

January 2013

Metadata and Lasting Collaborative Success

Felicia J. Williamson

Sam Houston State University

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/provenance>



Part of the [Archival Science Commons](#)

Recommended Citation

Williamson, Felicia J., "Metadata and Lasting Collaborative Success," *Provenance, Journal of the Society of Georgia Archivists* 31 no. 2 (2013).

Available at: <https://digitalcommons.kennesaw.edu/provenance/vol31/iss2/6>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Provenance, Journal of the Society of Georgia Archivists by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

Metadata and Lasting Collaborative Success

Cover Page Footnote

Many thanks to my graduate advisor Dr. Elizabeth Dow, who is a mentor and a dear friend. Thanks to Tom Dillard and Tim Nutt at the University of Arkansas who gave me my start in the field and still encourage me from afar. Thank you also to my husband and colleague -- James Williamson -- who helps me understand metadata better every day. Many thanks to my editor and library school chum, Beth Colvin. Thanks also to my team: Cheryl Spencer, Trent Shotwell and Shaneil Snipe for keeping Thomason Special Collections running smoothly while I have my nose stuck in a book.

Metadata and LAMs: Lasting Collaborative Success

Felicia J. Williamson

“Collaboration brings new users to collections.”¹

Introduction

As Muriel Foulonneau writes, “at the heart of collaboration lies the harmonization of collections and services.” As more and more material becomes available through cultural heritage institutions, it has become part of many institutions’ mission to make these materials available online. Indeed, “the ubiquity of online access inspires a vision of a single search across *all* collections, without regard to where the assets are housed or what institutional unit oversees them.”² It is an expectation at many institutions to have online exhibits that coincide with physical exhibits. Moreover, it has become apparent that better access can be accomplished when institutions share information to reach their audiences.

In today’s information environment – where new users expect to access materials online – libraries, archives, and museums (LAMs) face external pressure to increase their web presence. For cultural heritage institutions – large, and especially small – the cost is daunting. Nonetheless, “by digitizing their collections, cultural heritage institutions can make information accessible that was previously only available to a select group of researchers.”³ This is a benefit that has drawn many a LAM to the precipice of a collaborative effort based on metadata interoperability. This article will discuss the use of metadata in

¹ Liz Bishoff, “The Collaboration Imperative,” *Library Journal* 129, no. 1 (January 2004): 34.

² Muriel Foulonneau and Jenn Riley, *Metadata for Digital Resources: Implementation, Systems Design and Interoperability* (Oxford: Chandos, 2008): 118; Diane Zorich, Günter Waibel and Ricky Erway, “Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives and Museums,” (Dublin, OH: OCLC Research, 2008), accessed June 10, 2013, <http://www.oclc.org/content/dam/research/publications/library/2008/2008-05.pdf>.

³ Ibid.

LAMs, focusing on best practices resulting from American attempts to utilize uniform metadata standards to collaborate and offer the best, comprehensive access to materials in LAMs.

Metadata

The most common definition of metadata is that it is “data about data” – another way to understand metadata is that it is all the information necessary to identify and retrieve a digital object. Historically, catalog records, finding aids, and museum artifact descriptions have formed the metadata backbones of LAMs. Thus, “good metadata makes it possible to catalog and effectively present digital information to the public.” For metadata to be good, it must describe many aspects of the original object, whether born digital or not. Significantly, many metadata schema are currently in use and there is no single metadata scheme that is prevailing – the result is that a collaborative effort will often include multiple metadata schema that have to be reconciled.⁴ To collaborate effectively, LAMs must grapple with this and many other complex technical issues. Good metadata, whatever the final conclusion, is key to collaborative success.

At the most basic level, metadata allows LAMs to keep track of materials for both their own institutional needs and for resource sharing or collaboration. At its best “metadata allows various functions to be performed on digital resources, for example, discovery, interpretation, preservation, management, presentation and re-use of objects.” For metadata to allow for discovery, interpretation, and preservation and so on and also be functional across institutions, the metadata must be interoperable. “Interoperability, at its most basic level, is the ability of different systems to talk to each other.”⁵ If metadata does not transfer well from one system to another, it will either decrease the effectiveness of a collaborative effort, or in a worst case scenario force the collapse of the collaboration altogether. Indeed, as the following discussion of collaborative success will show – metadata interoperability is the cornerstone of a successful project.

⁴ Trevor Jones, “An Introduction to Digital Projects for Libraries, Museums and Archives,” <http://images.library.uiuc.edu/resources/introduction.htm>.

⁵ Foulonneau and Riley, *Metadata for Digital Resources*, 6, 119.

Dublin Core

Most collaborative projects utilize some form of the Dublin Core metadata element set. “The Dublin Core (aka the Dublin Core Metadata Element Set), created in 1995, is a set of fifteen generic elements for describing resources. These are: Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage, and Rights.” The Dublin Core was established at the outset of the internet era and has international reach. Significantly it informs the many metadata schema that have grown up in the archival field, including METS, MODS, etc. The Dublin Core describes “a wide range of networked resources ... by a cross-disciplinary group of professionals from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship.”⁶ The fact that a cross-disciplinary group created Dublin Core is perhaps foretelling of its use for LAM collaborations as inherently cross-disciplinary endeavors.

Diane Hillmann explains a concept that comes up but is often not explained in many of the collaborative project descriptions – the use of qualified versus unqualified Dublin Core elements. The Dublin Core has fifteen optional elements, all of which have a set of qualifiers which further identify that particular piece of metadata. Thus, the use of “qualified” Dublin Core metadata means applying elements that are more descriptive due to the use of these “qualifiers” while unqualified metadata use the elements in their original form. Earlier projects relied on unqualified metadata while more recent projects recommend the use of qualified elements.⁷

⁶ Diane Hillmann, “Dublin Core Metadata Initiative,” accessed November 26, 2010, <http://dublincore.org/documents/2001/04/12/usageguide/>; Carol Godby, Jeffrey A. Young and Eric Childress “A Repository of Metadata Crosswalks,” *D-Lib Magazine* 10, no. 12 (December 2004), accessed October 11, 2013 <http://www.dlib.org/dlib/december04/godby/12godby.html>; Hillmann, “Dublin Core Metadata Initiative.”

⁷ “The Dublin Core metadata elements fall into three groups that roughly indicate the type of information stored in them: (1) elements mainly to the Content of the resource, (2) elements related mainly to the resource as Intellectual Property, and (3) elements related mainly to the Instantiation of the resources... Content (Title, Subject, Description, Type, Source, Relation, Coverage), Intellectual Property (Creator, Publisher, Contributor, Rights) and

Further, Dublin Core is often built into crosswalks to enable metadata interoperability. As Katherine Timms writes, “because it [Dublin Core] can be commonly applied in all three cultural heritage sectors (libraries, archives and museums), it can also serve as the standard to which descriptions can be mapped using crosswalks for use in building integrated systems.”⁸ Thus, the core set of either qualified or unqualified Dublin Core elements are set up alongside either MARC or EAD or the legacy descriptive metadata standards used by the agencies involved in the collaboration. The crosswalk is put in place to link one common element to another from standard to standard, which allows for true descriptive depth and interoperability and has been shown to increase the usability, flexibility and worth of the metadata sharing operation. The reach of Dublin Core is expanded by implementing Open Archives Initiative Protocol for Metadata Harvesting, even though few institutions are taking this step.

LAM collaborations have the end goal that they will provide more content online for a wider audience. To do this, LAM collaborators are turning to new technology and have commonly relied on meta-mark-up to enable this functionality. “The most common way to associate metadata with web-accessible content is to embed the metadata in the identical object that it describes. If the object is an HTML document, metadata can be embedded by use of <meta> elements... the metadata can then be harvested and indexed by Internet search engines.”⁹ While this allows for in-depth access to collections, it also requires investment by the LAM collaborators to enrich their metadata through the use of standardized tagging. The long-term payoff is there, but there must be the drive to make this happen across departments and even across institutions. When evaluating true costs and benefits of a collaborative project, stakeholders should

Instantiation (Date, Format, Identifier, Language).” Sheila S. Intner, Susan S. Lazinger, and Jean Weihs, *Metadata and Its Impact on Libraries* (Westport, Conn.: Libraries Unlimited, 2005): 32-33.

⁸ Katherine V. Timms, “Arbitrary borders? New Partnerships for Cultural Heritage Siblings – Libraries, Archives and Museums: Creating Integrated Descriptive Systems” (M.Thesis, University of Manitoba, 2007): 108.

⁹ Priscilla Caplan, *Metadata Fundamentals for All Librarians*, (Chicago: American Library Association, 2003): 45.

keep this perimeter in mind. Further gain comes from implementing the Open Archives Initiative Protocol for Metadata Harvesting, though it requires an added level of planning and expertise.

Open Archives Initiative Protocol for Metadata Harvesting

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a system that enhances access to metadata for the purpose of sharing and thereby, increase interoperability. The OAI-PMH crawls xml-structured metadata produced by museums and archives, and streamlines the process for harvesting the metadata and producing search results in the web environment. To participate, a repository must sign up and “open” their metadata to the crawling process. Multiple sets or types of metadata records can be searched by the OAI-PMH as long as they are validated and adhere to XML structures. “The OAI ... stands for the Open Archives Initiative and seeks to develop and promote interoperability standards that facilitate the efficient dissemination of content.”¹⁰ The PMH takes the OAI several steps further. Once metadata meets a minimum standard, the harvester will collect it and return search results for a particular repository. It is, essentially, a metadata aggregator.¹¹

The strength of OAI-PMH is that it “allows OAI provider systems to serve up any metadata schema that can be validated against an available SML Schema Definition.” which facilitates a flexible, if complex, data combing structure for large quantity caches of metadata records. However, the fact that practitioners make decisions about “mapping metadata from one representation into unqualified Dublin Core” and then create crosswalks to existing metadata schema – for instance, EAD or MARC – which are then combed by OAI-PMH to produce web results explains how the theory of OAI-PMH becomes difficult to put into practice. Significantly, OAI-PMH may be of substantial use and applicability to those repositories which update their records and upload large batches of records often – this explains why OAI-

¹⁰ Intner, et. al., *Metadata and Its Impact on Libraries*, 54.

¹¹ Carl Lagoze, “The Open Archives Initiative Protocol for Metadata Harvesting,” accessed November 25, 2010, <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

PMH has been adopted by agencies like NASA.¹²

While these problems should be on the radar for any group of collaborators about to embark on a metadata project, Sheila Intner writes this summation:

“Although there has been progress toward a default global metadata standard – unqualified Dublin core – as well as toward a global meta-language in which to describe the digital objects of various communities – XML – and a metadata framework in which to wrap the multiplicity of metadata schema these communities created to describe these objects – RDF – implementing the OAI has shown, among other things, that the problem of interoperability still requires a variety of assessment activities to guide plans for the long-term sustainability of the services established.”¹³

Indeed, Hillman writes that “the flexibility and lack of precision inherent in simple DC also allow its inconsistent application. Our experience corroborates earlier work suggesting that ongoing efforts to map subject terminologies and harmonize ontologies are necessary to achieve a high level of functional interoperability.”¹⁴ The most successful, long-term collaborations built LAM-specific ontologies, metadata-crosswalks, and were able to adjust their technology to best serve retrieval needs.

Literature Review

The literature on metadata and collaborative projects within LAMs can be divided into two main subject areas: the technical issue of metadata and its use for LAM collaboration and specific

¹² Intner, et al. *Metadata and Its Impact on Libraries*, 55-56; Chu Churngwei, Walter E. Baskin, Juliet Z. Pao, and Michael L. Nelson, “OAI-PMH Architecture for the NASA Langley Research Center Atmospheric Science Data Center,” in ECDL Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries, 2006, accessed October 14, 2013, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.75.5304>.

¹³ Intner, et al. *Metadata and Its Impact on Libraries*, 55-56.

¹⁴ Diane Hillmann and Elaine L. Westbrooks, eds. *Metadata in Practice*, Chicago: American Library Association, 2004. 175.

metadata collaborative projects in American LAMs. LAM collaborative projects moved from relying on Dublin Core as a sole metadata standard to more complex technological applications. Priscilla Caplan provides a fundamental interpretation of metadata including excellent explanations of interoperability, controlled vocabularies, and syntax. Hillman, Foulonneau, and Trevor Jones¹⁵ take this fundamental understanding and apply it to more complex technologies and their application, explaining how the methods with which metadata is applied can enhance the long-term success of a collaborative project.

Throughout the literature, discussions of new approaches or technologies that can overcome the potential shortcomings of either Dublin Core¹⁶ or OAI-PMH¹⁷ emerge. Metadata crosswalks are a recurring theme as well as the need for federated searching: “Simultaneously searching multiple databases via a single interface or portal is known as federated searching or meta-searching.” There is a recurring interest or willingness to invest in the “development of high functioning federated search”¹⁸ capabilities. The needs of the end user drive technical innovation. Current researchers demand one-stop searching technology with an

¹⁵ Caplan, *Metadata Fundamentals for All Librarians*, 1-44; Hillmann and Westbrook, eds. *Metadata in Practice*, 20; Foulonneau and Riley. *Metadata for Digital Resources: Implementation, Systems Design and Interoperability*, 118; Jones. “An Introduction to Digital Projects for Libraries, Museums and Archives.”

¹⁶ Dublin Core Metadata Initiative, <http://www.dublincore.org/metadata-basics/>, accessed December 1, 2010. “Early Dublin Core workshops popularized the idea of “core metadata” for simple and generic resource descriptions. The fifteen-element “Dublin Core” achieved wide dissemination as part of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and has been ratified as IETF RFC 5013, ANSI/NISO Standard Z39.85-2007, and ISO Standard 15836:2009.”

¹⁷ Lagoze, “The Open Archives Initiative Protocol for Metadata Harvesting.” “The Open Archives Initiative Protocol for Metadata Harvesting” (referred to as the OAI-PMH in the remainder of this document) provides an application-independent interoperability framework based on *metadata harvesting*. There are two classes of participants in the OAI-PMH framework: *Data Providers* administer systems that support the OAI-PMH as a means of exposing metadata; and *Service Providers* use metadata harvested via the OAI-PMH as a basis for building value-added services.”

¹⁸ Timms, “Arbitrary borders?,” 99; Zorich, et. al., “Beyond the Silos of the LAMs,” 17.

intuitive interface, but the metadata infrastructure necessary for that sort of searchability requires substantial expertise.

In response to the changing needs of patrons in addition to shrinking budgets, more LAMs have turned to collaboration in the online environment. Thus, a second area in the literature focuses on collaborative projects in American LAMs. Many of these projects are IMLS funded and are meant to gather local or state-wide materials and provide increased access to materials through unified, searchable metadata. For an introduction to the basics of LAM collaboration including funding and patron expectations, see Jennifer Novia's work in *LAM Collaboration*. Novia explains that the ability to present online surrogates of the varied items in the collections of LAMs forced the issue of collaboration on to potential collaborative partners – and made the idea of sharing access in the online environment (as well as funding streams) seem not only possible but desirable. A recurring example of an ideal collaborative project is the Colorado Digitization Project, which is discussed in an article by Brenda Bailey-Hainer.¹⁹ This project is archetypal in many ways, but was phased out in 2010. As one of the first large collaborative digitization projects based on shared metadata and interoperability, the Colorado Digitization Program stood out as an example for other regional and intuitional collaborations that followed.

A current, successful statewide LAM collaborative is the Publication of Archival, Library & Museum Materials (PALMM)²⁰

¹⁹ Jennifer Novia, "Library, Archival and Museum (LAM) Collaboration: Driving Forces and Recent Trends," *Endnotes: The Journal of the New Members Round Table* 3, no. 1 (October 2012); Brenda Bailey-Hainer and Richard Urban, "The Colorado Digitization Program: A Collaboration Success Story," *Library Hi Tech* 22, no. 3 (2004): 254-262.

²⁰ "Publication of Archival, Library & Museum Materials (PALMM) is a cooperative initiative of the public universities of Florida to provide digital access to important source materials for research and scholarship. PALMM projects may involve a single university or may be collaborative efforts between a university and partners within or outside of the state university system. PALMM projects create high-quality virtual collections relevant to the students, research community and general citizenry of Florida." Publication of Archival, Library & Museum Materials (PALMM) (2012), accessed June 28, 2013, <http://palmm.fcla.edu/>.

project. This project, like a similar project in Texas – TARO²¹ – maintains a strong federated searching function that allows researchers to search across a multitude of state LAMs for materials through a simple online interface. PALMM is significant in that it presents a great deal of digitized content sourced from dozens of state agencies and repositories. It searches well and is easy to use and understand – and has incorporated interoperable metadata and a great deal of depth despite the diversity of source organizations. In contrast, TARO is an older project that simply searches online finding aids from participating institutions. TARO does not search digital images, and can only search the metadata of EAD finding aids – a limitation that excludes many potential institutional participants. Nevertheless, TARO provides searchable metadata for institutions across a large number of institutions and is easily searched.

There will likely be more projects like PALMM and TARO as regional organizations address the task of metadata unification as a group. Meanwhile, the next wave of U.S. collaborations are large institutional LAMs like the United States Holocaust Memorial Museum or the Smithsonian as well as university systems. Diane Zorich and her co-authors explain such projects in “Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives and Museums”²² in which the authors explain the movement of LAM administrators along a collaboration continuum as they work toward a unified search option. While online collaboration and increasingly flexible web environments make more resource sharing and online representation of collections possible, the need for communication and flexibility is evident. Historic, free-standing silos within the LAM community and within the metadata architecture make collaboration a challenge,

²¹ “TARO (Texas Archival Resources Online) makes descriptions of the rich archival, manuscript, and museum collections in repositories across the state available to the public. The site consists of the collection descriptions or ‘finding aids’ that archives, libraries, and museums create to assist users in locating information in their collections. Consider these an extended table of contents which describe unique materials only available at the individual repositories.” Texas Archival Resources Online, accessed June 28, 2013, <http://www.lib.utexas.edu/taro/about.html>.

²² Zorich, et al. “Beyond the Silos of the LAMs”: 10-16.

but the common goal of presenting collections online is a motivating force.

LAM Best Practices

First, the literature is clear in recommending that planners examine the needs of their user population and look at comparable projects – mining the literature for free advice before carefully choosing the metadata standard they will implement for the collaborative project. Indeed, while most of the literature mentions the use of Dublin Core as a basic template metadata scheme, recent articles are pushing for increased “technological and semantic interoperability.” As discussed above, to enhance interoperability LAMs will have to implement specific element structures based on a set of elements from the Dublin Core. Indeed, “stick to standards as much as possible, but if and when you diverge, document what has been done and why it was done.”²³ The current best practice is to tailor LAM-specific metadata set based on Dublin Core. Significantly, part of the lessons learned from other projects is that qualified Dublin Core might offer success for LAM collaborations.

Second, the use of a single metadata standard – Dublin Core – to map all other integral metadata records is a best practice. Successful LAMs take it further. “The dream of a single metadata standard is an illusion” and as such, “attempts to enhance consistency through the promotion of guidelines within communities and coordination across communities can be extremely valuable.” Thus, successful LAMs work through multilateral collaboration to encourage uniform application of the metadata elements that the institution itself deems most useful, and then the LAM sets up a structure to monitor and clean up the metadata records already in place. This enables the creation of uniform, good metadata from a variety of creator institutions or departments and, in the long-term, enhances interoperability. LAMs can take this even further if they are able to “anticipate future uses of your data.”²⁴

Third, it is important that any LAM collaboration take steps to build up the technical infrastructure that will allow for long term

²³ Timms, “Arbitrary borders?”: 96; Hillmann, *Metadata in Practice*, xvi.

²⁴ *Ibid*, xvi, 226.

success of a large technical undertaking, utilizing financial and human resources efficiently. To have technical infrastructure that will facilitate long-term success, collaborative partners should assess the state of their servers, choose a central management team and support staff and find a functional communication medium that works for all participants. Having the support and open communication lines with the IT department as well as the grant or departmental funding sources are two key elements for collaborative success.

Fourth, people matter. Like any team project, a LAM collaborative project is dependent on the people who work on the team. The complexities of a LAM collaboration demand flexibility and open-mindedness. “LAM professionals who understand issues surrounding different types of collections and collecting institutions, and who are not rigidly wedded to their own professional traditions, bring an open-mindedness that allows them to embrace ideas from other professions in the interests of the collaboration.”²⁵ Give and take will make or break a collaborative project.

It is imperative that a large, collaborative project involve the staff of all participating institutions or departments. Because staff members rather than department heads will often implement large projects on a day to day basis, their insights are invaluable. Moreover, if staff feel invested, their ongoing participation will increase. In addition, it is important to have a point person or people who are available and known to the program implementers. If those people are at the helm of a project and are either unavailable due to the demands of their other job duties or leave their position, the project will often fall on hard times. It is important to line up a trusted replacement and to always maintain open communication with all stakeholders. Transparency is important, as is the ability to ask questions and be confident that ideas, concerns and feedback will be heard and also responded to. Having a group email might be sufficient, as long as someone, or a group, take the responsibility to answer questions and concerns.

Finally, once the LAMs have put in so much planning and preparation, it is imperative to use the skills of great programmers

²⁵ Ibid, 27.

to produce an interface that allows for intuitive searching across collections. “One ideal feature of a landscape is that it should be transparent to the user. The professional and technical complications of collection versus item description and metadata format, content and aggregation should not be allowed to adversely affect the user’s interaction with the environment; their experience should be as seamless as possible.”²⁶ If the search interface helps the end user understand their results and increases the project’s visibility, it could help with ongoing sustainability through institutional buy-in and funding. Thus, a best practice for LAMs is to keep the end user in mind.

Conclusion

The issues of legacy metadata, institutional politics, and monetary and technical roadblocks are enough to discourage even the most ambitious information professional. However, the benefits to be gained from a successful collaboration are legion. Not only do new audiences gain access to collections, but an institution or set of partner institutions/departments, gain a much better understanding of, and thereby control over, metadata. This has lasting benefit to organizations and their patrons. By applying some best practices and spending more time planning and building an infrastructure that will last – collaborative partners can build online environments that facilitate research for wider audiences on a deeper level than was previously possible.

Felicia J. Williamson, MLIS, graduated with a BA in history, German and European studies and a minor in religious studies from the University of Arkansas at Fayetteville. She received a Master in Library and Information Science with an archives focus from LSU. Williamson is a certified archivist, the chair of the Professional Development Committee, Society of Southwest Archivists and is a member of the Walker

²⁶ Gordon Dunsire, “Future Information Environments: Deserts, Jungles or Parks?” (paper presented at Archives, Libraries, Museums 10 (AKM10), Porec, Croatia, 2006): 2, accessed December 1, 2010, <http://cdlr.strath.ac.uk/pubs/dunsireg/akm2006Futures.pdf>.

County Historical Commission as well as the Society of American Archivists. Since joining the faculty at SHSU's Newton Gresham Library as the new head of Thomason Special Collections in 2011, Williamson has instituted a program of instruction and outreach with the hopes of making its archival holdings more accessible to the campus and surrounding community.