https://doi.org/10.2352/issn.2168-3204.2019.1.0.21

Bibliotheca Philadelphiensis: Collaborative digitization and data management

Michael Overgard; University of Pennsylvania; Philadelphia, PA/USA
Anna Levine; University of Pennsylvania; Philadelphia, PA/USA

Abstract

In 2016, the University of Pennsylvania Libraries, along with fourteen partnering institutions in the Philadelphia area, was awarded a grant from the Digitizing Hidden Special Collections and Archives initiative of the Council on Library and Information Resources (CLIR) to produce the United States' largest regional online collection of medieval manuscripts. For the Bibliotheca Philadelphiensis *project, otherwise known as BiblioPhilly, partnering institutions had thirty months to digitize more than 160,000 pages from 450 European medieval and early modern manuscripts. According to the terms of the grant, the digitized manuscripts had to be made available in the public domain via a searchable digital interface, be easily downloadable at high resolution, and accompanied by both expertly compiled descriptive metadata and unique physical collation models that help researchers to date manuscripts, understand how codices were disassembled and reconstructed in different periods, recombine fragments, and much more. While the BiblioPhilly project required intensive data capture from photographers and catalogers along with specific, time-sensitive, and particularly careful handling conditions, the process functioned smoothly through project management and a cooperative spirit among colleagues. As a result of these efforts, researchers may now creatively interact with the materiality of a manuscript in a digital environment in a manner that would be impossible with the physical manuscript itself. In a manner new to the field, BiblioPhilly enables researchers to become not just assessors of, but participants in, a long history of manuscript repurposing, reconstruction, and transformation.*

Introduction

In 2016, the University of Pennsylvania Libraries, along with fourteen partnering institutions in the Philadelphia area, was awarded a grant from the Digitizing Hidden Special Collections and Archives initiative of the Council on Library and Information Resources (CLIR) to produce the United States' largest regional online collection of medieval manuscripts. For the *Bibliotheca Philadelphiensis* project, otherwise known as BiblioPhilly, partnering institutions had thirty months to digitize more than 160,000 pages from 450 European medieval and early modern (through 1599 CE) manuscripts. The vast majority of the digitization (132,000 pages from 358 manuscripts) was to be completed in a single photography studio on a single camera station in the Penn Libraries' Schoenberg Center for Electronic Text and Image (SCETI). The SCETI photography studio was also to host four content specialists from Penn's Kislak Center for Special Collections, who would create and verify all of the project's metadata. For BiblioPhilly, *metadata* means not only standard descriptive metadata (e.g., date, author, subject) and structural metadata (e.g., 1 recto, 1 verso, 2 recto, 2 verso, and so on), but also a host of codicological and paleographic characteristics that allow users of our project's website to easily discover manuscript features from notable bindings and palimpsests to historiated initials and colophons. Furthermore, content specialists developed physical collation models for all of the manuscripts, which help researchers to date manuscripts, understand how codices were disassembled and reconstructed in different periods, recombine fragments, and much more. As valuable as all of this data is, the recording of this data required catalogers to spend a lot of time with the physical manuscripts during the same brief, one-month period that a photographer and quality assurance (QA) assistant also needed to handle the manuscripts for digitization.

Challenges

The twelve institutions that sent manuscripts to SCETI typically delivered batches of ten-to-twenty manuscripts for a thirty-day loan period. The batch quantities did not exceed this both because of insurance policy limits and a desire to keep at least some manuscripts available to local users at owner institutions. To meet our more ambitious first-year production goals, we needed to process twenty manuscripts every thirty days, while preventing a photographer, QA assistant, and cataloger from reaching for the same manuscript at the same time. Digitization needed to be completed with 100% accuracy, because our production schedule would not tolerate the return of manuscripts for correction. Finally, all materials had to be handled safely, particularly through the digitization process. The challenge here was less a matter of adjusting to a gentler photographic setup, and more a matter of demonstrating the safety of our imaging process for project partners, making occasional accommodations in consultation with them, and leaving our partners with the final say in what would or would not be digitized.

Project Plan and Execution

Preparation

Before any manuscripts came to SCETI, staff from partner institutions would prepare their materials by creating spreadsheets with structural metadata. They would also record as much descriptive data as they were able. This preparatory work meant that when manuscripts did arrive to SCETI, there was a structural metadata record that indicated to the photographer how she should treat peculiar features of a particular codex (e.g. inserts, stubs, clasps), and a record to help the QA assistant confirm the accuracy of his image review. It also meant that Penn's content specialists could catalog manuscripts more expeditiously, because they needed to verify, edit, and augment existing descriptive metadata more than they needed to create new data from scratch.

Training for this preparatory work took place at two "boot camps" held at the Penn Libraries. As part of these boot camps, project partners visited SCETI. During this time, SCETI's imaging staff members were able to demonstrate their tools and methods for photographing manuscripts, to answer questions about the process, and to address any concerns of our project partners. As a result of the live imaging demonstration, curator anxieties about the stress under which digitization placed manuscripts in general, and worries about putting manuscripts under glass in particular, all but evaporated. What few clouds of concern still lingered will be described in more detail later, along with the physical setup for the photography of the manuscripts and a report of the post-digitization condition of the manuscripts. I will only add here that SCETI's manager visited each partner institution and conducted condition reviews of all manuscripts with local preservation specialists. Risk assessments were created for each manuscript, and the authority was left with the preservation specialist of each manuscripts' home institution to decide whether or not a manuscript would be digitized, and whether a manuscript would be digitized only under certain conditions (e.g., do not photograph under glass, only photograph with a conservator present, etc.).

Production

As manuscripts began to arrive at Penn, Penn's project team established that the first priority for our thirty-day processing window was digitization. If any work still needed to be done on a manuscript after it was returned to its home institution, it would be much easier to send a cataloger there with a laptop to record data than it would be to send a photographer with cumbersome and expensive camera equipment to capture quality images under less-than-advantageous conditions, such as rooms bathed in sunlight, or spaces that lacked sufficient electrical outlets to power equipment.

Once material arrived to SCETI, manuscripts were assigned a shooting order. While the photographer and QA assistant digitized manuscripts one through twenty, the catalogers worked in the opposite direction, processing manuscripts twenty through one. When the two teams met in the middle, somewhere around manuscript ten, priority was given to the digitization team to access the next manuscripts in the queue, and the catalogers moved on to process the manuscripts that had been through digitization in the early part of the month.

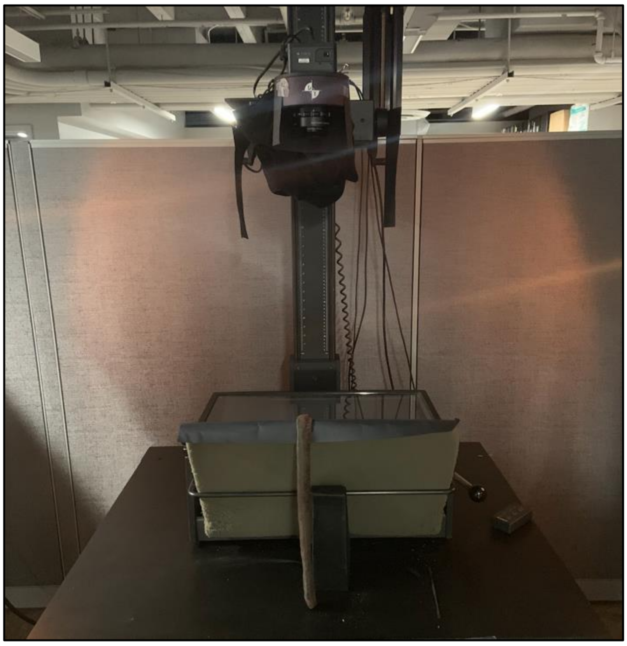


Figure 1. A Linhoff book easel atop a Digital Transitions RG3040 copystand.

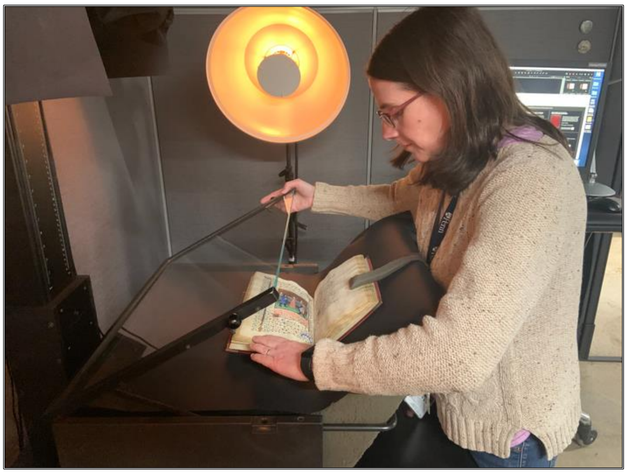


Figure 2. A view of the manuscript as positioned for digitization, as seen with photographer Andrea Nunez.

Our photographic setup for the BiblioPhilly project was the same setup that we regularly use in SCETI for the imaging of manuscripts. While project materials vary, SCETI photographs, on average, 200,000 pages of bound manuscript codices each year, and does so without injury to the manuscripts. Our setup is simple and commonplace, but nicely balances photographic efficiency with manuscript security.

While book easels are designed to place a book, in its entirety, underneath glass with a 180-degree opening, the majority of bound special collections materials we digitize, whether printed books or manuscript codices, cannot safely be opened so wide. As you can see in the image above, we only place one side of a codex under the easel glass at a time, while the other side of the codex is supported with archival book wedges. We capture all rectos in a first pass through the manuscript, turn the object around, and capture all versos in a second pass through the manuscript. We also captured the spine, fore edge, head, and tail shots for all manuscripts in our monthly batch on the first day of shooting, before incorporating the book easel into the digitization process. This routine meant that a QA assistant would not have to review a manuscript twice, first for boards and text block, later for edge shots; it also meant that a photographer would not need to return to a manuscript later in the month for further digitization while a cataloger was processing it. Taking edge shots on the first day after receipt of a new manuscript batch also meant that we provided the manuscripts with extra time to acclimatize to their new environment before we opened them.

An important aspect of the book easel, from a materials preservation perspective, is that glass is not brought down upon the manuscript, laying its full weight on the page. Instead, the operator manually raises the manuscript up to the glass, applying only as much pressure as is needed to smooth the page without compromising the condition of the page or larger codex. During our boot camp demonstrations of how we digitize a manuscript with the aid of a book easel, we allowed project partners to practice placing manuscripts into our setup. After our partners got a literal feel for the arrangement, SCETI was given the OK to digitize manuscripts with the book easel in all but three cases; not three institutional or collection cases, but only three out of the 358 we digitized for this project came with the stipulation that they were not to be shot under glass. In one case, this was due to the brittleness of severely cockled pages. In the other cases, this was because inks in two heavily illustrated manuscripts were determined to be too friable to sustain any pressure. Additionally, thirteen oversized manuscripts were not shot with the use of any glass. So of the 358 manuscripts SCETI digitized, 342 were shot under easel glass, two were shot under the cautious observation of conservators, and these manuscripts produced over 132,000 images. Of these 132,000 opportunities for damage, we suffered only two detached boards throughout the course of the entire project – an impressive feat for such old and delicate objects. In both cases, the condition reviews that took place before the project began had noted that board detachment would be expected with these manuscripts—regardless of what imaging techniques and equipment were employed for capture—and curators from the owner institutions had assumed the risk and authorized digitization.

Of course, our careful handling of materials would be for naught if our output – the digital images themselves – were inaccurate or of poor quality. Our quality assurance process not only required the QA assistant to review each imaging session to verify that our image files met FADGI standards, but required the QA assistant to compare each digital image with each physical folio *and* with the structural metadata record to assure that no part of the manuscript was missed or out of order.

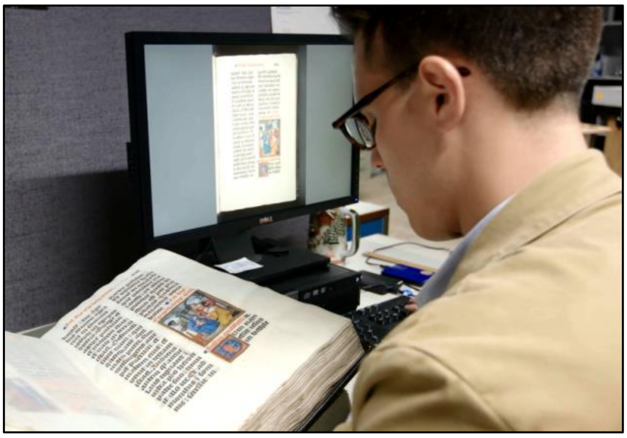


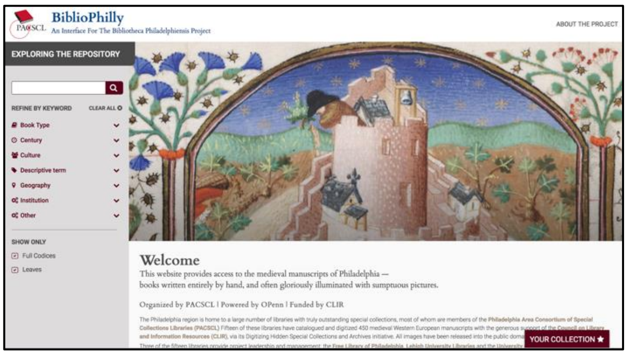
Figure 3. *SCETI staff performing quality assurance, checking a digital facsimile of the Free Library of Philadelphia's Lewis MS E160 against the physical manuscript.*

If there were any discrepancies, the QA assistant would either correct whatever errors existed in the structural metadata or would assign re-shoots to the photographer. Because of the great unlikelihood that a photographer, QA assistant, and structural metadata creator would all happen to overlook the same leaf, so that three people conducting three different processes would produce an error that somehow coincided, we are especially confident that our data is accurate.

When Penn's content specialists began processing a new batch of manuscripts, they started by recording all the data that they could ascertain only by handling the physical manuscript. If the manuscripts were returned to their home institutions before any further data was recorded, the content specialists would be able to complete their cataloging based on the images now captured. Because of this routine, content specialists have not had to make subsequent visits to partner institutions to complete their cataloging efforts except when a content specialist was away for an extended period of time while a batch of manuscripts was on loan to Penn.

Results

While the digitization of a manuscript might transform a fragile object into a better-preserved form, and while it may offer global, web-based access to a text that would otherwise be accessible in only one location, there is also much discussion about what is lost in such a transformation. A significant aspect of a manuscript codex's materiality that is often lost in digital environments is collation, which we have furnished in the form of metadata and a collation model visualization tool on our project's website.



Welcome

This website provides access to the medieval manuscripts of Philadelphia —
books written entirely by hand, and often gloriously illuminated with sumptuous pictures.

Organized by PACSCL | Powered by OPenn | Funded by CLIR

The Philadelphia region is home to a large number of libraries with truly outstanding special collections, most of whom are members of the Philadelphia Area Consortium of Special Collections Libraries (PACSCL) Fifteen of these libraries have catalogued and digitized 450 medieval Western European manuscripts with the generous support of the Council on Library and Information Resources (CLIR), via its Digitizing Hidden Special Collections and Archives initiative. All images have been released into the public doma
Three of the fifteen libraries provide project leadership and management: the Free Library of Philadelphia, Lehigh University Libraries and the University

Figure 4. *A screen capture of the BiblioPhilly digital interface's homepage.*

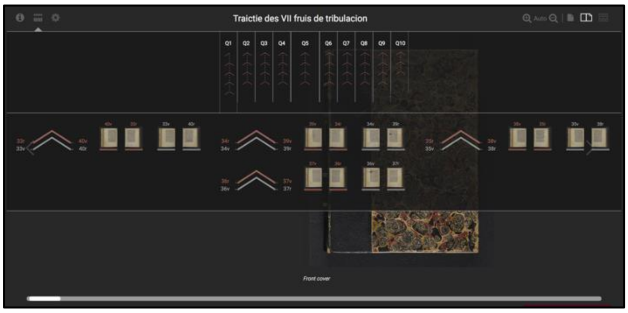
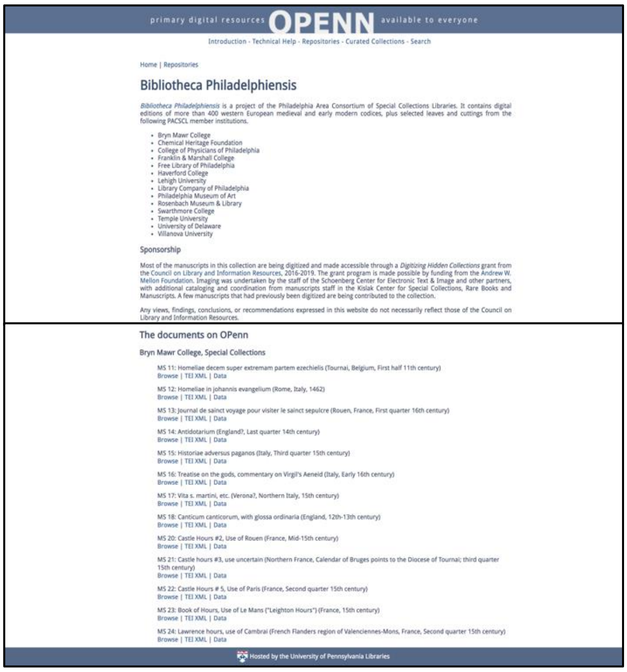


Figure 5. *A screen capture of the BiblioPhilly digital interface's collation model.*

While this information about a codex's materiality may be instructive to a scholar and may raise new research questions, it also makes possible the reuse of the digital manuscript in ways that would not be possible with the physical original. To take one simple example, a 13th-century manuscript that was disassembled and reconstructed in the 15th-century could now be split into both its original and present forms, something that would not be possible with the physical manuscript. In this way, contemporary researchers may become not just critics of but participants in a long history of document transformation. A manuscript is often the product of many hands constructing, inscribing, redacting, reconstructing, illustrating, and annotating. Such a tradition is closed to contemporaries because these activities disfigure rare and precious physical treasures. However, in a digital environment the long tradition of manuscript reconstruction and repurposing gains new life, where inventive and experimental creation need not live in tension with destruction.

To make data usable, it is not enough for them to simply exist on a server: ideally, the data must be accessible to students, scholars, and the public. All of our images (both TIFF master files and JPEG derivatives for web use) are freely available for download by page, by manuscript, or collection under a Creative Commons license for anyone to use. Our metadata is also available for download at the manuscript- or collection-level and is under the same Creative Commons license. Our descriptive, structural, and technical metadata exist in well-organized, machine-readable xml, which means researchers will not need to devote hours to heavy data transformation and remediation to make the data programmatically useful.


primary digital resources OPENN available to everyone
Introduction - Technical Help - Repositories - Curated Collections - Search
Home | Repositories

Bibliotheca Philadelphiensis

Bibliotheca Philadelphiensis is a project of the Philadelphia Area Consortium of Special Collections Libraries. It contains digital editions of more than 400 western European medieval and early modern codices, plus selected leaves and cuttings from the following PACSCL member institutions.

- Bryn Mawr College
- Chemical Heritage Foundation
- College of Physicians of Philadelphia
- Franklin & Marshall College
- Free Library of Philadelphia
- Haverford College
- Lehigh University
- Library Company of Philadelphia
- Philadelphia Museum of Art
- Rosenbach Museum & Library
- Swarthmore College
- Temple University
- University of Delaware
- Villanova University

Sponsorship

Most of the manuscripts in this collection are being digitized and made accessible through a *Digitizing Hidden Collections* grant from the Council on Library and Information Resources, 2016-2019. The grant program is made possible by funding from the Andrew W. Mellon Foundation. Imaging was undertaken by the staff of the Schoenberg Center for Electronic Text & Image and other partners, with additional cataloging and coordination from manuscripts staff in the Kislak Center for Special Collections, Rare Books and Manuscripts. A few manuscripts that had previously been digitized are being contributed to the collection.

Any views, findings, conclusions, or recommendations expressed in this website do not necessarily reflect those of the Council on Library and Information Resources.

The documents on OPenn

Bryn Mawr College, Special Collections

MS 11: Homeliae decem super extremam partem ezechielis (Tournai, Belgium, First half 11th century)
Browse | TEI XML | Data

MS 12: Homeliae in johannis evangelium (Rome, Italy, 1462)
Browse | TEI XML | Data

MS 13: Journal de sainct voyage pour visiter le sainct sepulcre (Rouen, France, First quarter 16th century)
Browse | TEI XML | Data

MS 14: Antidotarium (England?, Last quarter 14th century)
Browse | TEI XML | Data

MS 15: Historiae adversus paganos (Italy, Third quarter 15th century)
Browse | TEI XML | Data

MS 16: Treatise on the gods, commentary on Virgil's Aeneid (Italy, Early 16th century)
Browse | TEI XML | Data

MS 17: Vita s. martini, etc. (Verona?, Northern Italy, 15th century)
Browse | TEI XML | Data

MS 18: Canticum canticorum, with glossa ordinaria (England, 12th-13th century)
Browse | TEI XML | Data

MS 20: Castle Hours #2, Use of Rouen (France, Mid-15th century)
Browse | TEI XML | Data

MS 21: Castle hours #3, use uncertain (Northern France, Calendar of Bruges points to the Diocese of Tournai; third quarter 15th century)
Browse | TEI XML | Data

MS 22: Castle Hours # 5, Use of Paris (France, Second quarter 15th century)
Browse | TEI XML | Data

MS 23: Book of Hours, Use of Le Mans ("Leighton Hours") (France, 15th century)
Browse | TEI XML | Data

MS 24: Lawrence hours, use of Cambrai (French Flanders region of Valenciennes-Mons, France, Second quarter 15th century)
Browse | TEI XML | Data

Figure 6. *The BiblioPhilly data available for download via OPenn, the University of Philadelphia's digital repository of cultural heritage materials.*

While the BiblioPhilly project required intensive data capture from photographers and catalogers, the process functioned smoothly through fundamental project management and a cooperative spirit among colleagues. As a result of these efforts, researchers may now creatively interact with the materiality of a manuscript in a digital environment in a way that would be impossible with the physical manuscript itself. It is our hope that collation models will become a standard part of the data captured for all manuscript digitization projects, and that we may continue to consider not only what of the physical originals may be lost or occluded in digital reformatting but what of their materiality may be uniquely revealed in digital environments.

Author Biography

Michael Overgard received his BA in English from Eastern University (2005) and his MLIS from Rutgers University (2010). Michael has worked as the project manager of several digital projects for academic libraries, and, since 2015, has served as the manager for the Schoenberg Center for Electronic Text & Image at the University of Pennsylvania Libraries.

Anna Tione Levine is a Digital Asset Manager at the Schoenberg Center for Electronic Text & Image at the University of Pennsylvania Libraries. Anna works both as an asset and project manager, working with inter-departmental stakeholders and programmers to ensure the quality and relevance of our work.