary Work is Done

The Hill Museum & Manuscript Library acts as a partner to the institution holding the actual manuscript materials. As such, written agreements are created and signed before any manuscript digitization work commences. Before beginning, the technicians and scholars need to have a firm idea of the scope of the preservation work and what materials are to be digitized. For the photographic technicians at a given project site, this generally will have been determined before the technical work is

started.

Prepare the Workspace

Download and read this document:

HMML Digitization Studio Setup

A working environment needs to be prepared for the digitization work. In most cases, this workspace will be in place for months or perhaps years, so strive for a semi-permanent setup. Adequate electrical power supplies are needed—two PC computers will be running, as well as a set of studio flash lighting units. If possible, it's best to have the computers plugged into an uninterruptable power supply (UPS) unit. This type of unit will provide battery power to run the computers for a short while in the event of an electrical power supply failure. This gives the user time to suspend the work activity and properly shut down the computer systems without catastrophic crashes.

An area measuring about 3x2.5 meters in size will be needed for the photographic setup and the PC computer attached to the camera. The camera stand itself needs to be mounted on a very sturdy table (clamps are provided to allow the stand to be quickly attached without extra tools). A second table for the computer system is also needed in most cases.

In addition to this, it's usually good to have some additional space for a third table for the second computer system. Manuscripts awaiting photography can be sorted and queued up on this table. This second computer will used for metadata entry, image file handling, and recording the information to various media.

Prepare the Supplies, Tools, Etc.

HMML provides the tools needed for manuscript photography when a studio is put together. A book cradle system is provided, along with grayscale cards, plastic hold-down tabs, dusting brushes, tweezers, ruler, etc. In addition, there are some downloads the user should get:

arrows.pdf

metadata slate.pdf

The documents can be printed out on any laser or inkjet printer. The "arrows.pdf" documents has a bunch of arrow symbols that can be cut into individual arrows that are placed next to the manuscript photographed to indicate which way is up. The metadata slate document is used to record manuscript metadata if no electronic cataloging method is being used.

PC Computer Configuration

As of this writing, the software used to control the digital camera has been installed on computers running Microsoft Windows versions 7, 8.1 and 10. HMML generally prepares the computers before they are deployed or sets them up properly when a project is started. Users shouldn't have to install software or reconfigure the computers once they are in place. Computers are supplied with the following software:

- Microsoft Windows (operating system)
- LibreOffice (free office application suite)
- Camera Control Software for Digital Camera (Canon or Nikon)
- Image Browser (Canon or Nikon)
- Bulk Rename Utility (for renaming masses of image files)
- DVD Recording Software (to create DVD disks)
- Adobe Reader (for viewing PDF files)
- Internet Explorer or other web browser (for working with online timesheets)

There are often some other applications installed on the computers set up by HMML, but those listed above are the ones needed for HMML's work.

Some studios will choose to connect the two computers using Ethernet cabling to allow data to be transferred between them. Computers may also be connected to the Internet—if this is done, however, it is advised that the computers be equipped with anti-virus software.

Prepare Manuscripts for Photography

Generally, library specialists will gather the manuscripts to be photographed and bring them to the studio setup. Before they are photographed, data about the items needs to be entered into the second PC computer (the one not used for photography). HMML needs to have basic metadata about a manuscript in order to have something to put into its master manuscript database.

HMML has created data-entry tools for manuscripts:

http://www.vhmml.us/wtorborg/assets/Metadata.zip

This is a simple Microsoft Access database that allows the user to enter manuscript information into a form, creating a "flat-file" database of manuscripts. It has provisions for printing out a metadata sheet that can be printed out and photographed as part of the manuscript photography.

Once data about a manuscript is entered in one of these two systems and a metadata sheet is either printed or filled out by hand, the actual photography of the manuscript can take place.

Getting Ready to Photograph

To understand the rationale behind HMML's manuscript digitization methods, download and read:

Rationale.pdf

This document outlines the basic reasoning behind HMML's choice of digitization methods, covering cameras, lighting systems, book holding devices, etc. Some of the camera information is a bit outdated, but the reasoning is the same.

It's also essential to read and understand the document:

2016 Foliation and Filenames

This explains the methods and software used to create image files with the proper filenames—this will in turn allow a folder of images to be sorted by name, keeping the pages in the proper order.

Learn to Use the Digital Camera

A variety of digital single-lens-reflex cameras are suitable for manuscript photography. The following camera models are currently in use by HMML:

Canon 5D Mark II
Canon 6D
Nikon D810

Camera expertise is helpful but not absolutely necessary for technicians employed by HMML. User guides for these cameras can be found on the download page (address given at the beginning of this document).

Guides have been written for specific camera models explaining how to set them up for manuscript photography. In addition, the official camera manuals from Nikon and Canon are available here. Once you know which camera model will be used, it's best to read the guides and familiarize yourself with the camera. Currently, HMML is using the Canon 6D for its new projects.

Decide on a Project Prefix Code

The projects partnering with HMML are given a simple alphabetical code agreed upon by HMML and the holding institution. This is an important thing; this code is used in a number of ways:

-It's used as the filename prefix for naming the image files
-It is part of the HMML "source" number for a particular manuscript
-It's used to label folders to identify specific manuscripts

Some example of HMML project codes:

Project Code	**Code Translation**	**Location of Project**
OLM	Order of Lebanese Maronites	USEK, Kaslik, Lebanon
CFMM	Church of the Forty Martyrs, Mardin	Mardin, Turkey
MLRI	Muzeul Literaturii Române Iaşi	Iaşi, Romania

Each individual manuscript is assigned a "source number" based on this prefix. Examples:

OLM 00256
CFMM 00015
MLRI 00006

The Prefix is written in capital letters, followed by a space. The number following the prefix is always written out with five "placeholders," allowing up to 99,999 manuscripts to be identified for a specific collection. This is the information entered in the "source" field for HMML's metadata systems.

A particular file within a manuscript folder might have a name such as:

OLM_00357_025.JPG

This file is from the OLM project, number 357. The image is of a double-page spread. It is JPEG image file.

Here's an example of a filename from a manuscript photographed as individual recto and verso pages:

CCM_00025_015r.jpg

This file is from the CCM project, number 25. The image is of a single page, 15 recto. It is a JPEG image file.

Things that cannot be part of a HMML source number.

The source number is used as part of a database record for a number of interrelated systems at HMML. Programming languages and web scripting are used for querying and linking these records. Thus, there cannot be any alphabetic characters in the HMML source number that will cause the programming code to fail. Among these are:

( ) * & @ \ / < > | ^ $

Generally, folks tend to create a source number that matches the actual *shelfmark* of the manuscript in question. This is a good thing to try to maintain, but if shelfmarks contain any of these sorts of alphabetic characters, problems will arise. Work around this problem so that the HMML source number won't contain "illegal" characters. An example:

Shelfmark of Manuscript	**HMML Source Number Assigned**
SCAA 005(16)	SCAA 00005 16

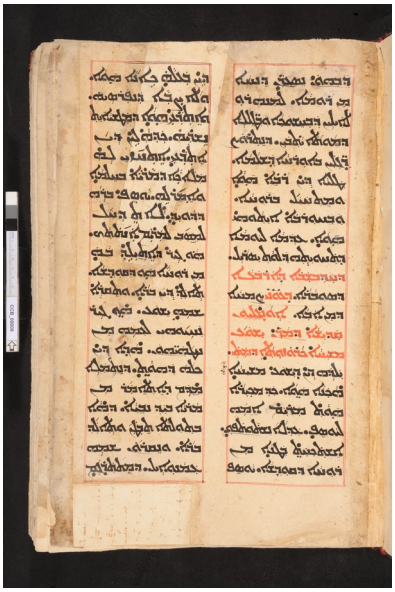
In this case, the parenthesis characters used in the actual shelfmark of the manuscript caused HMML's web programming to fail on these records. Simply replacing them with spaces fixed the problem.

Consult with HMML if further information is needed on creating viable HMML source numbers.

You'll see that in the case of naming individual files, an underscore character (_) is often used instead of a space. For the filenames, this is perfectly fine. For the actual database source entry, the form is to use the project code, a space, and then the five-digit number.

Photograph the Manuscripts

Other instructional documents contain the information for manuscript photography. Rather than repeat this, it might help to see a couple of good examples of the final result.



Manuscripts can be photographed as single pages or two-page spreads, depending on whether the book can easily lie flat without straining the binding. Some aspects of these photographs:

- Images are nicely framed. The book page takes up most of the available frame space without being cut off. Pixels are not wasted on empty space.
- Non-distracting background. HMML supplies a black fabric base as part of the book cradle system. This provides an attractive background for the book.
- The grayscale indicates proper color balance and exposure. The gray patch in the center of the scale is within tolerances specified in other “how to” documents. The white patch doesn't exceed 240 on the RGB color code scale.
- Both photographs have an “up” arrow positioned properly to allow people to know which way is up even if they are looking at the images at a reduced size or are unfamiliar with the script used in the manuscript.
- A small piece of paper with the HMML source number (described earlier) is included in the image. This way, if an image file is separated from the others in a given folder (or has its name changed), it is still possible to figure out where it came from.
- Clear plastic tabs are used to hold down the manuscript pages in the image on the right. The tabs allow the viewer to see the page contents underneath. They are unobtrusive and easy to retouch out of the image if need be.
- In the image at the right, white paper has been placed behind the pages that have holes in them, preventing confusion when reading the text.

One important note: Many books can and should be photographed as two-page spreads. This saves time and makes for less work in post-production. One thing for the camera operator to keep in mind is that if he or she photographs a book oriented “right side up” as they look at it while photographing it, the resulting images will actually be upside down when they are viewed from the computer! This is because the camera is usually oriented the opposite of how the technician views the book on the photography stage. The solution is to photograph two-page spreads “upside down;” this will produce images on disk that read correctly and won't require rotation afterward.

Foliate, Rotate and Rename Files,

Record Images to Disk

In general, the technician operating the camera spends his or her time photographing pages in accordance with the instructional documents mentioned. When a number of books have been imaged, the images are usually transferred to the second PC computer for these further operations. This often is accomplished by simply transferring an external hard disk drive from one computer to the other. This is faster than sending the files over a network, and frees up the camera computer for more photography.

On the second computer, a number of actions are performed on the image files. Among them:

- RAW and JPEG images are separated and placed in different folders.
- RAW and JPEG files are checked and possibly renamed so their filenames give an indication of their position in the book. This way, the files will list out in the proper order.
- Some holding libraries record the data to DVD discs. Others store their data on hard drives. From here, the holding institution can decide whether to also copy the files to a central server, or back them up in any other way.
- Images are copied to an external hard disk drive to send to HMML.

To understand these activities, refer to the document:

2016 Foliation and Filenames

Pay particular attention to the folder structure HMML specifies for storing the RAW and JPEG images created. HMML will be copying the JPEG images to a large server array when the data arrives here; the structure outlined makes this possible.

Send Finished Images to HMML

HMML is now using large-capacity external disk drives as the means for projects to send the finished manuscript images to HMML. Before sending images, take the time to make sure of the following:

- Check that the folder structure on the hard disk is in line with HMML recommendations.
- The image files sort by name in the proper order (representative of the structure of the manuscript itself).
- Each manuscript folder has an image file of the metadata sheet, along with photos of the spine, covers, and edges (the “additional” pictures, as described in the foliation and filenames document), unless the manuscript itself doesn’t have these things.
- Metadata exists for the manuscripts, either in the form of a handwritten form, the simple Microsoft Access database or in a more comprehensive system.