Onward to Omeka
A Migration Tale

By Annie Benefiel

INSTITUTIONAL CONTEXT

Grand Valley State University (GVSU) is a mid-sized comprehensive university in West Michigan, serving a student population of around 24,000. Its Special Collections and University Archives department (SCUA) is a part of the larger University Libraries. For many years, SCUA was staffed by one full-time librarian. In 2005, the unit hired another full-time archivist following the acquisition of a large literary archive of Michigan-born author Jim Harrison. Digitization of special collections material began in earnest around the same time, and the library adopted OCLC's [CONTENTdm](#) to host its digital collections online.

Several years later, SCUA agreed to become the repository for GVSU's Veterans History Project (VHP), a local contributor to the project coordinated by the U.S. Library of Congress. GVSU's VHP program director wanted not only to contribute the oral history recordings to the Library of Congress, but also to stream them online for ease of use in his own classroom instruction. At that time, CONTENTdm did not have the capacity to host streaming video or audio content, so SCUA enlisted GVSU's Web Services Librarian to write scripts and stylesheets to pull in the streaming videos hosted on the university's Ensemble media service into the item record pages in CONTENTdm.

In 2014, after nearly a decade of digitizing content, collecting oral histories, and collaborating with other campus partners, SCUA had exceeded its service tier capacity within CONTENTdm. Facing considerable additional cost to move to the next service tier and enable continued growth, GVSU Libraries began reviewing platform alternatives. By this time, an interdepartmental team had formed around the curation of digital collections, including the University Archivist, a new Assistant Archivist (the author of this article), and a Metadata and Digital Curation Librarian. Additional technical assistance and support was provided by the Library Technology Specialist and Web Services Librarian.

PILOTING PRESERVICA

The rapid growth of the libraries' digital collections had also necessitated a digital preservation plan, so the Metadata and Digital Curation Librarian began piloting the [Preservica](#) digital preservation system. In 2014, Preservica was still relatively new but promised a solution to digital preservation and access that could also preserve the original organization and file hierarchies of born-digital archives.

After testing ingest workflows for several months, the team had enough content in Preservica to open the site to public access and begin user experience (UX) testing. We reached out to our campus stakeholders, such as the Veterans History Project director, for reviews, and we conducted usability tests with GVSU students. Our UX testing and review process indicated that Preservica's Univer-

Annie Benefiel is University Archivist and Digital Collections Librarian at Grand Valley State University.

sal Access portal was difficult for users to navigate and didn't provide the same robust metadata searches that users of our CONTENTdm site were accustomed to.

When we provided our data to the Preservica support team, they indicated that improvement of the Universal Access portal was a considerable way down their development roadmap. To prioritize that development, GVSU would need to fund it themselves. Once again constrained by our budget, we began evaluating open-source alternatives for a digital collection access platform.

EVALUATING OTHER OPTIONS

In 2016, the University Archivist retired, and my role was changed from Assistant Archivist to Collection Management Archivist. A working group formed to evaluate digital collections options included myself, the Digital Initiatives Librarian, the Metadata and Digital Curation Librarian, and the Scholarly Communications Outreach Coordinator, who managed the library's [Digital Commons](#)-hosted institutional repository *ScholarWorks@GVSU*. We called ourselves the DOWG (Digital Objects Working Group).

After a review of the digital collection platform landscape, the team settled on a shortlist of options to review. These included our own Digital Commons repository; [CollectiveAccess](#), an open-source option used by GVSU's Art Gallery; [Omeka](#), a well-established open-source option; and [Hydra-in-a-Box](#), a new project in development through a collaboration between the Digital Public Library of America (DPLA), Stanford University, DuraSpace, and the Samvera community (formerly "Hydra project").

Our evaluation criteria included the following requirements:

- Affordable total cost of licensing or operation
- Affordable scaling of cost with growth of collections
- Hosting options or ability to self-host with existing staff and infrastructure
- Customizable search interfaces
- Ability to search or limit search results by metadata facets
- Native streaming of multimedia
- Browsing interfaces usable and intuitive
- Ability to present compound objects, such as a video accompanied by a text transcription, within the same record, preserving the relationship between all associated files
- Ability to view or stream digital objects within the web interface without downloading them
- Facilitates full-text indexing of PDFs and other subordinate files
- Enables OAI-PMH metadata harvesting so that we can contribute to the DPLA or other digital collection aggregators
- Bulk ingest of files and metadata
- Bulk update of metadata by collection
- Auto-generation of item thumbnails

Other criteria that we considered as not required, but nice to have, included:

- Responsive web interface for both computer and mobile displays
- Conforms to WCAG 2.0 web accessibility standards
- Supports embedding media from other sources (e.g. YouTube)
- Supports site-mapping or search engine optimization
- Ability to restrict file downloads
- Ability to require log-in to view certain items
- URLs and file/path identifiers that are logical, hierarchical, transparent, and understandable
- Platform supports virtual collections or exhibits
- Social media support and/or RSS
- User-created portfolios (e.g. "shopping carts") or persistent user-curated collections that they can annotate and share
- User-submitted metadata corrections, validated by site administrator
- Error logging and communication of errors to administrators in clear language

Each working group member independently demoed and reviewed the short-list options based on our established criteria. We then convened to rank and discuss our options. Though none of the options were perfect, the group settled on Omeka as our top choice. With our existing staffing and campus-provided server we could host our own instance of Omeka, and it met almost all our requirements via optional plug-ins. Some of the more sophisticated "nice-to-have" criteria were not available in Omeka, but both administrative and user interfaces were easy to use and fit our existing workflows well. We submitted our pilot proposal and rationale to the library's leadership team and established a pilot timeline.

PILOT AND MIGRATION PROCESS

During the summer of 2016, the Digital Initiatives Librarian worked with campus IT to install a test instance of Omeka on the server. In early fall, the Metadata and Digital Curation Librarian and I loaded test batches of several of our collections into Omeka. Then, at the end of the fall semester, we rolled it out to our campus stakeholders for review and conducted usability testing with students. Reviews and UX testing proved to be mostly favorable, and we received helpful feedback that helped us improve the search functionality.

With these favorable results, library leadership approved full adoption of Omeka, and we began full-scale migration. Because we'd chosen to self-host our platform, we did not have the assistance of a vendor support team to migrate our content. Our migration team consisted of all members of the DOWG, minus the Scholarly Communications Outreach Coordinator. We also added the additional metadata and ingest support of two library metadata specialists.

Our timeline was determined by the library's budget cycle – we would lose access to CONTENT-dm at the end of June 2017, as we were not renewing our subscription for another year – so we needed to have all the highest-use collections published in Omeka by July 1. We evaluated the collections we had to migrate by size, complexity, and importance, and divided them up amongst the

migration team. The most skilled team members each had one or more higher-stakes collections to migrate, and our support team members were assigned smaller, less trafficked collections that had fewer metadata issues.

Digital collection metadata in CONTENTdm followed the Dublin Core standard, though a few collections had been given custom metadata fields. Fortunately, Omeka also allowed custom fields, so these proved not to be a big problem. Digital object master and access files had been maintained outside of CONTENTdm on library network storage and were for the most part well-organized. Our migration process consisted of the following steps.

First, we exported our Dublin Core metadata from CONTENTdm in CSV (comma-separated-values) text files. Next, the metadata files were opened in Microsoft Excel and reviewed for consistency and formatting. We took the opportunity to adjust date formats to conform to the ISO 8601 time and date standard. This proved a bit tricky in Microsoft Excel, which autoformats dates seemingly according to its own whims. We quickly learned that we had to format date columns in Excel as text-only so that this auto-reformatting didn't undo our ISO 8601 compliant dates. We also reviewed our rights statements and changed them to conform with the Rightsstatements.org formatting, which was recommended to us by Michigan's DPLA Service Hub staff. Due to our tight timeline, we were not able to take the time for more extensive metadata cleanup prior to ingest in Omeka. A copy of the edited collection metadata was saved in Excel format to preserve our formatting and enable further editing as needed.

Next, the digital access files were uploaded to a staging bucket in Amazon S3 cloud storage. File URLs were copied into the metadata spreadsheets, and CSV copies of the spreadsheets were saved. These CSV files could then be ingested using Omeka's bulk ingest workflow. In this workflow, Omeka parses through the CSV, creates an item record for each line while mapping the metadata fields to column headers, and copies the files linked from the Amazon S3 bucket into the Omeka server, to be presented in the item record alongside associated metadata.

The team worked nearly full-time on this project from February to June of 2017, and successfully migrated about 80% of the digital collections to Omeka. The remainder were only offline for about a month before the job was complete.

POST MIGRATION CLEANUP AND LESSONS LEARNED

Following this intensive migration process, some staffing changes occurred that resulted in the dissolution of the DOWG team. A new library dean started in the summer of 2017, bringing along considerable organizational change as well. I became the University Archivist and Digital Collections Librarian in mid-2018, assuming responsibility for continued curation and development of SCUA's digital collections, while responsibility for digital preservation activities was assigned to the Metadata and Digital Curation Librarian.

The year following the migration, my direct-report metadata specialist and I focused our digital collection efforts on continuing the metadata cleanup process within Omeka. We were able to use Omeka's bulk metadata editing tools to establish consistency in our Subject, Type, Format, Publisher and Language fields. Unfortunately, a miscommunication during the migration process led to the Date field and Coverage field being transposed in some of the collections. In these instances, only a manual editing of each item record would fix the error.

In hindsight, I would take the time to create a very straightforward Dublin Core metadata guide and crosswalk for each team member to refer to during their migration work. We relied too heavily on the CONTENTdm export field mapping and didn't have any checks in place before each member ingested their assigned collections. Some team members with less metadata experience misunderstood the purpose of the fields, as well as the "Date" vs "Date created" fields from the CONTENTdm exports.

CONCLUSION

We're now approaching our 6th year in Omeka, and the platform is still meeting our needs. We've been able to add new items to old collections, including approximately 400 new Veterans History Project interviews. We've also created a number of brand-new collections through digitization and collaborative digital projects such as oral history projects and history harvests. While things are still going well with our digital collections, we are also in the stage of planning to sunset Omeka, as we recognize that migration is regular part of the digital collection lifecycle, and new technologies and platform options are continuously being developed.