

Annotated Bibliography: The Implementation of Digitization in Small to Mid-Sized Archives

This annotated bibliography focuses on the types of issues small to mid-sized institutions would need to consider in the implementation of a digitization project in their archives, sub-divided into topics representing the flow of implementation.¹ Work phases considered include strategy and project management, project costing and funding, software selection, metadata development, selection of records, online presentation, integration of socio-technological software features, and of course legal issues (see table of contents). Woven throughout are the twin rationale for digitization of existing physical material: access and preservation.

¹ Specifically I have tailored resource choices to institutions with more limited resources, for example, focusing on open source software solutions instead of proprietary options.

Table of Contents

Strategy, planning and project management.....	3
Digitization funding, project costs and collaboration.....	7
Choosing software.....	11
Digital imaging.....	16
Selecting material for digitization.....	17
Metadata and networked finding aids for digital material.....	21
Presentation of digital collections, finding aids and user participation.....	26
Legal issues (copyright).....	31
Other resources.....	35

Strategy, planning and project management

Much pre-planning must be undertaken before the actual act of digitizing records occurs. And, as the project proceeds, strong project management practices will make the difference between missing deadlines, communications breakdowns, and smooth execution. The following articles provide first-person/organization accounts and general project principles for digitization work.

Holley, R. (2004). Developing a digitisation framework for your organisation. *The Electronic Library*, 22(6), 518-522.

Written by the Digital Projects Librarian at the University of Auckland Library, this article describes the activities undertaken prior to beginning digitization at their academic institution. These activities included an inventory of digitization projects, staff awareness-raising and training, obtaining funding, enhancing the IT infrastructure and developing a digitization policy. The article provides lists of tasks accomplished at the University of Auckland Library, and lessons learned.

As a project leader at her institution, the author is able to share a great deal of practical information about digitization planning. Holley encourages libraries and archives to “feel the fear and do it anyway”—or to begin the work of creating a framework in advance of the implementation of a digital repository. Although her organization accomplished this last, she suggests that strategy and policy creation should come first, and that a digitization policy should “outline goals, guiding principles, selection criteria, management and access to digital collections, digital standards and guidelines to follow, intellectual property rights and digital preservation,” (p. 521). She

also suggests digitization should be instigated through a small pilot project that “addresses all the issues you are likely to face in a big project,” (p. 521). In conclusion, Holley suggests libraries and archives should take a number of steps to begin digitization, including launching a steering committee, finding an internal champion, and auditing staff skills, possible projects and IT capacity.

This article is helpful for any repository in need of an organizing framework to structure their initial digitization activities. I was able to use the article and the related digitization guidelines in creating resource information for the Royal Conservatory of Music Archives, where I recently completed a practicum. A high level overview, the topics Holley discusses are expanded further in other parts of this bibliography. A key learning from this article is the need to do groundwork in a number of areas, to whatever degree resources allow, prior to actually scanning items.

Related resources:

Holley, R. (2003). Digitisation guidelines - University of Auckland Library. Retrieved from <http://www.scribd.com/doc/54758971/ion-Guidelines>

These guidelines represent the actual product created at the University of Auckland to guide their digitization efforts.

JISC Digital Media (2012). Case study: learning lessons from other digitization projects. Retrieved from: <http://www.jiscdigitalmedia.ac.uk/case-study/learning-lessons-from-other-digitisation-projects/>

Tips on the various aspects of digitization work.

Eden, B. (2001). Managing and directing a digital project. *Online Information Review*, 25(6), 396-400.

Another first person account, in this case in diary form, the writer describes his lessons learned from a first-time project to scan/digitize historical images of Las Vegas held within Special Collections at the University of Las Vegas Libraries. While the writer walks through the steps involved in this project, the particular value of this article is its focus on collaboration and communication between team members involved in a digitization project. This was a first time project for all involved. Team members worked enthusiastically but separately on different aspects of the work, which led to conflict and challenges in conforming to timelines.

Eden provides four lessons learned, the first around managing creativity: “When dealing with individuals and groups involved in digital project creation, watch closely to make sure that creative minds and creativity do not overwhelm the ultimate project timelines and objectives,” (p. 399). He also emphasizes the need for communication between team members, the oversight of a team leader with good communication skills, and the importance of leaving time for user and team review before launch. Finally he suggests that, no matter the bumps along the road, “take pride in the work and accomplishments of the shared goal, and use the experience to make the next digital project that much smoother and enjoyable,” (p. 400).

Compared to the Holley (2004) article, this article is not as detailed in describing the technical issues and tasks of digitization project management. But, as is discussed in literature regarding the implementation of EDRM systems, Eden emphasizes the importance of communication to the success of any digital project—particularly when

the digitization task is new to the entire team and everyone is in learning mode. This fact should be top of mind as any small to mid-sized organization begins to digitize their records.

Related resources:

Downing, L. (2006). Implementing EDMS: putting people first. *The Information Management Journal*, 40(4), 44-50.

Maguire, R. (2005). Lessons learned from implementing an electronic records management system. *Records Management Journal*, 15(3), 150-157.

These articles emphasize the importance of on-going communication to EDRMS implementations, a transferable lesson for digitization projects.

Chapman, S. (2000). Considerations for project management. In M.K. Sitts (Ed.), *Handbook for digital projects: a management tool for preservation and access* (pp. 21-33). Andover, Massachusetts: Northeast Document Conservation Centre.

In the introduction to this book chapter, Chapman notes that digital projects come with their own challenges, including the constant need to adapt to technological change. Even so, he asserts that digital projects can be well-managed and structured, starting with clear goals and outcomes and working backwards. The chapter also includes detailed information about creating a plan of work and budget, and project staff roles for digitization work. Further, Chapman comments on project implementation and managing workflows, listing the numerous steps involved in implementation.

Ultimately, Chapman notes that, “the first measure of best practice is likely to be one of the ends justifying the means,” (p. 30), making the crucial point that project planning, goal- and software specification-setting must precede technology choice.

“Ultimately the project manager, not the technology manufacturer or distributor, must be the one to judge whether a given system will do the job that is needed,” (p. 23).

This chapter complements issues discussed in the previous two citations. The workplan, budget, roles and workflow information are provided at high level, but are a useful resource for an organization.

Related resource:

Downer, S., Medina, S., Nicol, B. and Trehub, A. (2005). AlabamaMosaic: sharing Alabama history online. *Library Hi Tech*, 23(2), 233-251.

This article provides another case study of a digitization project, in this case a collaboration of Alabama institutions. The study addresses many of the issues discussed in the above chapter at the level of one project. Further, that particular issue of *Library Hi Tech* is entirely devoted to digitization, so other resources and case studies can be found in edition 23(2).

Digitization funding, project costs and collaboration

Digitization projects are expensive and the demise of NADP funding in Canada will make it more challenging for small to mid-sized archives to fund their projects. Following are general discussions to assist organizations in the costing of digitization projects and suggested routes to obtain funding. Collaboration is also discussed as an approach that can assist archives to achieve sustainability for digitized collections, share costs and obtain greater visibility for their collections. Data preservation and costs are discussed further in a later section.

Anderson, C.G. & Maxwell, D.C. (2004). Funding. In *Starting a Digitization Center*, (pp. 125-137). Oxford, UK: Chandos Publishing.

When undertaking a digitization effort, costing is a significant factor in the ability to both launch and sustain a digital collection. This book chapter reviews at high level the funding considerations that must be addressed at the outset of a project. The authors note that the three main areas of digitization costing are related to: technology and workflows, intellectual property, and institutional (ongoing infrastructure) costs.

The authors make several useful suggestions in the area of budgeting. They argue that the best way to keep costs down is through economies of scale, whether internally or through collaboration with other institutions, as the most difficult and costly phase of a digitization project is at the outset. They also emphasize the importance of a pilot costing exercise to determine the actual cost per page or item for your institution. "Take a sample of the materials you plan to digitize and take them through the process of scanning, metadata creation, OCR, image editing, proofreading and text markup," (p. 128). However, this sample must actually be representative or you may find hidden costs at a later date. Another consideration is the use of outsourcing as a cheaper alternative to buying specialized equipment and finding internal labour for scanning, although this may not be viable for smaller organizations.

Three sources of funding for digitization are generally discussed. Grant funding is difficult to sustain in the long-term due to the resources required to continually re-apply for grants. Cost recovery through charges for printouts is also possible, but does require some administration. The authors also argue that institutional support over the long-

term is critical to ensure the budget for ongoing maintenance, additions to the collection and data preservation. This requires the organization, at whatever size, to see digital collections as providing global visibility for the library or archives, and value for users and the community.

The question of funding is an important piece of the project management process, as discussed in the articles in the project management section. Costing must be integrated into every project phase, and again, for long-term sustainability some part of the overall archives budget must be set aside for the administration of digital collections. A key link between this discussion and that of project management, as discussed by Holley (2004) and Eden (2001) is the stress around conducting a successful project that does not go over budget. This requires both realistic budgeting and strong execution—success in these areas then advances the cause of devoting a greater percentage of the archives budget to digital activities.

Related resources:

Bia, A., Munoz, R. & Gomez, J. (2010). DiCoMo: the digitization cost model. *International Journal on Digital Libraries*, 11, 141-153.

If conducting a large digitization project, an archives may want to consider this recent article. The authors have developed more complex equations and processes to estimate the costs and time required to digitize large amounts of textual material. From a library perspective, it does present a framework that appears more complex than would be required for a small digitization project, but may be of interest.

Hughes, L. & National Initiative for a Networked Cultural Heritage (2003). *Report on the price of digitization: new cost models for cultural and educational institutions*. Retrieved from <http://www.ninch.org/forum/price.report.html#top>

This link includes presentation summaries from a forum on the price of digitization co-sponsored by the New York Public Library and New York University. Although this symposium took place almost 10 years ago, the summaries provide a glimpse into a

number of issues and case studies regarding pricing, including the “Making of America IV project” and an address by Stephen Chapman on the costs of digital preservation and affordability of repository storage.

Bonn, M. (2001). Assessing the costs of conversion: making of America IV: the American Voice 1850-1876. Retrieved from http://www.lib.umich.edu/files/services/dlps/moa4_costs.pdf

This handbook focuses on the conversion of historical books into digital form. Again it may not be fully appropriate for the digitization of an archival collection but it does provide a very detailed discussion of project costs on a large digitization project.

Middleton, K. (2005). Collaborative digitization programs: a multifaceted approach to sustainability. *Library Hi Tech*, 23(2), 145-150.

In Canada, portals like Our Ontario and Canadiana.org make the digitized material of many smaller archives searchable and available more widely. This article introduces an issue of *Library Hi Tech* journal devoted specifically to collaboration around digitization in US states. In his overview of the issue, Middleton alludes to some of the financial benefits of collaboration, including the ability for archives to conduct business planning together for the sustainability of state-wide projects. He also points to the benefits that state-wide projects provide to accessing inexpensive training, digitization centres that provide scanning hardware for multiple institutions to use, and installations of the same content management software at different sites.

Middleton notes that “some of the most challenging aspects of collaborative digitization programs are social rather than technical,” (p. 146). Whether within an institution or through a collaboration of multiple institutions, professionals from IT, libraries, archives and museums must find common ground across their different vocabularies, practices and funding structures. An inclusive approach that respects

differences (for example in metadata needs) while opening participants to new resources, training and tools is required.

Collaboration can be an approach to address issues discussed in other areas of the bibliography, from user interface to software selection. More Canadian focused information is provided at the following links.

Related resources:

Davis, C. (2010). West beyond the west: digitization symposium. *British Columbia History*, 43(4), 10-14.

Davis summarizes a 2008 symposium on digitization held at the University of British Columbia, providing information on many digitization projects in BC and across Canada, including collaborative initiatives, and sources of funding being employed.

Bell, B. (2008). Canadiana.org: one of Canada's oldest and newest digitization initiatives. *Partnership: the Canadian Journal of Library and Information Practice and Research*, 3(1), 1-4.

Bell discusses the work of Canadiana.org and the Canada Project as a national digitization strategy (although I am unclear where this work stands in relation to recent changes at Library and Archives Canada).

Choosing software

As noted in the project management section, technology choice should not drive a project, but be the outcome of clear project goals and needs. However, project managers will ultimately have to make a selection based on criteria they create. For a small to mid-sized archives with limited budgets, open source software options provide free technology and the ability to join a community of users working together to improve functionality. The following specifically refers to software that manages and

displays digital content and does not refer to archival management system software.

While by no means a description of all choices available, some discussion of open source options and selection processes follows.

Goh D. et al. (2006). A checklist for evaluating open source digital library software. *Online Information Review*, 30(4), 360-379.

This article provides a set of criteria for evaluating open source digital library (DL) software, and applies those criteria to four DL platforms. The criteria reflect an effort to match software functionality to user needs, and fill a gap in literature to that point. The authors are associated with the Division of Information Studies, School of Communication and Information, Nanyang Technological University, Singapore and wanted to use their evaluation to fuel their own digitization project within the school.

The authors determined a number of functions of “good” DL technology including aspects of content management, user interface, user administration, system administration and other requirements. Software (Greenstone, Fedora, Eprints and CERN CDSware) was then evaluated on comprehensiveness, useability, flexibility and expandibility. The authors note that this work “represents a first step in the development of an exhaustive evaluation tool and...the checklist does not take into account factors such as hardware, time, manpower, money and other resources, as these may vary depending on the implementing organization or individual,” (p. 371).

While other project management articles cited in this bibliography mention technology choice, this article works to fill a gap in providing detailed guidelines for

technology choice. Although the scoring scheme is complicated and the criteria fairly technical, the digital library evaluation checklist in the Appendix does contain a wide-ranging set of qualities to select from in the evaluation of software for an organization. Further, the bibliography, now somewhat dated, may contain some other useful reading.

Voth, S. (2010). *Advantages and challenges of open source web publishing: a case study* (unpublished master's thesis). Graduate School of Library and Information Science, Queens College of the City University of New York: New York.
Retrieved from:
<http://www.pcnyc.edu/omeka/pdf/AdvantagesandChallengesofOpenSourceWebPublishing.pdf>

Open source software is often the best choice for smaller archival repositories, due to low cost (free), unrestricted licenses and continual improvement by a community of expert users and programmers. This Masters thesis project describes the transfer of an archival postcard collection from proprietary content management system CONTENTdm to the Omeka open source web publishing platform. Omeka is a relatively new platform that operates similarly to WordPress, and is increasingly being adopted by libraries, archives and other heritage organizations, particularly for its web publishing and exhibition capacities. While providing a good discussion of the merits of Omeka, this thesis is also useful as it details the steps the author took to migrate the digital objects and create the digital collection.

Voth discusses a number of areas of using Omeka. These include the experience of installing Omeka on his system, using the Omeka plug-ins, which provide various types of functionality, modifying Omeka's out-of-the-box themes (for look and feel),

batch import of content, metadata, search and retrieval testing, exhibit building, and tagging. Various tips and tricks he learned through the implementation are discussed and may be useful to an organization learning Omeka for the first time.

In conclusion Voth provides an important caveat for archives as they choose software solutions for their digital collections:

The open source model is an attractive alternative to proprietary Web publishing, but as the literature points out, adopting such a solution is not free (Trappler, 2009; Ven et al., 2008). Open source requires a level of expertise, which may involve additional staff training, or under some situations, employing consultants to do the groundwork of setup and customization. Additional expenditures of staff time may also outweigh any cost savings, but as some note (O'Reilly, 1999; Ahmed et al., 2009), the superior quality of the software more than compensates for the hours spent. Open source software is by nature extensible and customizable, and if such an approach is used for web publishing, an organization assumes full control and responsibility over its collections. Success depends upon the willingness of staff members to take on new roles and challenges, and to build upon their expertise. (p. 52)

Open source software may lack proper documentation, and support is often only available through web forums of users and developers. But, Voth notes that the democratic nature of open source means that the community can address issues and shortcomings of the programs over time, and as a user you can add your voice to influence that process. Narrowing down the discussion from the general guidelines in Goh et al. (2006), this thesis speaks to the needs of smaller archival organizations as they determine the best software approaches to meet their budget, infrastructure and staff capacity.

Related resources:

Kucsma, J.et.al. (2010). Using Omeka to build digital collections: The METRO case study. *D-Lib Magazine*, 16(3/4). Retrieved from <http://www.dlib.org/dlib/march10/kucsma/03kucsma.html>

This case study details the experience of the Metropolitan New York Library Council to build a digital collection using Omeka, and is complimentary to the above discussion by

Voth (2010).

Boggan, K. (2011). Creating digital archives using WordPress. In K.M.L. Jones & P.-A. Farrington (Eds.), *Using WordPress as a library content management system*. Library Technology Reports, retrieved from alatechsource.org.

Boggan discusses the experiences of archives that have used WordPress to create digital archives. Currently an ubiquitous open source platform for web publishing, one may not think of WordPress first when looking for a digitization/content management software. But this discussion points to the fact that software options are changing quickly, and do not all fall within the same software categories. If an archives has more technical expertise, some creative thinking could lead to an unexpected choice.

Boggan first discusses non-archival content management options Drupal and Joomla!, and archival options Omeka, ContentDM and Greenstone. She then describes archival use of WordPress in its traditional form as a blog. Finally she describes a WordPress plug-in called Scriblio, which adds content management abilities to WordPress. The author was hired by the Cleveland Colby Colgate Archives to customize Scriblio for their digital collection needs. While in 2008 this functionality was not particularly robust in terms of metadata and search, and required a great deal of technological skill to implement, the author notes that upgrades to WordPress in 2010 further integrated content management abilities into WordPress itself.

Again, this article shows that there is more than one option available, both from within software created specifically for libraries, archives and museums, and from more

general web publishing and content management systems. Software choice may at least partially depend on whether an archives has the budget to hire a consultant developer, or has the skills in-house to customize a solution—and how much customization is worth the time and effort given timelines, size of repository and future needs.

Digital imaging

There is a great deal of information on the technical processes involved in creating digital images of archival, library and museum material. A number of standards apply, as can be noted in the Federal Agencies Digitization Guidelines Initiative resource list (see below). A smaller organization may want to peruse the technical standards developed by large libraries and archives such as: National Library of Australia, Library of Congress, Canadian Heritage Information Network/Canadian Museum of Civilization and Online Computer Library Centre (OCLC). This information includes scanning resolutions and processes, file formats, and other aspects of digital imaging for various types of records (textual, photographic etc.). An archives will also need to consider the purchase of a scanner and/or photographic equipment, and the National Library of Australia link below provides a sense of the equipment choices made by a large organization.

Related resources:

CHIN and CMCC (2007). Digitization standards for the Canadian Museum of Civilization Corporation: scan and artefact photography. Retrieved from <http://www.pro.rcip-chin.gc.ca/sgc-cms/coursenligne-onlinecourses/index.php>

Federal Agencies Digitization Guidelines Initiative (2009). A resource list for standards related to digital imaging of print, graphic, and pictorial materials. Retrieved from <http://www.digitizationguidelines.gov/guidelines/digitize-standards.html>

Library of Congress (2006). Technical standards for digital conversion of text and graphic materials. Retrieved from <http://memory.loc.gov/ammem/about/techStandards.pdf>

National Library of Australia (2012). Image capture standards. Retrieved from <http://www.nla.gov.au/standards/image-capture>

National Library of Australia (2012). Digital capture equipment. Retrieved from <http://www.nla.gov.au/content/digital-capture-equipment>

OCLC (2008). Preparing digital surrogates for RLG cultural materials. Retrieved from: <http://www.oclc.org/research/activities/past/rlg/culturalmaterials/surrogates.htm>

Selecting material for digitization

Another key aspect of a digitization project is the process of selecting which resources will become part of a digital collection. As digitization of physical resources is expensive and time consuming, usually not all records in an archives will be chosen to be digitized. The authors below discuss the factors that must be taken into consideration in selection, and processes used to make those choices.

Vogt-O'Connor, D. (2000). Selection of materials for scanning. In M.K. Sitts (Ed.), *Handbook for digital projects: a management tool for preservation and access* (pp. 35-60). Andover, Massachusetts: Northeast Document Conservation Centre.

As digitization is expensive, there is a proliferation of content available on the Internet, and copyright issues come to bear, the selection process is very important to

the overall success of a digitization initiative. Vogt-O'Connor discusses a process for nominating materials, evaluating the nominated materials, and prioritizing what remains after weeding, based on value, use and risk.

Similar to the process of archival appraisal, selection for digitization involves a cost / benefit analysis. Vogt-O'Connor notes that archivists must take a number of factors into consideration, such as: appropriateness to repository mission and archival focus, audience demand, legal restrictions on use, accessibility of original material from storage, and barriers to digitization due to format (e.g., framed items). While "many organizations may be tempted to 'just digitize it all' because selection seems labor-intensive and expensive," (p. 36), providing web access to material is akin to publishing and some rigor must be applied before records are shared online (in relation to accuracy, cultural sensitivities etc.).

Vogt-O'Connor also details the benefits of the configuration of a Selection Committee, which can include discipline experts in the collection themes, technical digitization specialists, librarians and/or archivists, researchers/public and lawyers. Once a committee is created, the selection steps are reviewed in detail. Pre-selection tasks involve connecting with other organizations, conducting digital projects to gain expertise, and ensuring cooperation between institutions. Then staff (particularly librarians or archivists) nominate material for the Selection Committee to consider. Next, the Selection Committee would review the nominations and make initial de-selections using uniform criteria. Then if further de-selection is required to meet the goal number of digitized items, records should be reviewed according to value

(informational, administrative, artifactual/intrinsic, associational/relational, evidential and monetary). Risk factors include legal issues, social concerns and preservation issues due to damage through scanning. The third factor to consider is audience use, and the authors suggest conducting a small pilot project to gauge user response before beginning a larger initiative.

This chapter is useful as it provides a number of common-sense guidelines to help in the selection process, and includes sample forms to use in a project. Selection is very much connected and contingent with other digitization issues discussed in this bibliography: the price of digitization, and copyright issues. Getting selection right can place a project in a good position for success.

Related resources:

Buelow, A.E. & Ahmon, J. with Spencer, R. (2011). *Preparing collections for digitization*. London, UK: Facet.

A recent discussion of the preparation of material for digitization.

De Stefano, P. (2001). Selection for digital conversion in academic libraries. *College & Research Libraries*, 62(1), 58-69.

Given the similar audience of academic researchers in an archives and an academic library, this article by De Stefano is potentially relevant for archives. The author discusses the need to select for digitization due to researchers' increased expectation to find material online—a similar pressure coming from archival users. Harkening back to the collection building principles used in the formation of textual library collections, De Stefano suggests the best guide for selection are the needs of the

immediate community of users, as was the case with the selection of material for print in the 15th century when printing became widespread. She also highlights the importance of rigorous macro criteria and professional approaches, given the costs involved and the need to adhere to the overall mission of the parent organization. Then at a micro-level, she outlines six collection-building criteria: “(1) subject, (2) intellectual content, (3) potential use, (4) relation to collection, (5) bibliographic considerations, and (6) language,” (p. 62), arguing that these could be applied to digitization selection.

She then goes on to discuss some of the digitization for preservation approaches in use in libraries, finally concluding that, “salient points taken from the collection-building and selection decision-making models offered above settle most harmoniously around the overriding directive of research libraries to align their collection development practices with their parent institutions and the utility of doing so. The idea of use, especially high use, is fundamental to collection development and is the common thread in all selection decisions. Coupled with the extraordinary access capabilities permitted by digital technology, use holds significant promise as a guiding factor in selecting materials for digital conversion,” (pp. 65-66).

De Stefano’s common sense assertion of use as primary digitization criteria for libraries is complicated in archives by the competing archival priorities noted in Schellenberg’s primary and secondary values – legal and administrative evidentiary uses vs. informational and research uses. But given that it may only be possible to convert a small portion of archival material in a repository into digital form, I would suggest that audience demand and community engagement factors have greater weight in a digitized

archives. Mirroring the authors emphasis on not just use, but high-use, an archives needs to look through research value, communications and exhibition lenses when prioritizing from within their records. Further, the need to ensure a match with the overall mission of the institution seems universal across libraries and archives. One could argue that an archives should “put their best foot forward” and start a digital collection with the most exciting records for researchers and the public, to draw interest to the site. And finally, as an adjunct thought, response of users to digital records through social tools (Krause & Yakel, 2007) may provide even further feedback on which records are most useful after a collection of records are digitized.

Related resources:

Cox, R.J. & Pittsburgh Archives Students (2007). Machines in the archives: technology and the coming transformation of archival reference. *First Monday* 12(11). Retrieved from <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2029/189>

4

This article includes some discussion of the increased expectations of users to find digitized material, and the move to digital services at the archives.

Metadata and networked finding aids for digital material

Good metadata creation is crucial to the success of a digital repository for a number of purposes. Metadata chosen should support authenticity and preservation, track rights information, and provide good descriptions to ensure access. Following are descriptions of some of the standards used for digitized records. An archives may also choose to migrate from their existing metadata schema and/or choose some of the “out-of-the-box” fields included with software platforms.

Dublin Core Metadata Initiative (2008). Dublin Core Metadata Element Set, Version 1.1. Retrieved from <http://dublincore.org/documents/2008/01/14/dces/>

Anonymous (1998). Metadata and the Dublin Core: emerging standards for digitization projects. *Information Intelligence Online Libraries and Microcomputers*, 16(12), 1.

Work on the Dublin Core Metadata standard began in 1995, at an invitation-only OCLC workshop in Dublin, Ohio. As digitization initiatives by libraries, archives and museums proliferated there was the need for consistent and simple metadata to be used widely. Anonymous (1998) notes that while the MARC metadata standard was and is widely accepted in libraries, by 1998 Dublin Core was gaining ground outside the library community and internationally. The sponsors of the Dublin Core initiative represent major players in the field: Library of Congress, National Library of Australia and others. While the latest version is dated 2008, since 2001 the DCMI Usage Board has reviewed any changes made to the element set.

Benefits of Dublin Core include its simplicity and semantic interoperability: the standard includes 15 core elements that can be understood across professions, by non-professionals and across domains. These include contributor, creator, date, format, rights, subject, and title, among others (these can be cross-walked to RAD). The standard does not include controlled vocabularies, assuming that the domain area will supply their own.

Other benefits, as noted by Anonymous (1998) include extensibility “The Dublin Core provides an economical alternative to more elaborate description models such as the full MARC cataloguing of the library world. Additionally DC includes sufficient

flexibility and extensibility to encode the structure and more elaborate semantics inherent in richer description standards,” (p. 3). Dublin Core is used in the Omeka open source software, and is integrated into other packages such as CollectiveAccess.

National Information Standards Organization (2006). ANSI/NISO Z39.87-2006: Data dictionary - technical metadata for digital still images. Retrieved from: http://www.niso.org/apps/group_public/download.php/6502/Data%20Dictionary%20-%20Technical%20Metadata%20for%20Digital%20Still%20Images.pdf

Going down a level from a very general descriptive standard like DC, metadata standards are also available for specific types of records. NISO, the National Information Standards Organization, for ANSI, the American National Standards Institute, has developed this standard for technical metadata for digital raster still images. Originally created through a workshop in 1999, this metadata schema addressed the fact that cultural organizations were focused on descriptive metadata for still images, leaving a need for more robust technical metadata. This standard focuses on data related to quality attributes such as detail, tone, colour and size, as well as elements that could be used to assess the current value of an item (aesthetic and functional). Although originally focused on TIFF format, the standard has been expanded to include other file formats.

Although a smaller archives may not choose to utilize every one of the large number of suggested metadata fields (related to describing the basic file, image capture, image assessment, and change history) it provides a very detailed and technical menu of options if the archives has a focus on images.

PBCore (2011). *Public broadcasting metadata dictionary project*. Retrieved from: <http://pbcore.org/index.php>

Created by the global community of public broadcasters, PBCore is an audio-visual metadata standard that acts as a data model for media cataloguing and asset management systems. While public broadcasters can use this standard in its entirety to customize their media asset management systems, archives may choose to consult the robust technical metadata included in this standard. Like the ANSI/NISO standard, the public broadcasters have put a great deal of effort into developing metadata elements on a number of sub-themes: root elements, intellectual property, elements and instantiation. I am aware that content management systems, for example ResourceSpace, may include some audiovisual metadata fields “out-of-the-box” but this standards allows an archives to go to the experts and delve deeper as needed.

Related resource:

Thomas, S. et al (2007). *Paradigm workbook on personal digital archives*. Oxford, UK: Bodleian Library.

This detailed workbook arises from a text project working to preserve the personal digital archives of a group of politicians. It includes a very long chapter on administrative and preservation metadata that could be another metadata reference for archives.

Riley, J. & Shepherd, K. (2009). *A brave new world: archivists and shareable descriptive metadata*. *American Archivist*, 72(1), 91-112.

Going beyond earlier union catalogues and their online equivalents, this US article discusses the new world of shareable descriptive metadata. Currently this takes place through online partnerships equivalent to archivescanada.ca, and through services such as ArchivesGrid. The authors argue that archival metadata should be shared more

widely in a networked world, to allow archival material to be integrated into services such as Google Earth, Wikipedia or the Internet Movie Database. This represents a change to a greater openness in sharing metadata and requires pre-planning by archives.

The authors note that “Providing [an] optimized shared view requires two phases in implementation: 1) understanding and applying the principles of shareable metadata during initial descriptive metadata creation, and 2) providing the technical means of generating the shared record, ideally using automated rather than manual processes,” (p. 96). The article goes on to describe the qualities of shareable metadata, and discuss the need to strike the right balance in content and contextual information within the metadata descriptions to meet the needs of a much wider audience than usually served by the archives. The same spirit can be applied to both new metadata and older metadata within the repository, with the goal of creating metadata that can be easily translated into use within multiple standards, whether RAD/DACS, EAD, Dublin Core etc.

This discussion is very important for repositories building their metadata schema for a digitized collection, so that they begin the process with the awareness and openness to ensure their metadata is built in a form that can be shared with wider networks later on. It also encourages the on-going collaboration of single repositories with others to build shared standards and processes.

Related resources:

Open Archives Initiative (2002). *Open Archives Initiative Protocol for Metadata Harvesting*. Retrieved from <http://www.openarchives.org/pmh/>

This standard is the technical protocol that allows the harvesting of metadata from repositories as discussed in Riley & Sheppard (2009).

Library of Congress (2002). *Encoded Archival Description*. Retrieved from:
<http://www.loc.gov/ead/>

XML standard that encodes archival finding aids.

Elings, M.W & Waibel, G. (2007). Metadata for all: descriptive standards and metadata sharing across libraries, archives, and museums. *First Monday*, 12(3). Retrieved from
<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1628/1543>

This article talks about the parallel development of data standards by the library, archives and museum communities, concluding that each field did not reinvent the same tool, but created a rich group of tools that applies to their unique needs.

Presentation of digital collections, finding aids and user participation

Any archives working on digitization will also need to focus on the presentation of the digitized records and the addition of Web 2.0 functions that allow users to interact with the records and the archives. Useability—the ability to find the digital records—is key to the success of a digital repository.

Anderson, C.G. & Maxwell, D.C. (2004). Putting collections online. In *Starting a Digitization Center*, (pp. 139-153). Oxford, UK: Chandos Publishing.

After spending much time planning and eventually digitizing primary source material, that material must then be displayed in a manner that is inviting and usable for its audience. Anderson and Maxwell suggest “navigation, navigation, navigation,” (p. 139) must be the primary consideration when placing records within a website. The authors remind archivists of the design basics of audience, purpose and context—the

primary considerations when determining the look, feel and structure of the digital repository. Will records be accessed by school children or by scholarly researchers? Determine audience needs by reaching out and asking the audience how they will use the site and what they need from it. Accessibility for people with disabilities is also a consideration when designing a site.

As noted in the Choosing Software section, some of the open source software available provides standard themes that the site owner may choose from and customize. While this is easy for those with less website design skills, it also ensures that a site will look like many others. Familiarity with Cascading Style Sheets (CSS) will allow an archives to customize the basic options provided, and/or determine the branding of a site being built building through Dreamweaver or other web design tool.

The information in this chapter is extremely high level, but does give a starting point that can be followed up with the myriad of general web design information available online or in libraries.

Krause, M.G. & Yakel, E. (2007). Interaction in virtual archives: the Polar Bear Expedition Digital Collections next generation finding aid. *American Archivist*, 70, 282–314.

Another aspect of digitization to consider is the online presentation of digitized records to users in a way that allows accessibility and user interaction through Web 2.0 features. This article discusses a project out of the Bentley Historical Library, University of Michigan, which incorporated Web 2.0 tools and features into a digitized collection. The Polar Bear Expedition Digital Collection represents the records of American soldiers

who were stationed in Russia following the Russian Revolution in 1917.

The authors were inspired by socio-technological systems in everyday use, such as Flickr, as well as social software already being used by museums. The project group decided to incorporate the following features into their interactive digitized collection: bookmarking, commenting, link paths, browsing, searching and optional user profiles.

The common themes within the research conducted through the project were: accessibility (ability of users to make use of archival descriptions), common ground (the sense of shared place and agreement between users of a site), awareness (ability to see the presence of other users) and interactivity (to encourage informal social encounters whereby users support and help one another). The article discusses the need for two-way interaction between archivists and users at the site of the record, and transparency of archival processes through, for example, colophons and annotations—activities not traditionally undertaken in archives. Human-Computer Interaction research, specifically social navigation, is also discussed as relevant to the creation of an interactive finding aid.

The findings regarding user interaction in the site, while preliminary, could help archives to anticipate the ways users will interact with digitized material, and to determine which features are most used. The authors found that comments were more active than bookmarking, with many of the comments categorized as information correction, while others fell into the category of information sharing. The browsing function was also highly important to users, and accessed far more than search. This may highlight the need for a search function that achieves the results users expect,

every time they utilize the function.

The main research question discussed in this paper was around finding aid accessibility and the authors concluded, “Place-making within the Polar Bear Expedition site is a good way of increasing accessibility. The proximity of digital surrogates of original materials enhances remote access, and the intuitive categorization of the materials makes them more accessible to users. These features appear to increase accessibility the most,” (p. 309). Again, while an early attempt, this project does, as the authors note, provide a glimpse of a future where social tools are and will be a normal expectation of users of digital archival collections.

Related resource:

Svensson, M. and Höök, K. (2003). Social navigation of food recipes: designing Kalas. In Höök, K, Benyon, D. & Munro, A. (Eds.) *Designing Information Spaces: The Social Navigation Approach* (pp. 201-222. London: Springer.

The implementers of the Polar Bear study noted that they took inspiration from this chapter.

Lampert, C. & Chung, S.K. (2011). Strategic planning for sustaining user-generated content in digital collections. *Journal of Library Innovation*, 2(2), 74-93.

Moving beyond the initial efforts to integrate user participation into a digital collection as described by Krause and Yakel (2007), Lampert and Chung ask the question: what needs to happen at a strategic organizational level after an enthusiastic librarian or archivist launches a trial social media project (including one attached to a digital collection)? The introduction of Web 2.0 initiatives affects a number of areas of an organization (administration, systems/technology, middle management,

communications, and assessment), and as noted by Eden (2001), teamwork is required by multiple players to manage and resolve issues regarding digital initiatives.

The provision of content/comments on digitized objects by users adds a new layer of management and communication tasks—in terms of responding to comments and mediating user discussion. An organization must determine response roles and guidelines for interacting with users. Further, a digital project manager needs to be tasked with making strategic decisions about digital collections, addressing errors pointed out by users and generally ensuring the high quality of the digital archival collection and its descriptions. Further work is required around web interface design, integration of user-generated metadata and promotion of the interactive site to users.

This article is a worthy read as it highlights the overlap between digitization and socio-technological tools within an archives, and the impact on organizational strategy. The difference of four years since the Krause and Yakel (2007) article shows that thinking has moved from experimental to strategic. The authors note that, “Users will be willing to contribute to a site if they receive responses quickly and they see other users are contributing content as well,” (p. 90) and “ideally, during the planning process for any new project that encompasses Web 2.0, key decision points would be clearly defined and a process would be created to document the decision and rationale,” (p. 90).

The authors conclude with five key points for integration of Web 2.0: encompass new user expectations in organizational philosophy, reflect commitment in the strategic

plan, acknowledge that Web 2.0 tools will impact a range of staff, integrate new roles into workloads, and evaluate and assess Web 2.0 projects.

Legal Issues (Copyright)

Copyright is one of the main factors in an archives determination of whether it will select particular material for digitization. When digitizing, close attention must be paid to copyright law, fair use guidelines and other exceptions that allow use of material by the archives. Archives should look at the terms of donor agreements and determine if permissions must be obtained from the donor/copyright holder before displaying the material online. The costs or time involved in obtaining permissions may help make the decision of whether to attempt the digitization of those records, or whether to focus on records where the archives clearly holds copyright or where material has passed into the public domain. Obviously privacy considerations and sensitivities are also part of the decision-making process. Finally, there is an overlap between metadata and legal issues, as any metadata schema must provide clear information, preserved for the long-term, about the rights status of a record.

Bill C-11: Copyright Modernization Act. (2011). 1st reading, Sept. 29, 2011, 41st Parliament, 1st session. Retrieved from: <http://www.parl.gc.ca/HousePublications/Publication.aspx?Docid=5144516&File=9>

Given Royal Assent after passage through the Senate on June 29, 2012, this Bill updates Canadian copyright law to address issues in the digital world. It will not come

into affect for several months as it must go through an order-in-council process and be translated into regulation.

On the positive side, “fair dealing”, or exceptions allowing for the user’s right to use copyrighted material, has been expanded to include the new categories of education (in a structured setting), parody and satire. Users may also now use a publically available work to create a new work in a non-commercial context. It also allows users to “format-shift” or copy a work during the course of private study, “time-shift” to for example “PVR” something for later viewing, and make back-up copies of material.

Clauses 28 to 30 apply to libraries, archives and museums. Clause 28 allows a LAM to copy a work in its permanent collection in an alternative form in advance of the earlier form becoming technologically obsolete. Clause 30 applies to unpublished works in archives, (a) only allowing the archives to make a single copy of a record for a user and “(b) the archive informs the person that the copy is to be used solely for research or private study and that any use of the copy for a purpose other than research or private study may require the authorization of the copyright owner of the work in question (Bill C-11, 2011). This addresses a long-standing issue for archives (CCA, 2011).

Of concern, the Bill repeals the exceptional treatment of photographs, making the photographer and not the commissioner of the photograph the copyright holder. With many photographs in archival holdings “orphan works”, this makes it even more difficult for archives to determine copyright (CCA, 2011) and by extension to be able to digitize and display records online. The Canadian Council of Archives noted in its 2011

Brief from the Canadian Council of Archives (CCA) To The Legislative Committee on Bill C-32 (CC32) that this issue of orphan works of material in all formats severely limits the ability of archives to digitize records for online access.

The area of greatest concern from many quarters are the provisions around technological protection measures, or “digital locks” on copyrighted material, which have been criticized for not properly balancing user and creator rights. If there is a digital lock on a copyrighted item, its existence trumps the fair use provisions, meaning that archives cannot break a digital lock in order to ensure preservation of a record. The bill in its current form does not, therefore, fully support archives in making Canadian heritage available digitally, and time will tell whether pressure will bring modifications to this new law upon implementation.

Related resources:

Canadian Council of Archives (2011). *Brief from the Canadian Council of Archives (CCA) To The Legislative Committee on Bill C-32 (CC32)*. Retrieved from http://www.cdncouncilarchives.ca/copyright/BillC-32Brief_Jan2010_CdnCouncillofArchives_EN.pdf

Description of issues affecting archives within the updated copyright law.

Copyright bill finally clears Commons (2012, June 19), *CBC News*. Retrieved from: <http://www.cbc.ca/news/politics/story/2012/06/19/pol-copyright-bill-passed-commons.html>

Update on the recent passage through the House of Commons of controversial amendments to Canadian copyright law.

Murray, L.J. & Troscow, S.E. (2007). *Canadian copyright: a citizen's guide*. Toronto: Between the Lines.

This 2007 guide provides recent background on Canadian copyright law for the lay reader. One of the authors is a law academic while the other an English academic, bringing two complementary perspectives to the discussion, and writing with awareness of the complicated and challenged current copyright environment. The book provides history and contextual information, explains the scope of copyright, owner's rights, user's rights and then law in action. Of particular interest will be the chapters on digital rights management explaining content locks, and the chapter on libraries and museums explaining the increasing tension between "LAMs" (library, archives, museums) and copyright owners in the digitized environment. However, this chapter does not provide any detailed information regarding archives or archival digitization. As a general and accessible guide on Canadian copyright, though, this book is a recent and useful resource.

Related resources:

Hutchinson, T. (2011). Copyright and cultural institutions: guidelines for digitization for U.S. libraries, archives, and museums. *Archivaria* (71), 150-152.

This article is a book review of the above titled book. Hutchinson notes that some general aspects of the discussion may be useful to non-Americans.

Marelli, M. (2012). *Copyright Committee Information Bulletins, Canadian Council of Archives*. Retrieved from http://www.cdncouncilarchives.ca/copycomm_info.html

This series of brief bulletins provides working archivists with basic, Canadian focused information on issues that may be encountered in a digitization project. These include: putting copyright material on a website, how to get copyright permissions and the meaning of "public domain".

Noel, W. (1999). *Staff guide to copyright: National Archives of Canada*. Ottawa: National Archives of Canada, 1999.

While US resources on copyright are plentiful, there are enough differences between Canadian and American law to require a Canadian archivist to consult unique material when undergoing a digitization project. This 1999 guide is a notable Canadian manual on the subject (Hutchinson, 2011), but does require updates to account for changes in law. Created for the staff of the then National Archives of Canada, Noel provides detailed information on the range of copyright issues specific to archives: rules of protection, and their application to audiovisual material and sound recordings, cartographic, architectural and engineering drawings, manuscripts, art and photographs, and government archives. Again, while the manual is clear and written specifically for archives, it does not reflect *Bill C-11* (2011). Hopefully it will be updated given the significant changes to copyright law recently passed.

Other resources

Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2012). Bibliography. Retrieved from <http://brtf.sdsc.edu/bibliography.html>

Canadian Heritage Information Network (2012). Professional Exchange – Digitization. Retrieved from: <http://www.pro.rcip-chin.gc.ca/ressources-resources/index-eng.jsp?No=0&Ne=8109&N=8109&Qo=50>

Federal Agencies Digitization Guidelines Initiative (2009). Digital conversion – documents and guidelines: a bibliographic reference. Retrieved from <http://www.digitizationguidelines.gov/still-image/digconv.html>

Kenney, A.R. & Rieger, O.Y. (Eds.) (2000). *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group, 2000.

Lapotin, L. (2006). Library digitization projects, issues and guidelines: A survey of the literature. *Library High Tech*, 27(2), 273-289.

TechSoup (2012). Website. Retrieved from <http://home.techsoup.org/pages/default.aspx>