*v*HMML Data Management Plan

Data Types and Backup Procedures

The core of HMML's data preservation plan is to have backup copies of critical data in multiple formats and in multiple locations. Since the *v*HMML system is hosted by CSB/SJU IT Services (CSB/SJU=College of Saint Benedict/Saint John's University), HMML relies upon their workflows and processes for data loss prevention and restoration, as well as for system software backup and restoration. The *v*HMML system is tiered on multiple virtual machines (VMs) with the MySQL database and the image data as separate VMs. In addition, there are separate tiers for a *v*HMML Test platform as well as for the Production platform that is exposed to the public. Since the Test platform is regularly deleted and re-created, sometimes with bogus data for testing purposes, the data management policy outlined here deals only with the Production platform.

Two main types of electronic data are an intrinsic part of the *v*HMML system: (1) alphanumeric data such as catalog records and user data (registration details and access privileges), and (2) image data and associated image transfer metadata (IIIF JSON manifests). The first is stored in tables in a relational MySQL database; the latter is stored on a VM in a Linux directory structure organized by HMML sub-collection.

Data Type 1: MySQL Database

Backup Procedures

Currently, the *v*HMML MySQL database is backed up in four ways: 1) nightly through a MySQL dump to disk in distinct directories on the VM labeled with the date, as well as an off-campus dump to the data center at our partner institution, the College of Saint Benedict; 2) nightly taped incremental backup kept in the Saint John's University IT Services office; 3) weekly taped full backup (each Tuesday) kept in the Saint John's University IT Services office; 4) monthly taped full backup kept in a fireproof safe in another building on the Saint John's University campus. HMML plans to add a copy of the MySQL database to the backup tapes of our digital image collections sent periodically to off-site storage in Utah (see below). Every two hours throughout the workday, the MySQL database is dumped as a file on the VM, so that at most two hours of user and catalog data could be lost in case of a system failure. These bi-hourly data dumps are overwritten on a daily basis since they are redundant with the nightly taped backups.

Sharing and Dissemination

All catalog metadata is publically exposed in *v*HMML Reading Room and can be viewed without registration. This metadata is available for reuse according to a Creative Commons license (CC BY 4.0). Currently, a *v*HMML Administrator can export catalog metadata as JSON or comma delimited CSV files for specific collections or for the entire database. *v*HMML 3.0 would add tools for export and harvesting of metadata in encoded format, as requested by project partners. Sharing of user data is done under the limited conditions outlined in the Privacy Policy: https://www.vhmml.org/privacy.

Intellectual Property

All metadata creators agree in writing to make their work available according to the CC BY 4.0 license.

Data Type 2: Digital Images

Image Formats and Backups

The cameras used by HMML produce two types of image files simultaneously—a RAW digital image and a JPEG image of extremely high quality. Both types of files are retained after capture. A complete set of both RAW and JPEG images is recorded to a hard disk drive for the use of and retention by the owning library. Another hard drive is prepared for shipment to HMML.

Once the hard drive arrives at HMML, all of the JPEG images are copied to an on-site file server operating on an internal local area network (LAN). This storage area network (SAN) system makes use of redundant hard disk arrays to safeguard the data. The JPEGs on the SAN are backed up by CSB/SJU IT Services using LTO-5 tape cartridges stored in a vault on campus. The RAW images are copied to LTO-5 backup tapes that are shipped to a secure storage facility in Utah (Perpetual Storage: http://perpetualstorage.com/). The original hard drive is retained and stored in the secure, climate-controlled, and fire-protected HMML microfilm vault.

An additional set of JPEG images is produced for use in *v*HMML Reading Room. These high-resolution JPEG images, slightly more compressed than the "maximum quality" JPEG images produced by HMML's cameras, are far more usable for internet delivery but still have excellent viewing qualities. Images are rotated for the proper display orientation and JSON manifests are generated to allow them to be viewed in a IIIF environment. After preparation at HMML, these files are copied via secure FTP to the *v*HMML Reading Room Production image server. The original set of derivative images is retained on hard drives kept at HMML. All of the files on the Production server are backed up on tape by CSB/SJU IT Services as described above for the MySQL database. These images will also be taped for shipment to Perpetual Storage in Utah.

Throughout the process of making and storing backups, careful recordkeeping ensures that the status and location of all digital assets is known at all times. These records are likewise backed up and stored in redundant copies.

Sharing and Dissemination

Image files are being made available in *v*HMML Reading Room. Viewing images for most of the collections requires a one-time, no-cost registration that allows HMML to monitor usage and address violations of the terms of access.

Period of Data Retention/ Future Compatibility

HMML is a permanent institution with an endowment sufficient to guarantee ongoing operations at a level that would ensure continuing availability of data. The images and metadata in *v*HMML Reading Room will be available permanently even if in the future they are migrated to new systems. HMML staff are in continuous communication with CSB/SJU IT Services to ensure that hardware and software are available to retrieve data as needed. As storage formats or media used by HMML become deprecated or superseded, data will be copied to the preferred newer system, while in some cases retaining the data in older formats or media.