Digitization in the Real World

Lessons Learned from Small and Medium-Sized Digitization Projects

Edited by
Kwong Bor Ng & Jason Kucsma

M Metropolitan New York Library Council

Digital Treasures: The Evolution of a Digital Repository in Massachusetts

Dodie Gaudet (Central MA Regional Library System)
Kristi Chadwick (Central/Western MA Automated
Resource Sharing System)
Jan Resnick (Western MA Regional Library System)

Abstract

Digital Treasures is a digital library collection of the history of central and western Massachusetts. It is a collaborative project among Central/Western MA Automated Resource Sharing System (C/W MARS), Central MA Regional Library System (CMRLS) and Western MA Regional Library System (WMRLS). Initiated by C/W MARS in 2006, Digital Treasures began as a pilot program when C/W MARS purchased equipment and software and set up a scanning lab at its headquarters in Worcester. Currently Digital Treasures has 36 collections from libraries, with over 1,300 accessible images. C/W MARS, CMRLS and WMRLS continue to collaborate on ways to bring funding, selection guidance and metadata expertise to their member libraries and bring access to the wealth of cultural history of the Commonwealth.

Keywords: Central Massachusetts Regional Library System (CMRLS), Central/Western Massachusetts Automated Resource Sharing (C/W MARS), Digital Barn Raising, Digital Treasures, Digital Commonwealth, Western Massachusetts Regional Library System (WMRLS)

History & Background

In the early 1960s, three Regional Library Systems were formed in Massachusetts to provide a variety of services directly to the public libraries in the western, central, and eastern parts of the state. The services were (and still are) provided at no cost since the Regions are funded by the state. When automation began entering libraries some 20 years later, the WMRLS (Western Massachusetts Regional Library System, 2010) and CMRLS (Central Massachusetts Regional Library System, n.d.) were instrumental in helping to establish a fee-for-service automated network available to all types of libraries named Central/Western Massachusetts Automated Resource Sharing or C/W MARS (C/W MARS Inc, 2009).

In 1998, the Regional Library Systems became multi-type. All public libraries are automatically members of a Massachusetts Regional Library System (MRLS). All academic, school, and special libraries are eligible for membership if they meet basic qualifications. Any WMRLS or CMRLS member can join C/W MARS by paying the membership fees. Not all libraries choose to join C/W MARS: some are small and cannot afford the fees; others are larger, have their own network staff and prefer a stand-alone Integrated Library System. The number of libraries in WMRLS and CMRLS is larger than the number in C/W MARS. As of March 1, 2010 membership in the three organizations broke down as follows

	CMRLS	WMRLS	C/W MARS
Academics	17	19	11
Publics	72	103	130
Schools	120	164	2
Specials	26	25	4
Total	235	311	147

With the imminent development of Digital Commonwealth, a state-wide portal and digital repository, CMRLS and C/W MARS were both interested in setting up a small lab so that libraries could begin scanning photos, documents and other items which could also be

discovered via the state-wide portal. A Regional Library System and an automated network in the eastern part of the state had each purchased a scanner and offered instruction on its use, but few libraries took advantage of the opportunity. Considering this, C/W MARS proposed that one of its tech-savvy staff members who had an interest in digital photography be in charge of the lab and do the actual scanning if local library staff supplied the metadata. CMRLS had a cataloger on staff who could conduct workshops on Dublin Core and be available for metadata consultation. Since the offices of the two organizations are a mere seven miles apart, joining forces to support a single repository made the most sense and Digital Treasures (see *Digital Treasure*, 2009) was born.

The Scanning Lab

Due to some new memberships and some existing members upgrading, C/W MARS had enough discretionary funds to purchase hardware and software and create a new part-time position. The Executive Hardware for the new lab consisted of an Epson Expression 10000XL flatbed scanner with an 11" x 17" bed. A new server had recently been purchased for the network's regular activity and the old server was used for Digital Treasures. After much research, OCLC's CONTENTdm was chosen for the creation, storage and access to digital materials. Anticipating the future growth of Digital Treasures and the inclusion of text, OCR extensions were purchased at that time. Other programs loaded to the digital lab workstation included Adobe Photoshop CS2 for creation of derivative files, Adobe Acrobat for PDF conversion, and Datacolor Spyder 3 for display color calibration.

The Pilot Project

The principals involved in the Pilot Project were Michael Bennett, the new part-time Digital Initiatives Librarian at C/W MARS; Dodie Gaudet, Consultant for Bibliographic and Technical Services at CMRLS; and Jan Resnick, Assistant Regional Administrator for WMRLS. Dodie and Jan worked directly with the libraries in their respective Regions; Michael, who had previously attended the

Northeast Document Conservation Center's (NEDCC) School for Scanning, did all of the scanning and was available to answer the technical questions.

First some ground rules were established for the project. It was decided that the subjects of Industry and Agriculture would allow for maximum participation yet give the repository a focus. Concentrating on pre-1923 objects would avoid copyright issues. Each library would be limited to 20 scans. Materials were to be two-dimensional, no larger than 11" x 17", and in good enough condition to be handled without falling apart. Once library staff had created the metadata, they would be responsible for transporting their items to and from the C/W MARS headquarters in Worcester.

Information sessions were held in each Region to promote Digital Treasures and explain the Pilot Project. Next, Michael, Dodie and Jan developed an online questionnaire which was sent to the directors of all Regional member libraries. Directors were asked if they were interested in participating in Digital Treasures, how much material they owned, what kinds of materials they owned, how much they ultimately planned to contribute to Digital Treasures and related questions.

As the cataloger, it was Dodie's responsibility to learn Dublin Core and develop a two-hour workshop. In preparation for the project, she attended Introduction to Digitization offered by NEDCC in October of 2005 and their Persistence of Memory Conference in November of the same year. Using the Dublin Core Metadata Initiative (DCMI) site (Dublin Core Metadata Initiative, 2008) and the Bibliographic Center for Research (then Collaborative Digitization Program) Best Practices Document (CDP Metadata Working Group, 2006) she synthesized material on Dublin Core for Digital Treasures participants.

Simple Dublin Core was used so as not to overwhelm the staff creating the metadata since most of them were not catalogers. The handouts are available on the Digital Treasures site. Workshops were scheduled in both the Western and Central Regions. Participants

brought examples of the objects they planned to have scanned and practiced their Dublin Core descriptions on paper forms.

The first library to complete the metadata for 20 images was the Jacob Edwards Library in Southbridge. The descriptions were keyed into an online form and e-mailed to Michael at C/W MARS. The photographs were delivered to C/W MARS for scanning and matched to the metadata. By the end of the Pilot Project in August 2006, 24 libraries, twelve from each Region, each had a digital collection of 20 images on Digital Treasures. Libraries also received a CD of their images to use for publicity or any other purpose they chose.

To allow the repository reach beyond the Commonwealth's borders, Digital Treasures metadata have always been compatible with the OAI standard (Open Archives Initiative, n.d.). The metadata are harvested both by the Digital Commonwealth and OCLC's OAIster project.

When specifications were being developed for Dublin Core records, the principals decided to use Library of Congress Subject Headings so that the records would integrate well with the MARC records in the online catalog. After the Pilot Project launched, Michael began working with staff at OCLC, C/W MARS and member libraries to crosswalk the Digital Treasures Dublin Core metadata into MARC format that could be loaded into OCLC Worldcat. As a subscriber to OCLC, C/W MARS would not only have their digital records available online, but would be able to export the records into their own public catalog with referring links back to the collections. Collections available in October 2007 were harvested and their MARC records are now available through WorldCat and the C/W MARS online public catalog as well as Digital Treasures.

After the Pilot Project

Up to this point, all costs for Digital Treasures had been assumed by C/W MARS, but the organization could not justify supporting this service exclusively with membership fees. The decision was made for C/W MARS to subsidize Digital Treasures, and charge participants a portion of the costs involved in scanning and maintaining the digital

repository. With a vote of the membership, the following fee schedule was instituted and is still in effect:

All C/W MARS members are allowed 20 free scans. Each scan beyond the first 20 is $8.50 (50% of the actual cost to digitize an object). After scanning, the library pays $1.00/image annual maintenance fee. Libraries who are not C/W MARS members receive no free scans, pay $12.75/scan (or 75% of the actual cost) and $1.50/image annual maintenance fee.

The annual fee is assessed for the maintenance of each digital image's permanence over time through future platform migrations and regularly scheduled data refreshment cycles. It also covers maintenance of online delivery through CONTENTdm and the allowance of OAI metadata harvesting.

Once the Pilot Project ended, activity in Digital Treasures slowed down drastically. A couple of additional information sessions were held along with another Dublin Core workshop. Everyone liked the idea of having photographs and documents from their local history collections digitized, but had difficulty finding the time to select objects and describe them in a useful way. Even though the financial situation was relatively healthy at the time, some libraries found they could not justify the scanning fees. In the fall of 2007, Dodie applied for and received a grant from the Greater Worcester Community Foundation (GWCF) to cover scanning fees for libraries. The grant was called the Central Mass. Memory Project.

More importantly, the grant from GWCF funded a metadata specialist to work with the libraries selecting and describing their objects for digitization. Jeff Monseau, a Library and Information Science student at Simmons College who also worked part-time with the Archives of Mount Holyoke College, was hired as a "circuit rider." Because of the mission of GWCF, funds could only be used to help libraries within the Central Region. For those libraries, Jeff's experience, even the mere fact that he was available to help, was a tremendous boon and several new libraries added 20 images while other libraries increased the size of their digital collections. The grant

funds enticed some non-C/W MARS members to participate in Digital Treasures which broadened the collection.

In early 2008, Michael moved on to the University of Connecticut and Kristi Chadwick became the administrator of Digital Treasures.

Digital Treasures on the Road

Digital Treasures had been up and running successfully for over two years. Digital Commonwealth had developed a repository as well as a portal and was publicizing its services, stimulating more interest in digitization. For libraries that had not yet taken the step, the Technical Services Section of the Massachusetts Library Association (MLA/TSS) organized a one-day conference in the fall of 2008. The committee chose the topic *Introduction to Library Digitization* and invited several people from throughout New England to talk about their projects, how the projects were started and to describe the practical, day-to-day aspects. Kristi, Dodie and Jeff were asked to talk about Digital Treasures, its technical details, and the Central Mass. Memory Project.

A Digital Barn Raising

As one travels westward through the Commonwealth, the area becomes more sparsely populated. Most towns have libraries, but the libraries are often small and some are staffed by a single person. The median population in western Massachusetts is 1900. Springfield, the largest community, has over 150,000 residents, meaning some of the other communities are significantly below the median. The towns are also rather dispersed by New England standards, and there is little or no public transportation. The western Massachusetts approach to Digital Treasures – as with many projects - relies on team work.

In May 2008, Digital Commonwealth received a grant from the H. W. Wilson Foundation for conversion efforts. The timing was fortuitous as WMRLS was in the process of organizing a "digital barn raising", a cooperative effort to get libraries started. To follow the theme, everyone contributed. Kristi Chadwick brought an introduction to Digital Treasures. A member of the Digital

Commonwealth Board, contributed information about the Digital Commonwealth and described the grants available from the Wilson Foundation. A librarian from the Berkshire Athenaeum in Pittsfield reviewed the major copyright issues involved in digital projects. Dodie Gaudet presented the basics of Dublin Core and how libraries would use it. Member libraries brought a few of their local documents. Jan Resnick coordinated the program and arranged a few extra catalogers to assist in the process.

The program consisted of presentations, questions and discussion, and concluded with a work session in which participants from member libraries began to develop Dublin Core metadata. This structure worked well for the group. They had the opportunity to ask questions (there were lots about copyright issues) and then had the chance to develop initial Dublin Core records for their documents in a supportive environment. WMRLS hopes to offer additional "barn raisings" expanding the skill base to a wider group of small libraries. Their collections contain a broad variety of treasures to be digitized. Buckland Historical Society, which participated in the barn raising session, went on to apply for and receive one of the Digital Commonwealth's Wilson Grants. Collaborating with the Buckland Public Library, they submitted over 50 objects to Digital Treasures. The Erving Public Library also received a Wilson grant and contributed 20 objects to the repository. The idea of the barn raising generated so much interest that participants were asked to contribute to a panel discussion at the day-long Digital Commonwealth conference held in the spring of 2009.

Workflow

Libraries with an interest in having objects digitized contact Dodie, Jan or Kristi. Items are limited to two-dimensional, 11" x 17" maximum size for scanning, or audio files in MP3 format. The original scope of agriculture or industry was lifted, so the theme of the collection is up to the library. For libraries that are not familiar with Dublin Core, Dodie assists with training, if based in the Central

Region. Jan and Kristi assist libraries in the Western Region. This can be done by e-mail, phone, or in-person consultation.

When the library has selected the objects they wish to add, Kristi gives them access to the password-protected online form. This form can be printed as a worksheet for initial descriptions, then entered online and submitted to Kristi by e-mail. The library also receives a copy for its own records. Once the library submits the first record for review, then all metadata can be sent.

Once the metadata are received, an appointment is made to go to the C/W MARS headquarters in Worcester, where the scanning lab is located. Depending on the number of objects to be scanned, a two-hour or more block of time is reserved for the library. If the collection is new, preparation for the library's appointment will include setting up folders on the CONTENTdm server and workstation and verifying the administrative metadata, which are kept in a Microsoft Access database.

When the library representative arrives with the images and/or documents, all items are matched by the identifier assigned by the library to the one listed in the metadata. Scanning on the Epson printer is done with its native program. All images are scanned in TIFF format for the "master" files, with a long side of 3000 pixels. Resolution of TIFF files range between 300 to 600 dpi, to achieve desired pixel dimensions. During the scanning, information from the metadata workform is copied into the MS Access Administrative Metadata database, including the title given to the object and the identifier assigned by the library.

After initial scanning is complete, the TIFF files are manipulated in Photoshop and converted into the derivatives. This includes autoleveling and creation of JPEG files. Two separate JPEGs are created, a full-size image and one at access size for CONTENTdm with a long side of 600 pixels and 72 dpi. The full size JPEG is loaded to the CONTENTdm server directly, and the access images are loaded through the CONTENTdm workstation, where the rest of the metadata are loaded in from the workforms then uploaded to the server for web presentation.

Stumbling Blocks

From the beginning, there has been some resistance from a few libraries to the digitization of their collections. The reasons vary widely. Some staff and trustees are concerned that by making documents available digitally, people will no longer feel the need to visit their libraries. Only experience will breach this barrier.

There is fear among a few people that digitizing library holdings and making them available on the World Wide Web exposes a library's materials to being used fraudulently. There is no evidence of this happening, but those who feel this way are not easily convinced otherwise.

The copyright issue of post-1923 material is one that concerns library staff. Many photographs have been donated to libraries, and their origins are unknown. As Digital Treasures expands and libraries have more of their hidden collections scanned, this is an area that needs more research, more specific direction on what can and cannot be digitized and how to proceed if there is a complaint. In the meantime, concern over possibly breaking the law limits what objects libraries are willing to contribute to Digital Treasures.

The digitization process has developed into a "three-legged" stool, comprised of Selection of Materials, Descriptive Cataloging/Development of Metadata, and Financial Resources. Without support for all three, digitization does not happen. Carving out time for staff to select objects for scanning has become more difficult since Digital Treasures started. The world-wide financial crisis quickly trickled down to libraries and many libraries have had to lay off staff and reduce the working hours of others. Selecting materials for inclusion in Digital Treasures is time-consuming even when the local history collection is well organized and many are not.

There are fewer people who feel comfortable with original cataloging. Even after objects have been chosen, it is easy to get bogged down in the process of descriptive cataloging. Digital Treasures participants have always been encouraged to create thorough metadata so that their images are more easily found via

keyword searches. Unfortunately, the process is overwhelming to many, especially those who already feel time pressures.

No one thinks that the scanning fees, which are subsidized by C/W MARS, are unreasonable. However, with tight budgets, every penny is scrutinized. For a C/W MARS member library, adding 20 more images to its collection not only costs $170, but the annual maintenance fee is increased by $20 - a commitment that some libraries are unwilling to make. Financial incentives in the form of grants can help, especially when the grants include an experienced person to help with selection and metadata.

The Future of Digital Treasures

Currently the repository stands at 36 collections, with over 1300 items. The digital repository's contents are searchable through the native Digital Treasures front end, the Digital Commonwealth portal, OAIster, major Internet search engines such as Google and Bing, WorldCat and the local C/W MARS online public catalogs. New digital file formats, such as audio recordings, have been included in the repository. The Digital Commonwealth presented a grant round in 2009 to small institutions in Massachusetts, allowing some libraries to match funds to create new collections for the coming year. Another grant round is in the works for 2010.

Knowing how much personal assistance with the selection and metadata process can benefit libraries, C/W MARS has added the Digital Treasures project to the University of Rhode Island (URI) Graduate School of Library and Information Studies (GSLIS) Site Experience database for a possible internship. While statistics show that Digital Treasures is accessed, with nearly five million hits and over 550,000 unique visitors in 2009, the success for the future lies in continuing contributions and promotion of the repository.

References

C/W MARS Inc. (2009). *Connecting libraries in Central and Western Massachusetts*. Retrieved on May 2, 2010 from http://public.cwmars.org/

CDP Metadata Working Group. (2006). *Dublin Core metadata best practices*, version 2.2.1. Retrieved on March 31, 2010 from http://www.bcr.org/dps/cdp/best/dublin-core-bp.pdf

Central Massachusetts Regional Library System. (n.d.). *Homepage*. Retrieved on May 2, 2010 from http://www.cmrls.org/.

Digital treasures: A Central & Western Massachusetts digital library project. (2009). Retrieved on May 2, 2010 from http://dlib.cwmars.org/.

Dublin Core Metadata Initiative. (2008). *DCMI metadata terms*. Retrieved on March 31, 2010 from http://dublincore.org/documents/dcmi-terms/

Open Archives Initiative (n.d.). *Standards for web content interoperability.* Retrieved on March 31, 2010 from http://www.openarchives.org/.

Western Massachusetts Regional Library System. (2010). *Western Massachusetts Regional Library System - Libraries Cooperating Communicating Sharing*. Retrieved on May 10, 2010 from http://www.wmrls.org/