

Giving Voice to Ancient Texts: Manuscript Scholarship in the Digital Era

COLUMBA STEWART

Hill Museum & Manuscript Library, Saint John's University, Colledgeville, Minn.; e-mail: cstewart@hmml.org

doi:[10.1017/S002074381700099X](https://doi.org/10.1017/S002074381700099X)

The study of manuscripts was traditionally the preserve of scholar-curators in research libraries who devoted their lives to the exhaustive study and description of collections that were often gathered from several sources. The expected result of their solitary labor was a printed catalog that might describe at most a few hundred manuscripts from a particular linguistic or religious culture. That model has been challenged by the advent of large-scale digital projects that aggregate thousands or even tens of thousands of manuscripts from multiple libraries in a single database or portal.

This contribution is based on my experience with such a project at the Hill Museum & Manuscript Library (HMML) at Saint John's University in Colledgeville, Minnesota. I have been Executive Director since 2003, leading a move from the largely Euro-centric work of HMML's first decades to a new focus on manuscript cultures in the Middle East, Africa, and India, encompassing both Christian and Islamic traditions. HMML had already undertaken an extraordinary microfilming project in Ethiopia in the 1970s (the Ethiopian Manuscript Microfilm Library [EMML]). While the results were groundbreaking for Ethiopian studies, the 8,000 manuscripts were less than 10 percent of HMML's microfilm holdings. The current total of microfilmed and digital manuscripts approaches 200,000 items. Most are codices, with the remainder being archival materials. Thousands more digitized manuscripts are being added every year from digitization projects across the world.

In the microfilm era, HMML published catalogs of many collections, most significantly the ten volumes for the EMML. Some were simply inventories or handlists. Others, like the many volumes of the EMML catalog, provided detailed, text-level descriptions. Copies of microfilms were made upon request, typically requiring several weeks to process because of the need to seek permission from the owning institution and then to send the film to a processor for duplication. As the digital era dawned in the 1990s, HMML was part of a consortium funded by the Andrew W. Mellon Foundation to develop standards for encoding, storing, and presenting catalog descriptions of medieval and Renaissance manuscripts in electronic form. The project, Electronic Access to Medieval Manuscripts (EAMMS), created a complex metadata scheme consisting of nested tables in a relational database to represent: the complete manuscript object; its parts (if the object consisted of more than one text block bound together into a single object); and individual texts. Because the metadata scheme was designed by scholars of premodern western manuscripts, it was inevitably oriented toward the manuscript types familiar to them. The EAMMS scheme would become the basis for the Digital Scriptorium¹ initiative and the European Manuscript Access through Standards for Electronic Records (MASTER) project.² HMML's implementation of EAMMS went online in 1999 and continues to exist as a legacy catalog while its contents are being

conformed to new standards for inclusion in the *Virtual* HMML (vHMML) Reading Room (described below).

HMML's 2003 initiative in the Middle East introduced color digital imaging and a new workflow. Photographic work at the many field sites is done by local teams trained and paid by HMML. Their equipment is provided by HMML. Regional HMML field directors provide technical support and manage the flow of data from each site to Minnesota, where images and metadata are archived both on- and off-site. In the early digital years, HMML continued to catalog manuscripts using the EAMMS scheme with some adjustments to support description of nonwestern manuscripts. By 2009 the inadequacy of both the scheme and the underlying technology had become clear. An obvious need was stricter authority control for toponyms, institutional and personal names, languages, scripts, and other features of manuscripts to provide a more consistent search experience. New funding from the Andrew W. Mellon Foundation supported some adjustments to the cataloging procedures and the hiring of more catalogers. An adjunct online system allowed faster, Google-like, searching of the EAMMS-based catalog. For born-digital objects the online catalog now included links to sample images delivered via HMML's implementation of ContentDM, known as *Vivarium*. Catalogers at HMML used an MS Access form to add records to the database. External catalogers entered metadata using templates created in MS Excel. These were then reviewed by HMML curators and exported to a format that could be aggregated to the database. Manuscript images were provided to scholars on CDs upon payment of a service charge or, starting in 2012, free of charge via password-protected galleries in *Vivarium*.

In the early 2010s, HMML began to develop other online tools to support manuscript studies across a wide range of languages and cultures, with an initial focus on students and scholars just starting to use manuscripts in their research. The resulting vHMML platform was launched in 2015 with paleography lessons in Latin and Syriac, densely annotated sample images (an online paleographical album), and reference tools.³ It drew on the pedagogical expertise of HMML staff and outside consultants to complement, or in many cases, substitute for, classroom instruction on manuscript studies.

As the original vHMML platform was in development, HMML staff were at work on the next phase of the project, vHMML Reading Room, which would provide access to complete image sets of all digitized manuscripts (both born-digital and scanned microfilm) in HMML's collections. A grant from the Henry Luce Foundation supported planning and software development for Reading Room. The issue of authority control for names and features resurfaced with new urgency. Besides improving the search experience for users, rigorous consistency would make data more easily shareable with (and usable by) other Digital Humanities projects. Despite good intentions, HMML's previous efforts at enforcing strict metadata rules had proven inadequate, partly because of flaws in the metadata scheme. A thoroughly revised metadata scheme would be needed for vHMML Reading Room. Devising it required both traditional scholarly expertise and familiarity with developments in the Digital Humanities. The latter was not viewed as a mere adjunct to subject knowledge or as an auxiliary technology, but rather as the form of much present-day study of the humanities.

The new metadata scheme for vHMML Reading Room fully accommodates non-western manuscripts, and additionally provides alternative schemes for printed objects, hybrid manuscript-printed objects, and archival materials. Strict application of LC and VIAF authority control for toponyms, institutional names, personal names, and titles (where they exist) is presumed throughout.⁴ HMML's curators developed controlled vocabularies for languages, scripts, and various descriptive fields according to current standards. An online cataloging interface was created to support metadata entry by both HMML staff and external catalogers, with a range of permission levels designed to provide quality control for cataloging submitted by off-site catalogers. vHMML Reading Room uses the MIRADOR image viewer, a platform compliant with the International Image Interoperability Framework (IIIF), and chosen for its superior viewing experience, potential for sharing with other Digital Humanities projects, and robust user community.⁵

vHMML Reading Room was launched in August 2016 as part of vHMML 2.0. All metadata is freely searchable by anyone visiting the site. Viewing of complete image sets generally requires free, one-time registration. The implementation of Elasticsearch⁶ technology provides greatly improved searching and faceted search options. Each object now has a permalink for reference purposes. vHMML Reading Room does not yet support downloading of images, though for most collections this would not be available anyway because the owners would not permit it.

Imposing strict control of names and descriptions of features for objects in vHMML Reading Room unavoidably slows the process of metadata creation. To mitigate this problem, some collections are being initially uploaded to vHMML Reading Room with minimal metadata sets to make them available without delay. This serves scholars who know which manuscript they need (on the basis of previous catalogs or scholarly references) or who wish to browse uncataloged, or even unknown, collections. Such "stub" records contain basic and stable identifiers (city, repository, shelfmark, HMML project number) as well as the permalink. Over time the records can be enriched with additional metadata conformed to the standards of vHMML Reading Room.

vHMML currently supports data sharing with other projects via exports of the entire catalog database in JSON format and of the permalinks for all objects in CSV format. vHMML 3.0, currently in development with funding from the National Endowment for the Humanities, will allow export of metadata in encoded form (e.g., METS, OAI, and EAD). The new development phase includes exploration of technologies supportive of linked data. The Mirador viewing environment supports easy invoking of images from other projects via IIIF manifest URLs generated for each object. The IIIF manifest URLs are not generally exposed in vHMML Reading Room because of restrictions imposed by the owners of the manuscripts. They do appear in the metadata display for items from HMML's own collections and for scans of the EMMML microfilms. As our partners grow more confident about the benefits of image sharing, they may choose to allow exposure of the IIIF manifest URLs. Upon request HMML can supply trusted projects with the IIIF manifest URLs of currently restricted manuscripts or collections to permit viewing and annotation in other IIIF environments.

A major Digital Humanities project like vHMML offers several kinds of intellectual labor, whether for HMML staff, external catalogers, or the growing number of partner projects using vHMML Reading Room as the environment for cataloging and

displaying their digital objects. There is the obvious task of describing the manuscripts, which requires a different approach from traditional cataloging oriented toward creation of printed reference catalogs. In the online environment, emphasis is placed on discoverability, both of texts and of the material aspects of manuscripts. The latter is especially important for those interested in the production and circulation of manuscripts. The presence of several hundred manuscript libraries in vHMML Reading Room permits analysis of collections both as distinct entities and as components of larger data sets. Digitization of several libraries belonging to different communities in the same city or region (such as Aleppo or Jerusalem) allows mapping of intellectual microclimates and the exchanges between them.

Although the metadata fields in vHMML Reading Room allow inclusion of highly detailed descriptions, limited resources and the urgency of making as many manuscripts available as quickly as possible mean that, for the present, such traditional exhaustive cataloging is reserved for highly important manuscripts. Given HMML's priorities of digital preservation and access, development of richer descriptions will perhaps best be done by partner projects focused on particular manuscript cultures or textual traditions, or through contributions by individual scholars willing to share their findings with the user community. vHMML 2.0 provides basic support for user-contributed corrections or suggested additional metadata; one of the goals for vHMML 3.0 is an improved suite of tools for contributions and corrections of metadata.

One of the major imperatives for vHMML and similar projects is awareness of evolving expectations for projects in the Digital Humanities for content (both quantity and quality) and for the ease and richness of the user experience. Project teams must pay close attention to other initiatives while having a clear understanding of the purpose and focus of their own project. Any shifts or expansions of scope must be coherent and sustainable. The range of content in vHMML Reading Room is already vast, with some 23,000 objects in more than forty languages in the database as of September 2017, most of them with full image sets. More are added weekly. For this and other projects, ambitions for further development of the software and improvement of the user experience are necessarily constrained by limitations of human and financial resources. Wise decisions about choosing which features to add when a new development cycle begins require counsel from an external advisory board and regular communication with users. The dependence of most Digital Humanities initiatives on external grants for initial development can obscure the challenge of sustainability, a central concern for projects like HMML's that promise long-term access to digital assets.

In sum, vHMML and its Reading Room call for new kinds of humanists, equally at home in traditional textual study and in the newer forms of scholarship enabled by direct access to tens of thousands of manuscripts representing many cultures. This is surely intellectual labor of a high order.

NOTES

¹Digital Scriptorium, accessed 29 September 2017, <http://bancroft.berkeley.edu/digitalscriptorium/>.

²Manuscript Access through Standards for Electronic Records, accessed 29 September 2017, <http://master.dmu.ac.uk/>.

³Paleography lessons for Arabic and Armenian manuscripts will be released in 2018 as part of vHMML 3.0.

⁴HMML is contributing new names and titles to the Library of Congress to build up the nonwestern authority records.

⁵For Mirador, see <http://projectmirador.org/>, accessed 29 September 2017, and on IIIF, see <http://iiif.io/>, accessed 29 September 2017.

⁶For more information on Elastic Search, see <https://www.elastic.co/products/elasticsearch>, accessed 29 September 2017.