METAWARE.BUZZ ARCHIVE

Created April 12, 2019 from the metaware.buzz site. Posts are listed in increasing chronological order

Private: Welcome to Metadata Standards (June 30, 2014)

BY ERIK · PUBLISHED JUNE 30, 2014 · UPDATED OCTOBER 15, 2014

Project Scope:

The Metadata Standards public information site's goal is to inform and engage a the metadata community through resource sharing, analysis and discussion. The broad focus on metadata includes standards, best practices, current trends, and innovative examples the sharing of which is intended to advance research and practice in the metadata community. Our primary community includes

practitioners and researchers in the library, archive, museum and school communities. The platform will, if possible, be hosted on an ALA site and should employ a technology platform that encourages distributed editing and easy community engagement. The DH+LIB community site is used for inspiration in design, content negotiation and editorship.

The site will be curated/edited first by MSC members and later by the wider metadata community. This may include contributors from the Metadata Interest Group, the CAMMS Education group(s), and the Linked Data Interest Group as well as other community members. Engaging library school students may be another avenue for contributors (but likely not editor/curator activities)

The scope of the site includes republishing of content (e.g. gathering of tweets or news items/websites) for posting. In-depth articles or analysis of content or topics are acceptable and may even be a part of an outreach effort.

Twitter handle @mdatastandards reserved 'just in case'

editors pick who on twitter to follow

Project timeline:

1. December 2013 – January 2014: Explore DH+Lib site in detail and create a shared concept

2. January 26, 2014: Discuss community engagement site with ALA community at ALA Midwinter

3. February 2014: Explore logistics:

    1. Done – Document shared scope statement: Erik

    2. Research hosting options: Jenn, Eva, ?

        1. Erik will host based on hostmonster

    3. Begin considering design issues: Bonnie

4. March-May 2014: Design and implement technical and editorial structures

5. June 2014: Launch of site with

6. July 2014: Presentation of site at ALA Annual

7. Vote on domain name: July 15

8. Setup base site and have report back on what needs to be done – July 31

9. Soft Launch: August 15th

Building a Metadata Community (October 15, 2014)

BY [ANDREW WEIDNER](#) · PUBLISHED OCTOBER 15, 2014 · UPDATED MARCH 8, 2015

I recently returned from the [2014 International Conference on Dublin Core and Metadata Applications](#) in Austin, Texas, where I had the opportunity to mingle with a distinguished group of metadata minds. As a relatively new professional in the field, I was able to participate in the [Next Generation Metadata Specialist Program](#) that brought new professionals to DCMI to learn more about the history of Dublin Core as a metadata standard from those who helped shape it from the start.

The morning special session on the first day was a highlight of the conference as DCMI veterans took turns sharing their stories about the early days of metadata for the Web and how Dublin Core took shape in response to the need for a lightweight standard that could accommodate a variety of use cases. A common theme from those stories was that developing a standard is largely about conversations that take place within a community, often in conference center hallways and dimly lit bars; and that process is much more productive when the community is a friendly place that welcomes new ideas.

Over time, that is what we hope to accomplish with this project, [metaware.buzz](#), by providing a platform for the exchange of ideas about the theory and practice of metadata. We encourage you all to [get involved](#) and join the conversation!

Andrew Weidner [@metaweidner](#)

Linked Data for Libraries, Archives and Museums (October 21, 2014)

BY [ANDREW WEIDNER](#) · PUBLISHED OCTOBER 21, 2014 · UPDATED MARCH 6, 2015
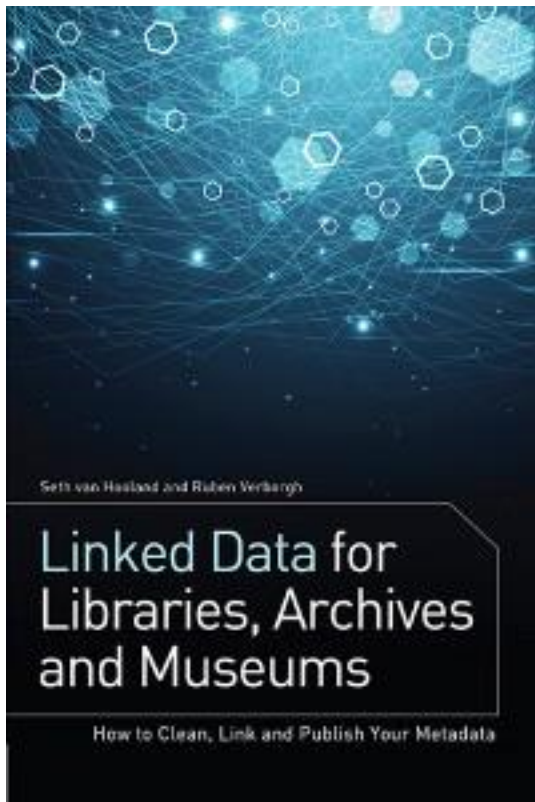
In [Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata](#),
Seth van Hooland and Ruben Verborgh provide an effective manual for those who wish to understand
and apply linked data principles to metadata for digitized cultural heritage collections. The book's five
main chapters cover the activities that must be undertaken in order to publish semantically rich
metadata on the Web: modelling, cleaning, reconciling, enriching and publishing. Each chapter ends
with a case study that describes the practical steps accomplished by a particular institution toward that
chapter's topical focus and introduces readers to a variety of tools and techniques useful for creating
linked data. In addition, the authors' companion Web site ([freeyourmetadata.org](#)) provides access to
numerous tools and data sets for experimenting and working with linked data.

The early chapters provide solid guidance for anyone currently creating metadata for digitized cultural
heritage objects. The later, more technical chapters provide a road map for the difficult task of
publishing linked data. Throughout the handbook, the authors reiterate that creating and maintaining
linked data is a difficult task, though not without its rewards. Practitioners in the field have yet to reach a
consensus on the best technologies and methods for publishing and consuming linked data. As such, it
can be challenging to choose an appropriate data model, serialization format, and delivery method for
linked data, let alone clean, reconcile and enrich existing legacy data in order to bring it into the linked
data environment. Van Hooland and Verborgh provide an accessible and useful road map for making
intelligent decisions about how to best create and publish linked data for cultural heritage collections.

Andrew Weidner [@metaweidner](#)

The full version of this review appears in the [Journal of Digital Media Management](#), Vol. 3, No. 3.

BIBFRAME and RDF vocabulary reuse (November 25, 2014)

BY <u>JENN RILEY</u> · NOVEMBER 25, 2014

The conversations surrounding BIBFRAME can be dizzying with unfamiliar terminology, questions posed in areas the library community has little experience with, and our community thinking aloud and learning as we go. These discussions and issues are deep, with many competing perspectives. An <u>announcement this week by the National Library of Medicine (NLM) regarding their BIBFRAME testing</u> highlights one of these issues: whether BIBFRAME should define all of the properties and classes (i.e., the whole metadata structure) needed for resource description in libraries in its own namespace, or define a core set in which libraries have particular expertise but rely on other specialist communities for other parts of the vocabulary.

Back in early 2013 this issue was <u>first discussed</u> on the BIBFRAME list. As part of that thread, Eric Miller from <u>Zepheira</u>, the company contracted to design the RDF model for BIBFRAME, <u>stated</u> the initiative's intentions were to take the former (define everything themselves) approach: "While the recommendation of a singular namespace is counter to several current Linked Data bibliographic efforts, it is crucial to clarify responsibility and authority behind the schematic framework of BIBFRAME in order to minimize confusion and reduce the complexity of the resulting data formats." With this approach, connections to identical or similar concepts in other RDF vocabularies can be made through mechanisms such as OWL's <u>sameAs</u> property. A vocabulary designed this way is slightly easier to implement on its own, but is more difficult for machines to process and perform inferences on, and is to some degree less likely to be used by other communities.

Questions have been posed from the community about this decision for the direction of BIBFRAME. One particularly cogent analysis of a number of related BIBFRAME issues comes from Rob Sanderson in a <u>discussion document</u> first released in the summer of 2014. While the issue of vocabulary reuse is not in and of itself the focus of Sanderson's analysis, it underscores many of his points. For example, when describing what he sees as unnecessary complexity in the BIBFRAME model that he describes as "predicate proliferation," Sanderson states "[t]he proliferation is made worse by not reusing predicates that could be reused from other ontologies." In Sanderson's analysis, not reusing vocabularies from other sources seems to be a symptom of what he sees as other problematic modeling practices within BIBFRAME.

Most recently, the issue of a single BIBFRAME namespace vs relying on specialist communities for parts of the vocabulary has been raised again through <u>NLM's November 21 announcement </u>regarding their future direction for BIBFRAME testing. In their post to the BIBFRAME list with this announcement, NLM expresses unease about aspects of BIBFRAME modeling: "…as NLM has experimented with BIBFRAME over the past several years, we are increasingly concerned that the vocabulary development, in attempting to become sufficiently aligned with traditional bibliographic cataloging, may hinder meeting all of BIBFRAME's goals, particularly those of flexibility and extensibility." The lack of reuse of outside vocabularies and the complexity of the BIBFRAME model that results is a particular area of concern for NLM. Attempting to reduce this complexity is the core of NLM's approached way forward: "We intend to draft a core BIBFRAME vocabulary for experimentation (we fully understand that a workable core vocabulary will require collaboration from many communities, but we need to start with

something) and extend it with RDA (using the RDA Registry Elements) and an NLM vocabulary for local data."

Vocabulary re-use is one of many issues central to the design of BIBFRAME, and like other of these issues, is still a topic of debate. It is an example of the tension between following traditional library models (including ease of moving legacy data forward) and adopting information models and practices from outside of libraries with the goal of leveraging work designed for the web and integrating library data into the broader information space. NLM's testing announcement reminds us that optimum place on this continuum for BIBFRAME, and library metadata in general, is far from resolved within our community.

DRAFT Checklist for Evaluating Metadata Standards (January 20, 2015)

BY [JENNIFER LISS](#) · PUBLISHED JANUARY 20, 2015 · UPDATED AUGUST 31, 2015

The [ALCTS/LITA Metadata Standards Committee](#) submits this draft document, Checklist for Evaluating Metadata Standards, to the library, archives, and museums metadata communities for discussion.

The committee will discuss this document at our next meeting at the American Library Association Midwinter Conference in Chicago on Sunday, February 1, 2015 from 1:00pm to 2:30pm in McCormick Place West, Room W194a. Registered conference attendees are welcome to attend.

We encourage those not attending the conference to comment on this post or send private comments via our [webform](#).

This checklist is intended for use by libraries, archives, and museum (LAM) communities for the development, revision, use, and assessment of metadata standards.

1. The future of metadata is in the network

Metadata, its standards, systems, and services, are most efficient when connected with the broader network of information. This requires openness, automation, computation, web-design, and responsivity. These features are the new value proposition for metadata creation and management in LAM institutions.

2. Metadata should only be created where there is value

For too long, LAM institutions have relied on a network of professionals to create metadata "just-in-case" it is needed. The ability of information systems to analyze and index digital objects themselves changes the value of traditional metadata. Heavyweight standards are paralyzing rather than transforming LAM information services.

3. Metadata and metadata standards should be open and re-usable

Open metadata is a foundational building block of information systems and next-generation research. Metadata standards, associated vocabularies, and the metadata records themselves need to be open for use and re-use.

4. New metadata standards should support new research methods

Traditional bibliographic metadata supports a narrow vein of research. As new research methods emerge (e.g., computational linguistics, computational bibliometrics, linguistic analysis, network analysis) metadata standards and exchange/access methods should support this new research.

5. A metadata schema without a maintenance community is of little enduring value

Metadata schema are only as valuable and current as its community of practice. Communities of practice are changing rapidly and, although metadata in LAM institutions have been very stable over the last 40 years, that is not necessarily the case for future standards.

6. Metadata standards of the future should be web-enabled by default

Newly adopted standards should leverage the web, to connect information from different sources, to support distribution to indexing and research services, and to support resource visibility on the web. "Web-enabled" standards leverage the building blocks of linked data by using HTTP URIs to reference objects, by publishing metadata-rich information, by using RESTful design approaches and by adding to data already available on the web.

7. Standards should be extendable with properties/classes/elements from other communities/standards

A key enabler in the successful deployment of a metadata standard is its ability to work with schemas and vocabularies from multiple communities. It is important that these schemas and vocabularies enjoy "first-class" status, in that their incorporation into a record or broader standard respects issues of granularity and specificity.

8. Standards should be applicable to multiple communities and support selective adoption

Metadata standards should follow a "plug-in" architecture in which they enable adopters to adjust the standard to the needs of their local community. "All encompassing" standards do not scale well over time and have difficulty engaging new adopters. As such, metadata standards should feature a "lite" implementation or should be narrow enough in scope to allow expert comprehension. Common libraries, archives, and museums (LAM) standards should not incorporate these details as a central part of their design.

9. Standards should support aggregation, exchange, automation, and computational analysis

The use of literal over referential metadata has created an environment in which LAM communities cannot easily aggregate metadata without considerable attention to normalization, disambiguation, and record unification. New standards should leverage "by-reference" models by default and should create a new web or cloud of metadata in which aggregation, exchange, and computational processing are core and easily accomplished tasks.

10. Metadata schema should follow the rules of "graceful degradation" and "responsive design"

Schema should support lossless and lossy interoperability with other standards and should be well described enough to support new and unexpected uses. Metadata should gracefully degrade by easily supporting simple or limited use scenarios. In addition, metadata should support automatic "up-sampling" so that more sophisticated uses (e.g., ontological inferencing) are easily achievable.

Tags: standards assessmentstandards development

2015 LITA Forum – Call for Proposals (February 11, 2015)

BY PARKS@UP.EDU · PUBLISHED FEBRUARY 11, 2015 · UPDATED MARCH 1, 2015

The LITA Forum Committee welcomes creative program proposals related to all types of libraries: public, school, academic, government, special, and corporate.

— LITA Blog

The 2015 LITA Forum Committee seeks proposals for excellent pre-conferences, concurrent sessions, and poster sessions for the 18th annual Forum of the Library Information and Technology Association, to be held in Minneapolis Minnesota, November. 12-15, 2015 at the Hyatt Regency Minneapolis. This year will feature additional programming in collaboration with LLAMA, the Library Leadership & Management Association.

The Forum Committee welcomes creative program proposals related to all types of libraries: public, school, academic, government, special, and corporate.

Proposals could relate to any of the following topics:

• Cooperation & collaboration

• Scalability and sustainability of library services and tools

• Researcher information networks

• Practical applications of linked data

• Large- and small-scale resource sharing

• User experience & users

• Library spaces (virtual or physical)

• "Big Data" — work in discovery, preservation, or documentation

• Data driven libraries or related assessment projects

• Management of technology in libraries

• Anything else that relates to library information technology

Proposals may cover projects, plans, ideas, or recent discoveries. We accept proposals on any aspect of library and information technology, even if not covered by the above list. The committee particularly invites submissions from first time presenters, library school students, and individuals from diverse backgrounds. Submit your proposal through http://bit.ly/lita-2015-proposal by February 28, 2015.

Presentations must have a technological focus and pertain to libraries. Presentations that incorporate audience participation are encouraged. The format of the presentations may include single- or multi-speaker formats, panel discussions, moderated discussions, case studies and/or demonstrations of projects.

Vendors wishing to submit a proposal should partner with a library representative who is testing/using the product.

Presenters will submit draft presentation slides and/or handouts on ALA Connect in advance of the Forum and will submit final presentation slides or electronic content (video, audio, etc.) to be made available on the web site following the event. Presenters are expected to register and participate in the Forum as attendees; discounted registration will be offered.

Please submit your proposal through http://bit.ly/lita-2015-proposal

More information about LITA is available from the LITA website, Facebook and Twitter.

Source: 2015 LITA Forum – Call for Proposals

Tags: LITALLAMA

ALA Midwinter 2015 LITA Preconference Review: Introduction to Practical Programming (Draft - never posted) (February 12, 2015)

BY PARKS@UP.EDU · PUBLISHED APRIL 12, 2019 · UPDATED FEBRUARY 12, 2015

Editor's note: This is a guest post by Anthony Wright de Hernandez

The Friday before Midwinter officially started, I attended the LITA preconference session Introduction to Practical Programming. As a first-time conference attendee with SQL, XML, PHP, HTML, and Visual Basic experience, I wasn't sure exactly what to expect from a session that encouraged attendance by participants with no programming background. I chose to attend because I want to learn Python and thought this session would provide a good introduction to the language.

The Instructor

Elizabeth Wickes, a graduate student at the University of Illinois at Urbana-Champaign, clearly knows programming in general and Python in particular. Her instructional style for this session was conversational and informative. Her passion and knowledge kept the daylong session engaging. The mix of basic programming information with Python-specific information ensured that no part of the day was wasted for anyone.

The Session

The session began with a brief overview of computing and programming languages. This was a great place to start for a class with a mixed level of experience. As someone familiar with programming, this provided a background for where Python fits in relation to other languages, why it was created, and how its general mechanics differ from other languages. For those with no programming experience, this overview gave a brief history of programming and included a fun introduction to the type of logical and literal thinking required when programming.

After the overview, we dove right in with an explanation of Python's core data types. Again, the content was presented for mixed consumption. The data type explanations were basic and clear enough for beginners while those with more experience could remain engaged learning the mechanics of how Python interacts with each of the data types.

We had some hands on fun with Python by creating Mad Libs involving Q, from Star Trek, a list of colors, and some randomizing functions. Those of us who brought computers were able to try the code ourselves while Elizabeth demoed it on a screen for the rest of the attendees. Our quick coding exercise resulted in fun outputs like:

Q asked me, "So what kind of pythons do you want?"

I don't know what kind of pythons I want! Who wants 4 pink pythons?

So I just said, "Give me whatever kind of pink pythons you have in stock, Q"

One great thing about the session was that Elizabeth took on our specific challenges. We all had an opportunity to present the challenges we are facing at work and then get specific feedback on how to create a solution using Python. For example, one of the attendees needed a way to compare two lists of 40,000+ items and identify any items in one list that aren't in the other. Elizabeth walked us through how to develop a Python script capable of doing the comparison and returning the desired results. There was some great practical demonstration during this part of the session but, sadly, there were only a few of us in attendance so we didn't get to see the variety of applications that a larger pool of challenges would have provided.

Further Study

Of course, a single day session isn't enough to become a master. At the end of the session, Elizabeth provided us with recommendations for further study, including:

[Her guided self-study lesson plan](#)

[The learnpython reddit (/r/learnpython)](#)

[PythonLearn](#)

[pyvideo.org](#)

Overall (for beginning programmers)

The session was well structured for beginners. There was no assumption of prior programming experience. Basic concepts were introduced smoothly and then built upon to bring beginners to a point where they could create something of practical use. Strategies were provided for researching answers to programming questions and specific recommendations for further learning were given.

Overall (for experienced programmers)

The session was a great introduction to Python. It was definitely designed for all experience levels but, as an experienced programmer, I didn't find any section a waste. As a way to start learning Python, this session was great value. I got a basic foundation for the language and expert guidance on where to look as I continue my learning.

Anthony Wright de Hernandez is a recent graduate from the University of Washington iSchool. He is the appointed librarian for his local church and is currently seeking employment in academic libraries. You can learn more at his website: [anthonywright.me](#).

Source: [ALA Midwinter 2015 LITA Preconference Review: Introduction to Practical Programming](#)

Discussion notes: Draft Checklist for Evaluating Metadata Standards

BY JENNIFER LISS · PUBLISHED MARCH 1, 2015 · UPDATED MARCH 1, 2015

[The committee discussed the draft document, Checklist for Evaluating Metadata Standards](#), at its American Library Association Midwinter Conference meeting in Chicago on Sunday, February 1, 2015. We would like to thank the guests who attended and contributed their comments–particularly given the inclement weather!

Below is a summary of the discussion of this document. The Metadata Standards Committee will continue taking comments on this draft through April 15th, 2015. [Provide your feedback](#).

Document Scope

- Document is good but perhaps too high-level to serve as a checklist; may be more useful as a declaration or manifesto

- Document is useful for providing background information on the environment in which metadata exists

- Authors should elaborate upon the document's intended usage

Document Clarity

- Definitions of terms are needed

- Document needs more specificity in areas

- Might include a discussion (via a link, accordion box, etc.) of each of the ten existing points

Comments on Specific Checklist Points (by number)

- Checklist preamble: might add that Checklist may be used for selection (in an LD environment), creation, maintenance, and governance

- Checklist #1 and #6: are related; combine, refine, or at least move next to one another

- Checklist #2 and #8: are related; same as above

- Checklist #3: does the committee interpret the word "open" to mean 'free to read'? Guests pointed out that some free to read content standards (the DCRM suite was cited) supplement a standard the exists behind a paywall

- Checklist #3 and #5: Does a standard being freely available mean that the standard isn't guaranteed to be maintained?

- Checklist #5: state that maintenance community must respond in a timely fashion; switch bolded text to a positive statement; is there something to add here about sustainable business models for maintaining standards?

- Checklist #6: explain/link to resources on "RESTful design approaches"; if document stays high-level/manifesto-oriented, identifying specific technology/model/protocol may make it more difficult to keep the document up to date

Suggested Document Use

- Could the committee create a table in which metadata standards are evaluated against this Checklist?

- Use Checklist to justify decisions to administrators, vendors, etc. (this was said in the context of the open=free discussion)

Perceived Community Needs

- Need for a resource that shows how standards interact

- Need for how to select standards; a list of criteria for decision making

Standardizing Metadata for Digital Humanities (May 18, 2015)

BY CAROLYN HANSEN · PUBLISHED MAY 18, 2015 · UPDATED JULY 13, 2015

Digital Humanities (DH) applications such as databases, digital editions, and data visualizations provide users with the opportunity to search and curate datasets in new and interesting ways. By using the power of computing technologies, DH applications can uncover patterns in data that shed light on previously untold stories. In order for these applications to be successful, they require high quality metadata that is based on standardization and consistency. However, the historical and literary documents that make up the datasets for these applications are often messy, ambiguous, and varied. As metadata specialists, how do we aid the DH community in creating metadata standards that maintain the authenticity and spirit of original datasets while providing enough standardization for DH applications to be successful?

As a Metadata Librarian who also works in DH, I have been struggling with this question. For example, my colleague James Van Mil and I are creating a database of intake recordsfor the University of Cincinnati's House of Refuge Collection. The database consists of over 6,000 child intake records from the 19th and early 20th centuries. The records provide rich description about the admitted children, such as ethnicity, religion, offenses committed, and location of birth. Creating index terms for this dataset and consolidating terms has been difficult, because doing so requires making assumptions about the data that may not be consistent with the historical context in which it was created.

For example, there are multiple terms in the original dataset that refer to children of Jewish ancestry, including: "German Jew"; "Hebrew"; "Israelite"; and "Jewish." Consolidating these terms under a broader term such as "Jewish" would be helpful for indexing purposes, but it might also lead users to make false conclusions about the data. In order to make indexing decisions thoughtfully, it is important to think of them as a form of data curation and make editorial policies accordingly. My colleague and I are still trying to find the balance between creating indexing terms that are searchable without being misleading.

One of the challenges of working with metadata and DH is that there are few discussions on how to create editorial policies around metadata standards. DH specialists are often focused on the Text Encoding Initiative Standard (TEI) metadata standard (which is the primary standard in DH). As a result, there has not been much research on other schemas or authority control, although there are efforts to incorporate linked data into DH, such as the RDF Textual Encoding Framework.

Librarians are also largely silent on these issues in the context of DH. My experience at conferences such as ALA has been that cataloging and metadata sessions focus on metadata standards in the context of cataloging bibliographic material for use in library systems. These sessions tend to be heavily MARC-oriented and non-technical issues such as ethics are not often discussed.

It would be very helpful if there were more cross-community discussions at ALA and other conferences between librarians specializing in metadata and cataloging standards and experts in digital scholarship. This is particularly important as the role of Metadata Services in academic libraries expands from cataloging-based services to consultation services. Thinking broadly, digital scholarship, whether in the

humanities or the sciences, requires standards to be successful and metadata specialists are the experts who can provide advice, guidance, and support.

Metadata for IRs: Fast, Easy, and Useless? (June 22, 2015)

BY [CAROLYN HANSEN](#) · PUBLISHED JUNE 22, 2015 · UPDATED JULY 13, 2015

I recently met with a faculty member at my institution, the University of Cincinnati, to discuss submitting content to our digital repository Scholar@UC. He completed the metadata input form quickly and was not interested in describing his resources. When I asked him what was important for him in terms of the discoverability of his content, he replied: "Nothing. I just need a link to put in my journal article." To metadata specialists, this is a disheartening (but not uncommon) response.

Institutional repositories (IRs) often rely on self-submission models in which users create descriptive metadata for their content. As metadata specialists, we understand the importance of consistent, high quality metadata for good indexing within an application and discoverability by search engines such as Google Scholar. However, our users may not understand the significance of their metadata, or creating metadata may not be important for their immediate needs. It is necessary for metadata specialists to understand and acknowledge that high quality metadata requires an investment of time and resources, and our users may have little of both to devote to description. For these reasons, it is challenging to write metadata guidelines for IRs and even more difficult to enforce those guidelines.

There should be a balance between fast and easy input and useful descriptive data. Faculty and students want simple submission forms, but their metadata is often not effective for discovery and reuse of data without some form of mediation. For example, a dataset with a title of "q.txt" that does not have a readme file would be very difficult to understand or use by outside researchers. If a repository contains thousands of files with inscrutable names, the repository loses usefulness. Similarly, if a repository contains files without subjects, keywords, or other descriptive information, the relationships between those files cannot be made explicit for use by an application. In short, the repository becomes university storage space rather than a rich source of institutional knowledge.

This is where the mission of an institution's repository becomes central to the discussion. If the mission is primarily preservation-based, an argument could be made that metadata is irrelevant. As long as the content is being preserved, the repository is fulfilling its mission. Yet this argument presupposes a repository full of simple, self-contained files. It is relatively easy to understand a single text file of a student's thesis or dissertation. But when we move into interconnected files such as datasets, metadata is essential for understanding the structure and content of the data. Without metadata, these items may be unintelligible to someone outside of the lab that created them, and they become difficult to manage over the long term. Preserving something that cannot be understood is a poor investment of resources, regardless of whether content preservation is consistent with the repository's mission.

So what then is the role of the metadata specialist in regard to IRs? Are we evangelists for users creating granular metadata? Do we take autonomy away from our users and enhance their metadata to fit established guidelines and standards? Are we simultaneously teachers and enforcers?

I don't know the answers to these questions. But I think that raising them is important because while creating metadata for its own sake is not useful, neither is a repository full of unintelligible resources. I

want to respect the autonomy and needs of my repository's contributors, but I also think it is appropriate for those contributors to share in the responsibility of creating a meaningful IR application.

This begs the question: How are we handling this issue at UC? The answer is complicated because Scholar@UC is in a transitional stage. Since we are only working with early adopters, faculty buy-in is valued more than high quality metadata. We do not have enforced guidelines, only general recommendations that submitters can choose to follow or ignore. Once Scholar@UC is available to the campus at large, my hope is that our metadata guidelines can be strengthened as faculty begin to see the value of their metadata in terms of discoverability. I hope that librarians and submitters can work together to create metadata that is fast, easy, and useful.

Tags: institutional repositoriesmetadata qualityscholarly communication

Some International Dimensions of Metadata (July 10, 2015)

BY CHARLES RILEY · PUBLISHED JULY 10, 2015 · UPDATED JULY 13, 2015

Fellow Metaware contributor Carolyn Hansen opened up some interesting areas of discussion in her piece, "Standardizing Metadata for Digital Humanities." She examines issues of context and ethics relating to Digital Humanities, where standards for the Text Encoding Initiative (TEI) predominate. She notes the need for "more cross-community discussions at ALA and other conferences between librarians specializing in metadata and cataloging standards and experts in digital scholarship," with some focused attention on non-technical issues in contexts outside of MARC-based cataloging. While my experience has been largely based in MARC and not TEI, there are sets of issues common to both that are only starting to be adequately addressed. MARC has not yet left us and initiatives like TEI, the Bibliographic Framework, and others are still evolving. Negotiating between the communities these standards serve is important.

One area that deserves some attention in this space is that of internationalization, which for folks who specialize in the field is thought of as an architecture, or "essential part of initial software design", rather than a feature (Deitsch & Czarnecki, 2001). For metadata practitioners interested in international aspects, some of the most relevant pieces of the architecture are those which can be extracted from a MARC record or other formats to feed into a structured language tag following BCP47: the language value, the country, and any data available on the scripts used in the record or the resource. While examining these factors in turn, we can also consider a few points about the conditions under which this kind of metadata is created, the context in which it is found, and to what end we can expect that a language tag might be applied.

The conditions for a library cataloger creating internationalized metadata are constrained at many levels. Catalogers have access to library cataloging software that is usually part of a proprietary Integrated Library System (ILS) such as Sierra, Alma, or Symphony, among others. Decisions regarding expanded support for a character repertoire within those environments are complex and can take several years of consultation with OCLC, the Library of Congress, and the Program on Cooperative Cataloging, and other stakeholders. In 2007, a specification was added to MARC to allow for the lossless conversion between MARC-8 and UTF-8 character sets through the use of hexadecimal

numeric character representations (NCRs). I won't go into more detail here, but referring to the specification may give you some sense of what the constraints have been for fully implementing Unicode in the MARC environment.

Contextually, there are also constraints on what language and country tags have been available to the cataloger. The 008 field in MARC is populated with language values from a different list (ISO 639-2/B) than the lists commonly used on the World Wide Web (ISO 639-1 or ISO 639-3). Country tags for the web are pulled from ISO 3166, but there is a separate MARC list of country codes used by libraries. While the language tags used in HTML might be familiar as part of a locale — en-us for English in the US, fr-fr for French in France, and zh-cn for Chinese in China — the equivalents in MARC coding may look more like this: eng-xxu, fre-fr, and chi-cc. Language tags are not always evenly mappable, especially when it comes to more obscure special collections material, translations, or content with more than one language represented.

The reason for extracting metadata from a record in order to build a structured locale tag is to match search results more evenly with language and locale preferences as detected from the user's web browser settings. Whether this would also have the effect of making a given user more identifiable to law enforcement is an open question, so appropriate privacy and minimization measures should be considered in light of continued efforts to collect metadata at scale for government surveillance projects. Data on any non-Latin scripts used in a record can be detected in some cases by querying the 066 field for content, but in most cases it would be more useful to rely on the Common Locale Data Repository (CLDR) library of exemplar characters for a language, and determine the script from its usage in the 880 fields.

In general, MARC offers a fairly high degree of granularity with respect to requirements for the interoperability of internationalized metadata; in many cases it offers a richness that deserves to be maintained. Unpacking its intricacies can be of interest for re-use and for delivering to the user desired content across platforms.

Metadata and Project Hydra (July 20, 2015)

BY CAROLYN HANSEN · JULY 20, 2015

In April 2015, a working group of Project Hydra partners and implementers was formed to provide metadata recommendations to increase interoperability among current users and help new Hydra implementers. The Hydra Metadata Working Group, under the direction of Karen Estlund, quickly established subgroups and initial deliverables to be presented at the Hydra Connect 2015 Conference. The subgroups address a wide-range of metadata types and issues, such as implementing Linked Data Fragments, mapping MODS to RDF, and writing FITS recommendations. The subgroups include the Applied Linked Data Subgroup, the Descriptive Metadata Subgroup, the Rights Metadata Subgroup, the Structural Metadata Subgroup, and the Technical Metadata Subgroup. Since transparency is an important component of Project Hydra's mission, meeting notes are freely available and comments are welcome.

I serve on the Hydra Metadata Working Group and also facilitate the Descriptive Metadata Subgroup. Descriptive metadata is a somewhat controversial issue in Hydra because of the diversity of Hydra partners and implementers. Members include various types of libraries and institutions, including academic, corporate, museum and special collections, and aggregators. As a result, the subgroup must work to create metadata recommendations that balance specific institutional needs with system constraints and interoperability. This challenging work is a later deliverable to be completed as part of the group's second phase.

What descriptive metadata schemas are you using?

- ☐ DPLA Application Profile
- ☐ Dublin Core
- ☐ MARC/MARCXML
- ☐ MODS
- ☐ PBCore
- ☐ VRA Core
- ☐ Local schema
- ☐ Other: [_____]

Sample Question from the Hydra Descriptive Metadata Survey

The Descriptive Metadata Subgroup is currently working on the first phase of its charge, which is to conduct a survey of descriptive metadata used by Hydra implementers. To my knowledge, this is the first attempt to perform an environmental scan on descriptive metadata within the Hydra community; it represents an important opportunity to discover current practice as well as the roadblocks that are preventing institutions from doing all that they wish to in their repositories. Questions about current practice include topics such as encoding standards, descriptive metadata schemas, local metadata fields, domain specific metadata, controlled vocabularies, sources of metadata, workflows for metadata creation, export formats, and more. Survey results will be presented at Hydra Connect 2015 in September.

Additionally, the Descriptive Metadata Subgroup has a new sub-subgroup, the MODS and RDF Descriptive Metadata Subgroup. This group's charge is currently being formed and will address how to handle "MODS XML in a linked data / RDF world."

If you're interested in learning more about Project Hydra or connecting with metadata group members, visit the Hydra Metadata Working Group website.

Tags: descriptive metadataProject Hydrastandards development

Introducing Student Created Content to Metaware (July 27, 2015)

BY MIKE BOLAM · PUBLISHED JULY 27, 2015 · UPDATED OCTOBER 26, 2015

At the 2015 ALA Midwinter meeting of the ALCTS/LITA Metadata Standards Committee in Chicago, attendees brainstormed methods for increasing the flow of content on the Metaware blog. From the blog's inception at the 2014 ALA Annual Meeting, it had only seen three posts published. At the time, I was dealing with revising my syllabus for the metadata course I teach at the University of Pittsburgh, and I suggested that I could add an assignment that would allow students to create content for the blog. The idea was popular with the committee members, and I moved forward with incorporating the assignment into my syllabus.

The course, LIS2407 – Metadata, is offered at the University of Pittsburgh School of Information Sciences, where I am an adjunct faculty member. It is an elective course in the Library and Information Sciences Master's Degree program. The course descriptionreads as follows: "Principles and application of metadata for networked information-resource organization, representation, retrieval, and interoperability using a variety of schemes and tools." This semester, there are 25 students enrolled in the course. About half of them are in the archives specialization and the others are primarily in the academic libraries specialization or individualized program. A majority of the students are in their third (and final) semester.

Using the contributor guidelines from Metaware as a model, I developed the assignment. It counted as 10% of the student's grade, and publication on the site was not required to meet the expectations of the course. Students were able to opt out of publication after submitting their post for the course. The students were required to select a topic of interest and submit a 500-1000 word essay that matched one of the Metaware categories: Analysis (longform discussions of a topics of interest in the metadata community), Editorial (opinion pieces for sharing personal observations related to metadata practice), or Review (Reviews of books, tools for metadata creation, and other resources of relevance to the metadata community).

The other assignments in the class required the students to work in pairs, and I decided to continue that policy for the blog post. Since this was a new initiative for Metaware, I was concerned about flooding the blog with 25 posts from students, or that we might see some duplication in the selected topics. Working in pairs gave students the opportunity to bounce ideas off of one another and edit each other's work. It also seemed to alleviate some of the stress they might have felt about creating content for a publicly accessible blog. The students submitted their posts via Blackboard, and I provided feedback and assigned grades. The posts were then moved to Google Docs, and access was granted to the students and the Metaware editor. The editor made recommendations and edits that were turned back to the students for approval.

The greatest challenge we experienced during the semester was the lack of example content on the blog. When I offered to make this part of my course, I was hoping more to see more original content published on the blog. During the time between the Midwinter meeting and assigning the project, only three posts with original content were added to the blog, and two of those were related directly to the work of the Metadata Standards Committee. Some of the students expressed concern about their lack of expertise in the area, and were not sure where they could contribute. I encouraged them to explore

other blogs, readings assigned for class, and assignments completed for other courses. At the time of writing, eleven of the thirteen posts will be ready for publication after minor editing. The remaining two posts require substantial editing before posting.

Over the next three to four months, I will be adding the student posts to Metaware. Assuming I'm teaching the course again next summer, I hope to include a similar assignment. I gave a status report on the project at the Metadata Standards Committee's meeting at ALA Annual in June 2015, and we are hoping that other instructors consider including a Metaware writing project in their courses. There is great potential in giving students professional writing opportunities and allowing them to showcase their work to the greater metadata community. If you would like to consider incorporating a Metaware writing project into a course, please contact the Metadata Standards Committee via the Metaware Contributor Volunteer form.

- The "F-Word": Folksonomies (Give them a Chance!) by Amy Berish and Amy Dinkins

- Practicalities of Standards Adoption by Megan Massanelli and Mary Phillips

- Telephony Metadata and the Rights of U.S. Citizens by Leah Geibel and Erin Scrimger

- How I Explain Metadata to the Non-Metadata World by Angelina Spotts

- Metadata in the Real World by Reba Sell and Emily Schoenlein

- More Metadata, Less Process? by Kira Condee-Padunova and Laureen Wilson

- Metadata Interoperability Among LAMs by Samantha Cabo and Sara Purifoy

- Metadata between Archivists' Toolkit and ArchivesSpace by Dominique Luster and Jon Klosinski

- "Metadata for All": Looking Back to Metadata Standards by Eleanor Godbey and Kathleen Donahoe

- oXygen Review from New Metadata Learners by Gesina Phillips and Christie Kliewer

- Searching is Not the Answer by Rose Chiango and Katelyn Quirin

Tags: metadata educationMLIS students

The "F-Word": Folksonomies (Give them a Chance!) (July 27, 2015)

BY MIKE BOLAM · PUBLISHED JULY 27, 2015 · UPDATED JULY 27, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see theintroductory post.

By Amy Berish and Amy Dinkins

As individuals we are constantly trying to organize and categorize the world. Most of the time, these specific duties are left in the hands of trained information professionals such as archivists, librarians,

and other professionals who create and collect metadata. What if this task was left to the users of this information? Folksonomies, or user created metadata, increase the discovery and overall use of collections by creating more access points. "[A] folksonomy evolves when many users create or store content at particular sites and identify what they think the content is about" (Gartner 2015). Folksonomies usually take the form of tags created in a social structure where the consumer of the information is the entity creating the tag. Folksonomies are unique and attractive because they put the power of description in the hands of the user. Often times, the controlled ways metadata and information professionals describe objects may not coincide with the general language people use to search for and retrieve information. Folksonomies help with this as they allow information seekers to use natural language to describe items. With a specific local structure, information professionals can enhance their metadata by outsourcing descriptive gaps to users.

The beauty of folksonomies is that they are completely uncontrolled. This usually does not jive well with the work of information professionals since they are usually the ones in total control. Often, "traditional, hierarchy, chaos, classification and authority are all words that swirls around the talk of …. folksonomy" (Edmunson-Morton 2009). So why should information professionals give folksonomies more attention? According to OCLC, "of the user-contributed content that would most enrich the metadata created by libraries, archives, and museums, more than half improve description. Almost half contribute content to the resources already offered" (Smith-Yoshimura 2012, 5). Additionally, user-generated metadata can help institutions bridge descriptive gaps, increase discovery and use of materials, and save time and money.

Oregon State University, in their recent metadata project, used folksonomies to assist in describing photographs in the Gerald Williams Collection. The project utilized the built-in tagging structure provided by Flickr. The project resulted in "increased visibility and access to our collections… [provided] avenues for further study or research, and [gave] our users a unique opportunity to interact with the Archives and other users" (Edmunson-Morton 2009).

A suggested way to bridge the inherently messy nature of folksonomies with the interests of information professionals is to utilize a local vocabulary where terms are suggested and users can suggest additional terms for the tagging system. At the beginning of the project, the OSU Archives staff developed a policy of persistent and consistent tagging for their Flickr accounts. Since the goal was to encourage user participation, the staff limited provided tags in order to reduce their influence on users (Edmunson-Morton 2009). It is a delicate balance, but using folksonomies does not have to mean dropping authority files completely. Controlled vocabularies are created from the needs of archivists, librarians, and information professionals – the people who use them. The way users categorize and retrieve information is often absent in these structures, leaving a very noticeable gap. Folksonomies are one way to fill it. With metadata generated by users, information professionals can identify new relationships, incorporate terms used by the community, and make it easier for people to discover the materials. Scary as it may be, information professionals should loosen their grip and give users a space where they have power over description. There are considerable benefits to folksonomies, and it is up to the professionals not to waste them.

References:

Edmunson-Morton, Tiah. June 19, 2009. "Talking and Tagging: Using CONTENTdm and Flickr in Oregon State University Archives." The Interactive Archivist. Society of American Archivists. http://interactivearchivist.archivists.org/case-studies/flickr-at-osu/

Gartner IT Glossary. "Folksonomies." http://www.gartner.com/it-glossary/folksonomy. Accessed June 2015.

Smith-Yoshimura, Karen. 2011. Social Metadata for Libraries, Archives, and Museums: Executive Summary. Dublin, Ohio: OCLC Research: http://www.oclc.org/content/dam/research/publications/library/2012/2012-02.pdf

Tags: folksonomiesMLIS students

Practicalities of Standards Adoption (August 3, 2015)

BY MIKE BOLAM · PUBLISHED AUGUST 3, 2015 · UPDATED JULY 28, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the introductory post.

By Megan Massanelli and Mary Phillips

The "DRAFT Checklist for Evaluating Metadata Standards," submitted at the beginning of the year by the Association of Library Collections and Technical Services (ALCTS) / Library and Information Technology Association (LITA) Metadata Standards Committee of ALA, provides something akin to a mission statement for the metadata community. It succeeds at providing theoretical scaffolding to guide practice. The "Discussion notes: DRAFT Checklist for Evaluating Metadata Standards," posted in March 2015, points out that further specificity in regards to the themes and issues from the Checklist would provide better guidance in practical application.

In seeking an actionable solution to these questions, a case study in the problems with navigating the quagmire of the "multiple communities" mentioned in #7 of the Checklist may prove a helpful point of departure for an ongoing discussion of metadata standards in practice. In addition, the issue of selective adoption and the utility and scalability of standards across disciplines found in #2 and #8 can benefit from an on-the-ground study that unifies the recurrent themes into a single, succinct point. We aim to provide a more granular example of how the Checklist can support consistent metadata practices across disciplines.

Issues in creating a metadata schema for the digital collections of the Center for PostNatural History (CPNH), a small museum of biological specimen and research materials affiliated with the Frank-Ratchye Studio for Creative Inquiry at Carnegie Mellon University in Pittsburgh, Pennsylvania, provides such a case study. Megan has been working with the CPNH since May 2015 to establish a standard schema for museum collection materials. The CPNH states that its mission is "to acquire, interpret, and provide access to a collection of living, preserved, and documented postnatural organisms" (CPNH Brochure). The term PostNatural refers to living organisms, whether plant, animal (including humans) or microbial, that have been genetically altered by human intervention through

selective breeding, or transgenetic manipulation. The CPNH collects, studies, and displays organisms that have been genetically altered by humans, or human-made processes, as well as published books, periodicals and photographs. As a museum of both scientific and cultural heritage significance, the Center does not seek to take a stance or influence the judgement of visitors on the topic of PostNatural, but simply to provide a space for the presentation of information and exploration of concepts and ideas that are largely absent from institutions of natural history.

Tags: MLIS studentsstandards assessment

Enabling Access to Resources in Non-Latin Scripts (August 10, 2015)

BY CHARLES RILEY · AUGUST 10, 2015

Somewhere between the realm of the symbolic and the Lacanian Real, there exists something akin to Michael Binkley's Anxiety Closet of resurgent Bloom County fame: characters living there may all have identifiable names, but they have a tendency to remain hidden and only emerge to torment us when we are otherwise most at peace. So it is with romanization. Resources that need to be described are identifiable as distinct from one another—whether both in their original script and transliterated form, or just through the romanized version of their metadata—but they still have a tendency to keep us awake at night.

Representing non-Latin scripts in the Latin alphabet often calls upon a need to use special characters or diacritics to keep everything well-accounted for and, in many cases, reversible to its original script. Many of the romanization tables in use by the Library of Congress and the ALA have, as a design goal, the intent of one-to-one correspondence between characters on each side of the table. This goal is not always possible to achieve. For most scripts, the conversion process readily lends itself to automation, but there are a few scripts (e.g. Japanese kanji and Malayalam) that defy automation and require dictionary lookup or more sophisticated algorithms to process.

From time to time, romanization tables come under review for different reasons. Sometimes user expectations have drifted away from a traditional transliteration method, and sometimes a newer method receives government approval, ISO acceptance, or lends itself to automation—though rarely all three. In the case of the recently approved revision to Tibetan, the motivating factors were really about meeting user expectations with a method that is easier to automate. Batch re-conversion of older records is possible, but the difficulty is prohibitive with current technology. This was a consideration raised in the debate over whether to revise the romanization table.

Including the original script into metadata has been getting easier, especially since about 2006. OCLC has done its part in expanding coverage from the traditionally supported JACKPHY scripts (Japanese, Arabic, Chinese, Korean, Persian, Hebrew and Yiddish). It announced support for Bengali, Devanagari, Tamil and Thai in 2006 and has since expanded to cover Syriac, Armenian and Ethiopic. More initiatives are taking place that address, for example, Georgian (British Library), Gujarati (D. K. Agencies), Lao (National Library of Australia and National Diet Library in Japan). Cases like extended Latin (a.k.a. International Phonetic Alphabet), extended Arabic (a.k.a. Ajami), Cyrillic and Greek are also much closer to being ready for wider acceptance in the metadata ecosystem.

Direct support for Tibetan and Mongolian scripts received priority consideration from respondents to a Script Priority Survey that was carried out in 2013 under the auspices of the ALA's Committee on Cataloging: Asian and African Materials, which I chair. While Tibetan should be relatively straightforward now with a new romanization table in place, Mongolian has some complex text rendering and layout requirements that might put its direct support on a slower track, at least for now, until things like vertical text layout become more widely supportable.

There are, even so, anxieties in Binkley's closet remaining to be provoked, but we will leave those for a future discussion.

Telephony Metadata and the Rights of U.S. Citizens (August 17, 2015)

BY MIKE BOLAM · PUBLISHED AUGUST 17, 2015 · UPDATED AUGUST 18, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the introductory post.

By Leah Geibel and Erin Scrimger

Many U.S. Citizens do not realize that telephony metadata is available to the government through the Patriot and Freedom Acts. The legality of this bulk collection is questionable, however, and its continuation is suspicious. The Foreign Intelligence Surveillance Act (FISA) of 1978 requires that telephony metadata only be made available if records being sought are relevant to an investigation, a subpoena is obtained, and pen registers are used to trap and trace devices. The Patriot and Freedom Acts run contrary to these stipulations and the government has given themselves the right to access telephony metadata carte blanche. FISA has not been updated to include new technologies, so metadata collection is often a loophole that is ultimately unconstitutional. Definitions concerning the difference between foreign affairs and foreign intelligence must be articulated, and limits about what can be collected must be enforced in order to ensure the privacy of U.S. Citizens (Donohue, 2014). Absent these limitations, the government is free to extract information without any legal cause or acknowledgement.

We might not like the idea of governments collecting our private information, and in fact it may not even be legal, but that doesn't mean it's going to stop anytime soon. The National Academy of Science (NAS) recently issued a report titled Bulk Collection of Signals Intelligence: Technical Options which determined that "refraining entirely from bulk data collection will reduce the nation's intelligence capabilities" and that controlled usage of data collected in bulk can be a way to protect privacy (2015). There are two ways to control usage: manually and automatically. The NAS study suggests a heavier shift toward the automatic by using software to determine if queries from intelligence analysts are allowed. However, this raises the question of how do we know what we don't know? If analysts aren't provided access to the broader picture, in this case bulk metadata, then how can they gain the context from which to derive the right kinds of queries? Can an algorithm determine as well as a human what is relevant?

Before continuing, it should be made clear that the authors do not pretend to know or understand the finer workings of signals intelligence (SIGINT) or the systems and algorithms used to interpret it. What

follows is merely thoughts that have arisen from contemplating the implications of moving toward a more strictly controlled automatic system of data analysis and how metadata standards could in some small way play a part in personal privacy.

The NAS report points toward a shift from bulk collection to targeted collection, where collection is defined as having happened only after information is moved and stored by the government (as opposed to remaining on corporate servers). If we are moving toward a system of algorithms that assess queries, then we should consider the use of controlled vocabularies (Tucker, 2015). At some point there would be no need for manual control, which is what most people are concerned about after all. Intelligence collection is necessary, but I don't want someone listening to my phone calls or reading my email. To answer the "how do we know what we don't know" question, we might not need to know. Presumably analysts are searching for different things, sometimes similar, and through searches they might find different pieces that together provide important information (Tucker, 2015). If software can be programmed to digest speech or text in real time and identify key phrases or words from a controlled vocabulary, it could lump those data sets together and assign subject tags. This, in a way, would be performing targeted collection. From these sets of data, analysts could only view relevant information on designated targets. The example given by the NAS report is a specific phone number.

This brings up two further points relating to controlled vocabularies. First, there would be the need for a committee whose sole purpose would be to create and update this list of controlled subject terms and make it available to analysts as new intelligence is gathered. This committee would be responsible for defining the subject tags that the software would place on data sets and also those terms that would be flagged as holding possible information. Second, systems and practices of intelligence gathering must have ethical checks and balances, and this is where manual (or human) usage control factors in. There would need to be some kind of elected advisory council that would review these controlled vocabularies before entering them into the algorithms to ensure that our rights and privacy as citizens were being upheld. For example, vocabularies dedicated to political or social issues should never be created for the purpose of tracking and monitoring the free speech of citizens, as was done during the Civil Rights Movement. Fear is often bred out of mistrust and the unknown. We should not fear our government, and this is why transparency in SIGNIT is so important. With tightly controlled and supervised usage, transparent targeted collection of metadata for the purpose of intelligence should not interfere with the privacy of our daily lives.

References:

Donohue, Laura K., "Bulk Metadata Collection: Statutory and Constitutional Considerations". Georgetown Law Faculty Publications and Other Works. Paper 1350. (2014). http://scholarship.law.georgetown.edu/facpub/1350

National Academy of Science. Bulk Collection of Signals Intelligence: Technical Options. 60. Washington, D.C.: The National Academies Press, 2015.

Tucker, Patrick. "What the End of Bulk Metadata Collection Would Mean for Intelligence Collection." May 13, 2015. http://www.defenseone.com/technology/2015/05/what-end-bulk-metadata-collection-would-mean-intelligence-collection/112519/

How I Explain Metadata to the Non-Metadata World (August 24, 2015)

BY [MIKE BOLAM](#) · PUBLISHED AUGUST 24, 2015 · UPDATED AUGUST 17, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the [introductory post](#).

By Angelina Spotts

For the few years that I've been working with metadata, I've had to answer that question that most librarians who don't work with reference questions and stacks of books dread: "What do you do?" I do admit that, at times, I've used the trite phrase "data about data" knowing full well it went a bit deeper than just that. Recently I have begun to improve the description of my work by being more whimsical in my answer, thereby avoiding the stress and frustration that comes with explaining metadata work to people who probably would not understand it anyway, no matter how much technical explaining I offer. Nowadays, I just sing the following song to them:

[The Rattlin' Bog](#)

O-ro the rattlin' bog,

The bog down in the valley-o

O-ro the rattlin' bog,

The bog down in the valley-o

And in that bog there was a tree,

A rare tree, a rattlin' tree

With the tree in the bog

And the bog down in the valley-o.

(Chorus)

Now on that tree there was a limb,

A rare limb, a rattlin' limb

With the limb on the tree

And the tree in the bog

And the bog down in the valley-o.

(Chorus)

Now on that limb there was a branch,

A rare branch, a rattlin' branch

With the branch on the limb

And the limb on the tree

And the tree in the bog

And the bog down in the valley-o.

(Chorus)

Now on that branch there was a twig,

A rare twig, a rattlin' twig

With the twig on the branch,

And the branch on the limb

And the limb on the tree

And the tree in the bog

And the bog down in the valley-o.

And you get the idea. The song goes on to refer to a twig, a nest, an egg, a bird, a feather and a flea. Until you end up with a final stanza that reads like this:

Now on that feather there was a flea,

A rare flea, a rattlin' flea

With the flea on the feather,

And the feather on the bird,

And the bird in the egg,

And the egg in the nest,

And the nest on the twig,

And the twig on the branch,

And the branch on the limb

And the limb on the tree

And the tree in the bog

And the bog down in the valley-o.

O-ro the rattlin' bog,

The bog down in the valley-o

O-ro the rattlin' bog,

The bog down in the valley-o

This is awfully wordy, but then so is metadata. It's a lot of information. So what I usually say, after singing the entire song and enjoying the worried looks of fear and confusion, is this: suppose you were looking for a particular tree and there was a huge database filled with trees of all types. You want one that contains an egg in a nest. You search "egg, nest, tree," and it might pull up a bunch of trees because there are many trees containing eggs in nests, but you want the one where the bird has a flea on its feather. So you search again using words like "flea, feather, bird, nest, tree, etc.," and the system, whether Google or a library catalogue, narrows it down to this tree and a few others. Without all of the descriptive information about the tree included with the item, which is a very specific tree in this case, you would struggle to find the exact tree you wanted.

Now we can go even further to say meta-metadata and start narrowing it down to what kind of flea or bird, but I usually won't go into that. The song seems to at least get folks to understand that metadata is all of the information about that one item. Usually I am not working with trees, but someone out there is, so this is not irrelevant. The same can be said for an archival item, a book, research data or whatever it is that a particular metadata librarian may be working on. A rare artifact and a rattlin' artifact, no matter what kind of artifact, metadata will help you get to the artifacts as well as preserve the information surrounding that artifact for someone who might need to use it in the future.

This, in my mind, is metadata for the non-metadata familiar.

Tags: metadata educationMLIS students

Metadata in the Real World (August 31, 2015)

BY MIKE BOLAM · PUBLISHED AUGUST 31, 2015 · UPDATED OCTOBER 26, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the introductory post.

By Reba Sell and Emily Schoenlein

As we continue to advance through the Digital Age, more and more information is shared online in various formats. Anyone can create blogs, share photographs, or spin a new record in order to share it with the world. Our personal documents, images, videos, and sound recordings are important to us,

and yet we often neglect to name them in a controlled manner or save them in locations that would help combat digital obsolescence. Many people believe that uploading an image to Facebook or Instagram means that their image will be online and accessible forever. As archivists and information professionals, we know that this is not the case. Social networking sites allow us to add tags, which are actually metadata, to our photographs, tweets, statuses, and other social media products. Most people happily tag their social media output, often using hashtags, but for most people this is just for the purpose of sharing their photographs, posts, etc. in hopes that more people will see and 'like' it instead of for any type of cataloging system. As the number of files that we are able to store increases with ever-increasing computer storage capabilities, our information can easily get lost without proper identification.

Why don't most people tag their data on their personal or work computers? Perhaps the simple answer is that people do not realize that they have the ability to add these tags. Or perhaps people do not think of tagging as a necessary step to finding their information at a later date. Adding this simple form of personal metadata will allow people to organize and search through their work more easily. Archival documentation and library materials require metadata in order to make these resources searchable and accessible. If individuals would begin to implement their own metadata, it would allow future archivists to better understand the collection creator's thoughts, and the subjects within the documents that they examine. At the very least, adding metadata to personal files would allow people to more easily organize and retrieve their own information. For example, this essay could be tagged as "metadata", "blog", "University of Pittsburgh", and "graduate school". This would enable a future archivist to see any work done while in graduate school at the University of Pittsburgh or items relating specifically to metadata. The addition of identifying information to the names of image files can also be very helpful. Most devices such as cameras or smartphones use the same basic file naming system, such as "DC0001" and onward, to name images when they are transferred to a computer. Changing the name of the image file, or even just the folder containing the images, to include metadata such as the date and name of the location or event will save you much more time later on.

While it might be too much to expect every individual who creates any kind of record to create metadata, it would certainly be a useful undertaking to perform for those who hope to share, save, or archive those records. It is the responsibility of information professionals to not only decipher personal naming conventions but also to teach and assist people with using metadata for their files. This might mean the development of tools like desktop search software such as the now out-of-date Beagle++ for Linux and UNIX. It at least means that people should be more aware of how to better name their personal and work files instead of just dropping them in folders. It is not enough for people to rely on social media to save their important information. Even though you may think that your pictures are safe on Facebook, it is still a for-profit company that may not be around forever. It is imperative that people learn how to effectively organize and store their personal files, and this starts with the help of archivists and other information professionals.

Tags: [metadata educationMLIS students](#)

More Metadata, Less Process? (September 7, 2015)

BY [MIKE BOLAM](#) · PUBLISHED SEPTEMBER 7, 2015 · UPDATED SEPTEMBER 1, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the [introductory post](#).

By Kira Condee-Padunova and Laureen Wilson

While there are compelling arguments for the use of standardized metadata schemas, many institutions still choose to rely on local metadata creation rules. This blog post discusses the pros and cons of standardized and local metadata schemas using the idea of "More Product, Less Process: Revamping Traditional Archival Processing," first introduced by Greene and Meissner in a 2005 article in The American Archivist regarding processing standards in archival repositories.

Greene and Meissner's article has had a major influence on the archives field since its publication ten years ago. However, the concepts suggested by MPLP, as the concept is often abbreviated, have not permeated into the related, and sometimes overlapping, metadata field. Both archivists and metadata creators deal with backlogs of work that needs to be completed in a timely manner, and neither archivists nor metadata creators start projects to lock items up in some kind of processing purgatory. The idea is to make information accessible to users. Of course, many of the issues Greene and Meissner cite with archival processing procedures, such as the fastidious removal of metal fasteners, have no relevance for metadata creators. Even so, it may be worthwhile for metadata creators to consider the bigger implications of MPLP and figure out what would constitute a "golden minimum" for metadata creation.

The purpose of this blog post is not to recreate Greene and Meissner's work in the metadata field, but, rather, to reconsider common practices among metadata creators and question whether or not the results produced are worth the time it takes to perform them. Perhaps the most common practice among metadata creators is the preference for using standardized metadata schemas rather than locally created metadata schemas. Standardized metadata schemas have obvious benefits, but can also consume large amounts of time and create problems for their users.

An institution that wishes to use a standardized metadata schema when creating its own records must first choose which standard they plan to use. This requires research on the part of staff, which will likely be time-consuming, especially if the institution has no staff members with experience in the metadata field. Even with careful research, there is no guarantee that the standard chosen will be the best choice for the institution. Institutions may have a difficult time finding a standard that contains all the information relevant to their particular needs. Committing to a standard also requires that the institution remain up-to-date with any changes made to that standard.

Smaller institutions are more likely to lack a dedicated employee for metadata and often rely on a small number of employees or volunteers to cover all types of work. Because of this, the time and effort involved in choosing and applying a standardized metadata schema may not be worth the potential benefits. Large institutions may also have their own problems with standardized metadata schemas. Although more likely to have staff trained in and dedicated to creating metadata, large institutions are also more likely to rely on their institution's IT department to assist in metadata creation – departments who might not even be familiar with standardized metadata schemas.

On the other hand, relying on a local metadata schema can create its own problems for the institution and delay the release of items for users. The beauty of standardized metadata schemas is that everything is already set-up. Instead of having to discuss which fields to add to a record, the most common fields are already there, waiting for data. There is not a risk of new metadata creators coming in and complicating the record by adding in more descriptors. Once each field has been populated (or not, as the case may be) the record can be published and the item made available to users. A standardized metadata schema might not provide the most in-depth information about an object, but it gives enough for users to find the object they need.

Even in situations where a local metadata schema is already in use, switching over to a standardized metadata schema should improve the ability of users to discover relevant results. Indeed, metadata schemes are often very friendly to both creators and researchers. Dublin Core is lauded for using easy-to-understand descriptors in its schema so that those filling out a record can immediately understand what should fit in each field. This allows for more simplicity on both ends, meaning that users can not only access the products faster, but also understand the titles for each description to see if the item is what they need. Since they're also clear from the creator's side, it means that the creation of metadata records can be left to volunteers or interns – after a sufficient training period – so that metadata creators can focus on curating and maintaining their collection.

If the focus is less on perfection and more on production – essentially More Product, Less Process – then standardized metadata schemes offer more freedom to institutions. With any new project there will be growing pains, but, once over the hump of implementing a new scheme, standardized metadata schemes allow institutions to collaborate more efficiently. The simplicity of the generalized descriptors may require some institutions to try to fit into a more narrow terminology, but the end result could be more helpful to the people who would use the resources that those institutions are offering.

Greene, Mark A., and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." The American Archivist 68, no. 2 (2005): 208-63. Accessed July 10, 2015. http://www.archivists.org/prof-education/pre-readings/IMPLP/AA68.2.MeissnerGreene.pdf.

Tags: metadata qualityMLIS students

Metadata Interoperability Among LAMs (September 14, 2015)

BY MIKE BOLAM · PUBLISHED SEPTEMBER 14, 2015 · UPDATED SEPTEMBER 1, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the introductory post.

By Samantha Cabo and Sara Purifoy

As data that describes digital or non-digital resources, metadata is integral to the management of items within institutional media repositories such as libraries, archives, and museums (LAMs). These institutions have long used the MARC standard as it includes an effective metadata schema in addition to encoding, storage, and exchange systems. However, the recent growth of digitized and born-digital resources complicates the metadata creation process across information communities. This is

particularly prevalent in terms of metadata interoperability or "the ability of two or more information systems to exchange metadata with minimal loss of information," which is beneficial because it can be used by interconnected computer systems to communicate and transfer data (Neiswender, 2009). Web-scale interoperability of metadata eliminates the isolation factor inhibiting information systems from maintaining and integrating their local database systems; descriptive information can now be stored in the form of a Hypertext Transfer Protocol (HTTP) Uniform Resource Identifier (URI), along with institutional holdings and access information (Seeman & Goddard, 2015).

The move toward implementing linked data within LAM databases is gaining momentum, mostly because of the varying degrees of inconsistency among databases and catalogs. However, with the constant uncertainty of what standards the future may hold regarding metadata interoperability, this transition has largely been theoretical rather than practical. Luckily, there are LAMs who have taken the plunge and emerged triumphant, or well on their way to success. Although this analysis references only a few of these examples, it aims to discern what best practices may look like for LAMs interested in bridging the gap.

One of the biggest challenges LAMs face when it comes to interoperability is the issue of systems integration. LAMs are generally used to operating as silos, each with a localized catalog or database, and little time or manpower to dedicate to importing and reformatting records from other institutions. To facilitate maintenance and integration of local systems, and support the need for cross-domain integrated searches, time should be invested in the development of excellent metadata (to be adopted across all LAMs) in lieu of updating local records that only serve their parent institution and users (Seeman and Goddard, 2015). As this metadata is developed, domains will become more linked across the web. Other institutions, projects, and LAMs will begin to reuse these links, or URIs, thereby improving the interoperability of data within the domain. On a high semantic level, this means that both humans and computers must be able to read and interpret the metadata correctly across platforms (Haslhofer and Klas, 2010). A call to arms aimed more particularly at LAMs, however, is to advantageously use this process as an opportunity to expose their unique resources and special collections not only to one another, but also to a wider group of users (Seeman and Goddard, 2015).

In order to gain wider exposure for LAM resources, structural and semantic heterogeneities, or the differences between metadata models across institutions, need to be eliminated. However unlikely, this could be achieved if a powerful organization such as the World Wide Web Consortium (W3C) were to recommend a standardized metadata model to be used across LAMs. Haslhofer and Klas suggest that metadata mapping is the best technique for achieving metadata interoperability (2010). Metadata mapping is the process by which metadata from one repository is successfully copied, translated, and then used within a second sister repository. However, metadata interoperability requires a mapping schema that can account for all technical levels of metadata: M2 schema definition languages, M1 metadata schemes, and M0 metadata instances (Haslhofer and Klas, 2010). If a metadata mapping schema can incorporate techniques that recognize and address these instance levels, it is the best (albeit most technical, time consuming, and costly) bet for a consistent and comprehensive solution.

Although the transition to linked data may seem daunting and clouded in risk, it is up to LAMs to invest in their expertise in metadata generation and management, and to look to one another for help in this effort. There has been great technological change and innovation in a brief period of time. LAMs must consider where they want to stand in the new information society, and how they are going to get there.

For LAMs to survive, their practices, services, and resources need to flow to their users on the web. "[T]hey must be of the web not just pass through the web but live and interact with web resources" (Zengenene, 2013). The mission of LAMs has always been the unbiased provision of information and resources; Zengenene (2013) quotes Sir Tim Berners Lee (creator of the World Wide Web) as saying, "data isn't worth much until it's free…freed from the silos in which it is locked up, and used in a mash-up that creates valuable new resources for you and others."

Works Cited:

Haslehofer, Bernhard and Wolfgang Klas. 2010. "A Survey of Techniques for Achieving Metadata Interoperability." ACM Computing Surveys, 42(2): 1-37. DOI:10.1145/1667062.1667064.

Neiswender, C and E. Montgomery. 2009. "Metadata Interoperability–What Is it, and Why Is It Important?" In the MMI Guides: Navigating the World of Marine   Metadata. http://marinemetadata.org/guides/mdataintro/mdatainteroperability.   Accessed June 15, 2015.

Seeman, Dean and Lisa Goddard. 2015. "Preparing the Way: Creating Future Compatible Cataloging Data in a Transitional Environment." Cataloging & Classification   Quarterly, 53(3-4): 331-340. DOI: 10.1080/01639374.2014.946573.

Zengenene, Dydimus. 2013. "Global interoperability and linked data in libraries." New   Library World, 114(1-2):84-87. http://dx.doi.org/10.1108/03074801311291992.

Tags: MLIS studentsstandards assessment

Metadata between Archivists' Toolkit and ArchivesSpace (September 21, 2015)

BY MIKE BOLAM · PUBLISHED SEPTEMBER 21, 2015 · UPDATED SEPTEMBER 16, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the introductory post.

By Dominique Luster and Jon Klosinski

Archivists' Toolkit (AT) is an open source content management system designed to assist archival institutions in asserting intellectual control over their archival collections and representing them in the most discoverable way possible. These representations are generally presented in archival records or in a public interface in the form of a finding aid. According to the AT website, the purpose of this system is to "support archival processing and production of access instruments, promote data standardization, promote efficiency, and lower training costs." AT offers substantive movement forward in the means of controlling archival metadata. Like most content management systems, the idea of mandatory descriptive fields is a strength for AT. Yet the lack of standards within those mandatory descriptive fields is a significant drawback.

Most institutions utilizing AT rely on supplemental metadata standards to determine how they create their finding aid records. A few of these standards include the General International Standard Archival Description (ISAD(G)), Describing Archives: A Content Standard (DACS), Encoded Archival Description (EAD), International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR (CPF)-EAC) and as well as more commonly known bibliographic standards such as AACR and MARC.

Despite the fact that AT cannot stand alone and leans heavily on the usage of other metadata standards, the system is flexible enough to be compliant with various types of outputs. For example, AT has the ability to both import and export EAD markup language and MARC XML records. Additionally, it has the ability to export MARC XML, MODS, DC, and METS (with MODS or DC). AT also has a number of metadata limitations. Its primary limitation is that there is no authority control or reinforcements between mandatory internal standards and external standards.

While AT does have the ability to export EAD, HTML, and PDF versions of finding aids, it lacks a central interface necessary for providing online access. ArchivesSpace, the open source successor to AT, attempts to extend the metadata management functionality AT users are accustomed to into a new public interface which is accessible to non-institutional or account-authenticated users. The public interface includes plugins for additional output formats of resource and digital object records. In comparison with AT, ArchiveSpace's public interface provides a very convenient way to retrieve various object representations. Supported formats for resource retrieval in ArchiveSpace include EAD, HTML, MARCXML and EAD PDF. Supported formats for digital object access include DC, METS and MODS.

ArchivesSpace has also improved upon AT's digital object metadata management functionality. AT contains a digital object module that can export a MODS records of a selected object. Institutions can configure their EAD finding aids to contain links to digital objects managed in an external system (CONTENTdm, Omeka, DSpace, Islandora, etc.). However, ArchivesSpace has rectified many issues that AT users have cited such as container-level URIs not correctly exporting, as well as digital object IDs not appearing in the exported MODS file. In terms of importing digital object metadata, the ArchivesSpace digital object management tool has corrected the limitations imposed by AT on importing creator and subject metadata, along with ensuring automatic linking between digital object and resource record metadata upon import.

Tags: archivesEncoded Archival DescriptionMLIS students

"Metadata for All": Looking Back to Metadata Standards (September 28, 2015)

BY MIKE BOLAM · PUBLISHED SEPTEMBER 28, 2015 · UPDATED SEPTEMBER 21, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the introductory post.

By Eleanor Godbey and Kathleen Donahoe

Mary W. Elings and Günter Waibel's article published in 2007 in First Monday provides an easily understood review of metadata standards and systems. The article has since become a useful teaching

tool for metadata beginners from all walks of life, though it's age may be an issue for some. For this reason, we chose to look back and see how relevant the article is for metadata students today.

The article begins by introducing standards through the metaphor of bottles. Data fields are compared to the actual bottle, data content becomes the contents of the bottle, data format is the crate in which the bottles are packed, and the person, truck, plane, or ship delivering the bottles is likened to data exchange. Different standards, vocabularies, formats, and sharing languages and methods are inserted into this metaphor to illustrate how the different pieces fit together and function. After this illustration of the basics of metadata, Elings and Waibel break down the history of metadata in libraries, archives, and museums, respectively. Each history has the need for a standardization of organizing and labelling in each institution. For example, in libraries AACR, RDA, and MARC were developed to meet cataloguing needs. In archives, AMC developed as an archives specific MARC, and EAD and DACS cropped up. In museums, CDWA and CCO were developed to deal with museum collections and needs specifically. After breaking down this history, the authors look at current metadata trends, bringing up the point that a majority of changes to metadata systems are more user focused. Libraries, archives, and museums are striving to better serve the user by making their metadata systems more user-friendly. The article concludes by making the point that these different cultural institutions share and learn from each others' metadata. The differences in standards really lies with the type of record or material being described.

Elings and Waibel's article is incredibly easy to approach and is written in simple terms. It isn't too technical and the metaphor used really helps to illustrate the very basics of metadata. Speaking as beginners to metadata standards and structure, we felt that the article provided the proper balance between formalized technical thought and introductory explanation. The article did seem to shy away from pointing out any negative aspects of the different formats and schemas, which could have been helpful in deciding between different standards. There are some small mentions, such as in the discussion of the library community, where it was pointed out that the Metadata Object Description Standards (MODS) operated more efficiently than MARC.

Considering the fact that this article was written in 2007, which was more than enough time for new developments to occur in the library, museum, and archival fields, we were careful to note shifts that have occurred in the field since its publication. For example, the comments about archival shifts toward more product, less process (MPLP) have been criticized on many levels. It is now often used as a first step to make a collection useable before more item level processing occurs. Also, the article was written before huge booms in social media and online use of collections, and therefore is less able to comment on one of the bigger trends of today's world, which is creating metadata to display collections online. Overall while we believe that the article was useful to introductory students of metadata such as ourselves, the article requires some updating in order to fully educate users on the specifics of metadata in the library, museum, and archival professions today.

The article under review is available to read at http://firstmonday.org/article/view/1628/1543.

Article Citation: Elings, Mary W., and Gunter Waibel. "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums by Mary W. Elings and Gunter Waibel." First Monday 12, no. 3 (2007). http://firstmonday.org/issues/issue12_3/elings/index.html.

Tags:

oXygen Review from New Metadata Learners (October 5, 2015)

BY [MIKE BOLAM](#) · PUBLISHED OCTOBER 5, 2015 · UPDATED SEPTEMBER 22, 2015

The following post was submitted by students enrolled in LIS2407 – Metadata at the University of Pittsburgh School of Information Sciences. For more information on the series, see the [introductory post](#).

By Gesina Phillips and Christie Kliewer

As this editorial is based on our particular experience with oXygen, it is important to first provide some context for our analysis. We are two MLIS students in our final semester, during which we are taking a class focusing on metadata. The metadata course in our program is an elective (as is cataloging). The class size is far larger than is typical, as many students have chosen to take this class to gain a working familiarity with metadata standards and schema before they enter the job market.

In this course we have the benefit of being able to gain some familiarity with a range of metadata structures in a programmatic environment. In many cases, however, it seems that the primary environment for learning about metadata is on the job. This is also the case for a number of professional competencies, but given the comparatively recent emergence of metadata as a branch of librarianship, it may be challenging for those who are encountering it for the first time to find resources to develop their skills. To that end, it is important that librarians dealing with metadata rely on good tools and effective training from within the library community. Even those fortunate enough to have a metadata librarian working with them would benefit from library-specific support communities and open metadata schema and documentation.

In our attempt to review oXygen, we sought to understand whether there were other resources widely used and accepted among metadata librarians. While our limited experience with cataloging has led us to test (and struggle with) the suite of tools available through the Library of Congress and OCLC resources, we found very few practical resources for learning and exploring practical metadata. We learned to use oXygen through guided work in our course, where we used the tool to create and edit records for a digital collection. We have found oXygen to be frustrating in some ways but also very helpful in others. Overall we felt that the documentation and beginner's guides were lacking and that the field would benefit from contributing to the realm of online documentation resources.

What oXygen does get right is its constant supportive feedback regarding your work. Considering the relative youth of the metadata field, having a resource that checks whether your records are valid and well-formed without necessarily reading through every record is a huge step towards beneficial computer-assisted librarianship. The software is also very eager to offer assistance upon opening. Under the default installation settings a popup page generated when oXygen is opened offers details on oXygen events, a quick link to the user guide and discussion forums, and quick tips for successful use. The links direct to pages on the company's website for documentation. Although the guides themselves are lengthy and verbose to the point they were not helpful to us, the design hints that the company is interested in providing assistance.

oXygen is intuitive, but it has the same difficult learning curve as other expert software tools where icons are often only useful when explained by mouseover text. For many of the icon buttons, users are required to click through and experiment with the options to understand what the icon does, and often a web search is required to understand entirely what the option allows for. This type of opacity within the program underlies many of our frustrations with the resource; considering that experienced metadata librarians use this tool, we hope to see librarian-created guides in the future to support development of the field. The software is very responsive in terms of indicating whether your record is valid and well-formed, through the small red-green light found in the upper right corner of the record page, but lacks many intuitive features that would help new metadata librarians become comfortable. Our belief is that if the field seeks to benefit from open schema and standards initiatives, we need to spend time documenting the tools we rely on.

Tags: MLIS studentsXML

Of Metadata and Confidentiality (October 19, 2015)

BY CHARLES RILEY · PUBLISHED OCTOBER 19, 2015 · UPDATED OCTOBER 27, 2015

Don't join the book burners. Don't think you're going to conceal faults by concealing evidence that they ever existed. Don't be afraid to go in your library and read every book as long as that document does not offend our own ideas of decency. That should be the only censorship.

–Dwight D. Eisenhower, Dartmouth College commencement exercises, June 14, 1953

In the first part of this quote, Eisenhower's imperatives from more than sixty years ago could be seen as sounding a warning against today's environment of censorship, revisionism, and fear. He offers us an invitation to inform ourselves widely and to gain understanding, and he appears to be advocating for the benefits of a free society.

Indeed, the first part of the quote, taken by itself, would appear at first to be strongly at odds with the spirit of Section 215 of the USA PATRIOT Act, which was interpreted by the Federal Intelligence Surveillance Court to allow the FBI the ability to collect massive amounts of metadata, at least before its recent replacement by the USA Freedom Act. While Edward Snowden's revelations focused largely on telephony metadata, another category he has given attention to is library patron records. The FBI has imposed gag orders accompanying national security letters to obtain metadata; under the Patriot Act, according to the Electronic Frontier Foundation, over 300,000 national security letters have been issued.

Whether the provisions of Section 215 that relate to library records have been effectively renewed under the USA Freedom Act is an open question. Gag orders in two high-profile cases, those of the Connecticut Four and of Nicholas Merrill, have been lifted. But a climate of fear persists, and the chilling effects are evidenced by cases of official paranoia not only here, but in the UK and Ethiopia.

Long before the Patriot Act, there have been efforts to preserve the confidentiality of library patron records. The American Library Association has offered guidance on the subject since at least 1971;

forty-eight states and the District of Columbia have laws to protect patron privacy but each law is different. The other two states, Kentucky and Hawaii, have attorney general opinions addressing the issue.

The protections are subject to interpretation and in many cases are limited; their applicability has yet to be fully tested under current federal law. In lifting the gag order for Merrill, the US District Judge characterized the FBI position as "extreme and overly broad." In light of the cautionary tone of the second part of Eisenhower's quote, what is perhaps needed is a federal law to ensure harmonization between the states on setting a floor for the level of privacy protection afforded to library patrons. The recently passed CalECPA may serve as a useful model.

Summary of Comments Received on MSC Principles for Evaluating Metadata Standards (April 18, 2016)

BY JENNIFER LISS · APRIL 18, 2016

The ALCTS/LITA Metadata Standards Committee posted a second draft of the Draft Principles for Evaluating Metadata Standards on October 27, 2015, on the metaware.buzz site. This second draft incorporated feedback received on the first draft. The committee held a second open comment period from the time this draft was posted through the ALA Midwinter 2016 meeting in Boston. We received comments directly on the post itself, on a web form set up for this purpose, from the audience following a presentation on the Principles at an ALA Midwinter 2016 sessionsponsored by the ALCTS Metadata Interest Group, and at the committee's in person meeting at ALA Midwinter 2016.

This document summarizes the feedback received on the second draft in those venues.

General

We received a number of general comments on the document and its approach. Several commenters requested more detailed information on why the Principles exist and what use cases they are intended to fulfill. We will write an introduction to the Principles document that addresses these and related issues. This introduction will define the scope of the Principles to be the standards and how standards construction impacts metadata being created, rather than best practices for creation of metadata itself, recognize that the information in the Principles is a distillation for easy consumption of lessons learned by many metadata communities over time; and acknowledge that some of the ideas in the Principles document are difficult to achieve, or that indeed methods for achieving them are still being defined by various metadata communities. The Principles as such are in some sense aspirational.

We received requests for clearer definitions for some terminology such as "add value" and "network", which we will achieve through linking to an established, externally-maintained glossary, and for including examples in the document, which we will endeavor to do to illustrate application of individual principles wherever possible. We will also ensure that the document adequately covers content standards, controlled vocabularies, and system- or machine-created metadata.

Principle #1

Comments regarding "metadata standards should be part of the network," centered around the concept of "by reference" models. We made an embarrassing error in language that caused this principle to originally say the opposite of what we intended, and have corrected this. We will enhance the text of this principle to focus on standards providing context and relationships to other vocabularies as a core expression of network-connected metadata, and cite the "by reference" model as exemplifying this principle. One commenter noted that occasionally implementations might choose to store both "strings" and "things", which we will address through language in the introduction covering pragmatic considerations taking over when circumstances warrant.

Principle #2

We received a number of comments on the concept of openness in this principle, "Metadata and metadata standards should be open and reusable." We will clarify that openness here refers to the metadata and not the resource, that explicit licenses for metadata promote reusability, and that openness is tied to governance and maintenance community actions. As we are clarifying that metadata creation practices are out of scope for this document, we will further edit this principle to reflect that, and remove the recommendation that all metadata itself should be open. While committee members believe strongly in the utility of open metadata, we do agree with a commenter that the focus of this document is elsewhere.

Principle #3

We received one comment on the principle "Metadata creation should benefit user communities," requesting clarity of its meaning. We will enhance this text to indicate that depth of metadata defined by a standard should directly fulfill a real world use case, and ensure that this principle and its heading focus on the role of the standard and not the practice of metadata creation.

Principle #4

Commenters on the principle "Metadata standards should support new research methods" requested greater precision and explanation of how this would occur, and to clarify that metadata can be reused for other purposes as well. We will expand this principle to focus on flexibility of metadata and the ability to do interesting things with it, with new research methods as one example. In addition, as suggested by a commenter, we will mention formal expression of metadata models through languages such as RDF Schema, RelaxNG, or XML schema as an example of making metadata more machine-processable.

Principle #5

One commenter on "Metadata standards should have an active maintenance and governance community" pointed out that it's a stretch to claim that metadata standards in LAM institutions have

been stable for 40 years. We agree, and will adjust the text to focus more tightly on the increasing pace of change.

Principle #6

Feedback on the principle "Standards should be extensible, embeddable, and interoperable" pointed out that modularity of metadata standards is affected by differing underlying intellectual models between standards. We agree, and will update the language here to reflect that.

Principle #7

Comments on "Metadata standards should follow the rules of 'graceful degradation' and 'responsive design'" focused on two areas. First, a commenter noted that graceful degradation is balanced by the concept of progressive enhancement. We will consider including both here, perhaps referencing Tim Berners-Lee's progressive levels of implementation for Linked Data. Second, a commenter requested greater clarity of meaning for the phrase "more sophisticated uses/ontological inferencing." We will endeavor to achieve this in a revision.

New Ideas

A few new ideas emerged from comments that we will incorporate into the Principlesdocument. First, we will cover the role of documentation for metadata standards, that they should provide sufficient information about the standard, be useful and usable, have examples, and be friendly to multilingualism. Second, we will introduce the concept of inherent bias within any human-created construct, recognizing that good metadata embraces diverse viewpoints and that metadata structures themselves likely encode some (typically unconscious) biases of their creators. While we cannot eliminate this bias, we can raise awareness of it. To assist with effective presentation of these issues, we will consult with ALA's various ethnic caucuses to review draft language. Finally, privacy issues surrounding metadata implementations are significant enough to deserve mention in this document. We will treat de-identifying metadata as a use case for the graceful degradation principle (#7).

We thank all who took the time to comment on this second draft of the Principles for Evaluating Metadata Standards. The comments are extremely useful and will result in a stronger final product. The Metadata Standards Committee is currently working on making the changes described here and in doing a final review of the document. We anticipate releasing a new version of the Principles in spring 2016.

Tags: standards assessmentstandards development

Bridging the Gap between Metadata Librarians and Art Conservators (August 1, 2016)

BY PEGGY GRIESINGER · PUBLISHED AUGUST 1, 2016 · UPDATED AUGUST 3, 2016

During the summer of 2014, I started a nine-month residency at the Museum of Modern Art (MoMA) in New York City. This project was part of the National Digital Stewardship Residency in New York, a program that takes recent graduates from LAM programs and places them at cultural heritage host

institutions to work on digital preservation projects. I will not go into extensive detail about my project here (you can read through my residency blog [here](#) to get more information), but the basic goal of the project was to create a metadata profile to record the conservation activities performed by museum conservators on time-based media artworks. Time-based media is a sort of catch-all term that refers to materials like film, software, and slides that "have duration as a dimension and unfold to the viewer over time" ([Guggenheim Conservation Department: Time-Based Media](#)).

When I entered MoMA on the first day, I could have filled a multi-volume set of books with what I did not know about museums, conservation, and working with museum professionals. I had no idea if the information architectures of our respective professions would match up enough to have coherent conversations, or if we would constantly be explaining basic tenets of our profession to each other. I was unsure to what degree metadata was understood and used in the museum world, and what kinds of systems museum professionals were using to interact with and store metadata. I also had almost no knowledge of audiovisual materials and the type of technical information needed to properly care for these types of materials, especially when they must be preserved as works of art (meaning maintaining aesthetic authenticity was hugely important). In short, I had a lot to learn.

To my relief, I found conservators to be willing and excited to participate in an information exchange with the traditional library world; we both had much to learn from each other, after all. Conservators helped me to understand their domain, and this in turn allowed me to translate that domain into a metadata profile that would adhere to digital preservation standards. So how did I bridge this gap between metadata librarians and museum conservators? In three ways: 1) Embedding myself in the domain of museum conservation; 2) Learning to appreciate the domain of fine art and conservation work; and 3) Translating everything I said out of library jargon.

The first of these, embedding in the domain, was greatly helped by the fact that my workspace was in MoMA's painting and sculpture conservation lab. This meant that I was able to interact with the conservators, watch their work, and observe their documentation process as it existed. I also shadowed the media conservators through the digitization/transfer process of one artwork. This gave me a clear sense of their process and the significant properties they needed to be able to record in a metadata profile. The conservators also walked me through current exhibits at MoMA and other museums to show me their work in context, so I could understand how the information in the metadata profile would actually assist conservators in the future to most accurately exhibit an artwork. Finally, I attended any meeting I could that was at all relevant to my project, allowing me to slowly absorb the museum terminology that I would need to know to create a useful metadata profile.

The second way I reached across the aisle to museum conservators was through really exposing myself to time-based media artworks, and gaining an understanding of why these pieces are so special that a person would devote their career to preserving them. I had not been particularly interested in modern art in general before this residency, and had very little knowledge of time-based media art. It was not until I saw these pieces on exhibit that I was fully able to appreciate their uniqueness and importance, and also what type of information would need to be recorded to ensure that these pieces could be exhibited this effectively in the future. Pieces like Feng Mengbo's [Long March: Restart(2008)](#), an interactive video game that surrounds the viewer on both sides like a tunnel, made it clear how innovative and affecting time-based media art could be. This type of art also demonstrated just how difficult the task of properly preserving these types of works would be, and how important it would be to

create a metadata profile that was both comprehensive and extensible, able to adjust to the new innovations sure to come in the future of time-based media.

Lastly, I translated everything out of library jargon, and emphasized how what I was creating could practically help them with their work, rather than focusing on the importance of digital preservation and good metadata practices from my perspective. These are proudly held beliefs in the library world, but can sometimes over-complicate a pitch for improved documentation to museum professionals. Some conservators are certainly interested and invested in digital preservation, but most are simply looking for a way to ease the burden of very heavy, time-sensitive workloads; explaining how a concise metadata profile could save them time now and in the future is a useful way to get conservators invested in such projects.

Museum conservators share many common interests with metadata librarians, but it can be difficult to see that from a cursory look at these two cultural heritage domains and their current metadata and digital preservation practices. My time at MoMA creating a metadata profile for time-based media conservation demonstrated that much of the work we do in the library metadata world translates very usefully to the museum conservation world; it is just a matter of knowing how to speak to each other to reach our shared goals that stands in the way of more comprehensive collaboration. I hope that this post will offer some useful guidance about overcoming these barriers to those considering such collaboration across museum-library borders.

This post was adapted from a presentation given at ALA Midwinter 2016 during the ALCTS CaMMS Cataloging Norms Interest Group meeting.

Tags: [digital stewardshipmuseumspreservation metadata](#)

Principles for Evaluating Metadata Standards (August 4, 2016)

BY [JENNIFER LISS](#) · AUGUST 4, 2016

The [ALCTS/LITA Metadata Standards Committee](#) has developed the Principles for Evaluating Metadata Standards for use by the library, archives, and museum (LAM) community. The principles were initially developed as an internal document as a distillation of the many lessons learned by metadata communities over time; however, the committee soon realized they could be useful to a broader audience.

The principles are intended to inform and support the development, maintenance, selection, and assessment of metadata standards. They may be applied to metadata structures (field lists, property definitions, etc.) and with content standards, [controlled vocabularies](#), and standards intended for both system- and machine-created metadata.

It is the hope of this committee that these principles can help guide the refinement of current metadata standards, and the creation of new metadata structures and vocabularies. These principles are aspirational and may not be fully implementable as a pragmatic concern for all standards.

1. Metadata standards should be part of a shared data network

Metadata—its standards, systems, and services—is most efficient when connected with the broader network of information as manifested in the W3C's [Linked Data](#) initiatives and related data sharing efforts. Metadata standards should provide mechanisms for expressing relationships to other vocabularies. For example, standards that allow for the use of URIs, in addition to or instead of relying solely upon strings of text, may connect information from different sources, support distribution to indexing and research services, and increase resource visibility on the web.

2. Metadata standards should be open and reusable

Open metadata is a foundational building block of information systems and computationally-enhanced research. Metadata standards and associated vocabularies need to be open for use and re-use, free of charge, and managed with openness and transparency. Metadata standards governance and maintenance bodies should consider explicit licenses that promote reusability of standards and vocabularies.

3. Metadata standards and creation guidelines should benefit user communities

The volume of information in a web-enabled world, along with the ability of information systems to analyze and index digital objects themselves, changes the value of traditional metadata. Metadata standards should be created with the intended audience, its related communities, and the public in mind. The depth and granularity defined by metadata standards should prioritize data elements that support clear use cases defined and contributed by their user communities and allow implementers to use only features relevant to their needs.

4. Metadata standards should support creative applications

Up to the 21st Century, bibliographic metadata has supported a relatively narrow vein of research, involving reading, viewing, listening to, or otherwise engaging with individual resources by humans. As new research methods emerge (e.g., computational linguistics, computational bibliometrics, linguistic analysis, network analysis), and as new data modeling methods geared toward machine actionability mature, metadata standards and exchange/access methods should easily support evolving modeling and use practices. Encoding of metadata standards in machine-processable formats such as [RelaxNG](#), [XML Schema](#), or [RDF Schema](#) can help to promote computational use of metadata in emerging fields.

5. Metadata standards should have an active maintenance and governance community

Metadata standards are only as valuable and current as their communities of practice, which are changing with increasing rapidity. The needs of implementers should be a driving force for the development of standards over time. Governance bodies should support transparent, timely development and revision processes by publishing standards development and revision protocols. Such protocols should include mechanisms for community contributions to draft standards. Governance

bodies should change and develop over time to reflect the diversity of their current and potential user communities.

6. Metadata standards should be extensible, embeddable, and interoperable

Central to the successful deployment of a metadata standard is its ability to be used effectively alongside other standards. Standards should be designed in a modular way to allow relevant parts to be incorporated into local systems together with parts of standards from other sources, with each given equal preference. Relatively complex standards should provide "lite" style implementation options, to allow for wider adoption and utility, even if this means lossy data transfer in some cases. "Lite" style implementations are helpful in enhancing interoperability of metadata standards that have differing underlying intellectual models.

7. Metadata standards should follow the rules of "graceful degradation" and "progressive enhancement"

Metadata standards should be designed such that, as the complexity of the standard is stripped away, the metadata degrades gracefully. Removing complexity, for example to make metadata that is encoded according to a robust standard useful to a less specialized audience or to remove sensitive information, should result in metadata that is still useful. Metadata standards that are designed from a [progressive enhancement](#) approach prioritize simplicity while being flexible enough to support enrichment of the standard to accommodate specialized metadata use cases.

8. Metadata standards should be documented

Alongside the metadata standard, information about its use in practice, examples of the standard as applied to a real-life resource, its governance and management structure, rationale behind design decisions, and its connections to metadata standards from other domains should be published. Efforts should be made to allow easy adaptation of documentation for multilingual environments and to make the documentation understandable to implementers from other fields. Applying a license to the documentation will clarify its use and reuse.

9. Metadata standards should be inclusive and transparent about historical and cultural biases

Metadata standards development is not neutral; human beings unavoidably assign value judgments when making (and not making) assertions about a resource, and in defining the assertions that can be made about a resource. Metadata standards developers should be aware of these value judgments, make them explicit to the degree possible, and take as a guiding principle not neutrality, but rather inclusivity of worldviews. A diversified team approach can be considered in the creation, implementation, and further enhancement of the metadata standards. Metadata standards and vocabularies should reflect changes in language.

Tags: [standards assessment](#)standards development

Metadata evaluation – NISO STS Draft comment (May 22, 2017)

BY ERIK · MAY 22, 2017

Introduction & Context

The American Library Association's Metadata Standards Committee (MSC) contains representation from three of its topical Divisions: the Association for Library Collections & Technical Services, the Library Information Technology Association, and the Reference & User Services Association. The MSC welcomes the opportunity for public comment on the draft NISO Z39.102-201x, STS: Standards Tag Suite.

The MSC has published a set of "Principles for Evaluating Metadata Standards," http://metaware.buzz/2016/08/04/principles-for-evaluating-metadata-standards/, which we have used to structure our comments. Please note that our principles are tailored more towards structured metadata, rather than markup of full text documents. Nevertheless, we hope that our comments are useful to the NISO STS committee.

1. Metadata standards should be part of a shared data network

As a document markup language defined as an XML DTD, W3C XML Schema, and RelaxNG schemas, it would be difficult to expose data encoded in NISO STS as Linked Data on the Web. For example, the current NISO STS element and attribute definitions do not seem to make a systematic effort to provide places to encode URIs for named entities. Given the Web community's focus on Linked Data today, the NISO STS standard would benefit from some thought in how to make the standard a bit more Linked Data friendly.

We note that most, if not all, of the NISO STS document markup could be done in the Text Encoding Initiative (TEI) community standard. We recognize that TEI is not nationally or internationally formally standardized and there is, in many communities, a strong need for a more official standard for encoding of standards documentation. However, the TEI does have wide adoption and use among certain communities, and clear statements on the relationship between the two would benefit both current and potential user communities for both languages, and promote linking between them.

Crosswalks for descriptive metadata features of NISO STS to other popular metadata standards, especially in the bibliographic realm, would promote use of STS encoded documents by other communities. We do note that the decision to not require a default namespace in the interests of retaining backwards compatibility does somewhat limit its ability to play nicely with metadata in other namespaces.

2. Metadata standards should be open and reusable

NISO makes its standards open for reading and download, and that is appropriate for this standard. Making the standard open is an important part of encouraging adoption and use.

It is slightly unclear to us what the exact relationship is between the ISO STS, version 1.1 and this NISO STS draft. The NISO standard would be easier to select, adopt, and use if this relationship and the differences between the two were made clearer in the documentation.

As with all NISO standards, the governance and oversight of STS is clear and structured through regular NISO procedures, which promotes confidence in the standard. The detailed documentation in the NISO "workroom" for this standard further makes it accessible.

We note that NISO standards are issued as copyrighted documents as a matter of course, which comes with an implicit statement that all rights are reserved. A more open license for these standards documents themselves would increase their utility to the community. The DTDs, XML Schemas, and RelaxNG schemas would also benefit from explicit and permissive licensing terms.

3. Metadata standards and creation guidelines should benefit user communities

Per the comment in item 6, below, the standard may benefit from use case exploration that would develop approaches to lighter-weight use cases. The general introduction does provide an overview of the defined use case "standards bodies, standards producing organizations, publishers, commercial vendors, and archives can publish and exchange standards documents." Are there other use cases or other user communities that would benefit and whose needs should be considered? For example, a discipline-based community formalizing its own metadata standard might not even know NISO STS exists, or if it does, would likely find it difficult to implement due to its size and complexity. It might be useful to consider how to make NISO STS easier to implement for use cases such as these.

4 . Metadata standards should support creative applications

Modern metadata standards are at their best when they promote computational and other derivative uses of the data and documents encoded in them. The definition of NISO STS in three formats — DTD, W3C XML Schema, and RelaxNG — increases flexibility in implementation, allowing integration into multiple types of technical platforms. This formal document structure easily promotes more advanced applications such as text mining. While definition of modern standards in RDF-friendly technologies is desirable to participate in Linked Data communities, we recognize this is difficult due to STS's nature as a document markup language.

5. Metadata standards should have an active maintenance and governance community

NISO's open, community-based, and formal governance and revision practices foster active engagement with standards such as this one, and this approach is commendable. The nature of STS as a new standard formalizing long-held practices among a stable implementation community speaks well to its ability to remain relevant over time.

6. Metadata standards should be extensible, embeddable, and interoperable

The incorporation of MathML into NISO STS is an advantage, as it does not re-implement the features of MathML in a new standard, but rather relies on a specialist community to maintain in its area of expertise. We note that NISO STS does make its own definitions of metadata features that other

established communities have standards for, including bibliographic information, text formatting, geographic places, rights information, and names.

NISO STS is a very large and complex standard. Presumably implementation of such a standard requires significant expertise and resources. Providing both the Interchange and Existing Tag Sets, while adding to the standard's flexibility, also increases its complexity. While we note that the "Scope" section in the STS draft for comment indicates that the tag set may be restricted to meet the needs of a given project, no formal mechanism is given for doing this, and as such does not meet our definition of "extensible". The committee might consider an official "lite" version that is more accessible, to promote wider adoption of the standard.

It appears that some elements have recommended vocabulary and abbreviations (e.g. http://niso-sts.org/TagLibrary/niso-sts-TL-0-2-html/element/doc-type.html) that do not correspond to external standards, but are rather embedded in the DTD/Schema for STS. It may be that defining external vocabularies at this level is impossible for this specific standard but could be a challenge to future stability of the standard. There are multiple ways in which a standard community could engage with this issue. By further defining the vocabulary this specific standard could make contributions that other standards might adopt.

7. Metadata standards should follow the rules of "graceful degradation" and "progressive enhancement"

As a markup language for standards documents, rather than only a metadata format, graceful degradation of the STS standard is difficult. Some features, such as <std-doc-meta> do make it conceivable that some data might be extracted (here, a bibliographic record for the standard), but the committee might consider what other use cases there are for automatically extracting data from STS and ensure the standard is structured in such a way to support this.

The principle of progressive enhancement relies upon a design whereby a format starts at a relatively simple state and then allows complexity to be added as needed. The design of STS as a pair (Interchange and Extended) of DTDs, W3C XML Schemas, and RelaxNG schemas containing all allowed elements and attributes does not embody this principle. Other implementation strategies, even with these technologies, could be used to make the vocabulary more modular, such that implementers could choose only features in categories useful to them. See, for example, customization options for the TEI as one method by which this might be done.

8. Metadata standards should be documented

In addition to the primary alphabetical listing of STS elements and attributes, at http://niso-sts.org, there is more helpful documentation available about the standard. The hierarchical view of the schema is especially helpful in enabling those new to the standard to more easily learn it. This page also offers helpful examples of elements and attributes used in context. Documentation grouping elements by theme would have benefited us as readers to be able to more quickly understand the features that the standard covers. In addition, full encoded examples would be helpful.

9. Metadata standards should be inclusive and transparent about historical and cultural biases

The <glyph-data> and <glyph-ref> elements provide useful means for expanding the content of an STS encoded document beyond Unicode characters.

We note that the content model for <name>, with its sub-elements for <surname> and <given-names>, may not work for all cultural traditions. <surname> is particularly problematic as a "family" name is not always a "surname", and even then a "family" name is not a universal construct.

Similarly, the content model for <address> seems to assume Western-style addresses.

The standard includes a few elements that begin with the word "trans-". We suggest either using the full term "translated" or a different abbreviation, to avoid using a word that carries with it other cultural implications in the English language, as a way of showing sensitivity to the LGBTQ community.

Other issues

@xlink:href is listed as an available attribute on <media> in the PDF but not the tag library
http://niso-sts.org/TagLibrary/niso-sts-TL-0-2-html/attribute/xlink-href.html

MSC / Metadata Interest Group presentation ALA 2017 (June 25, 2017)

BY ERIK · JUNE 25, 2017

At ALA Annual, 2017 Jenn Riley, Lauren Corbett, Erik Mitchell and Michael Bolam presented the case study from the MSC evaluation of NISO standard tag suite standard at the ALCTS Metadata Interst Group meeting

The group shared their experience from applying the evaluation standards against the NISO tag suite as well as lessons learned and thinking about how to approach this type of work in the future. Links to the presentation and evaluation guidelines are included below.

The discussion within the IG presentation focused on how the principles could be applied and how the community overall could provide feedback to standards communities. One idea from the floor was to have some type of coordinated event, a webinar or other similar meeting, to provide background and context on the standard. The discussion highlighted the value of targeted as well as comprehensive feedback and reaffirmed the notion that having a group purposefully provide feedback is a step in the right direction. Suggested future standards to evaluate include an upcoming DPLA application profile refresh.

The discussion also included a reference to a recently released NISO technical report that explored, among other things, the notion of semantic versioning and how such an idea could influence principle 7 (progressive enhancement / graceful degradation)

Presentation

MSC feedback on DPLA metadata application profile update (4.1) (November 1, 2017)

BY [ERIK](#) · NOVEMBER 1, 2017

Summary of MSC feedback to DPLA

The Metadata Standards Committee (MSC) sees that the standard is not a standard for creating new metadata but a reformatting of the metadata that DPLA is harvesting. As this metadata is not intended for human creation but rather automatic mapping we believe that the appropriate level of feedback is on the specific changes proposed in the 4.1 revision. Overall, the application profile approach of DPLA aligns well with the principles that the MSC promotes. Given this close alignment the feedback below does not dive into detail where there is good alignment.

The principles highlighted by the MSC focus on higher level issues related to metadata including design and use implications. As such the feedback provided here does not dive into the technical details and changes in the MAP 4.1 revision.

Overall, the recommendations below affirm the direction of the 4.1 MAP and the overall approach of DPLA in developing metadata standards.

1. Metadata standards should be part of a shared data network

Principle context: Metadata—its standards, systems, and services—is most efficient when connected with the broader network of information as manifested in the W3C's [Linked Data](#) initiatives and related data sharing efforts. Metadata standards should provide mechanisms for expressing relationships to other vocabularies. For example, standards that allow for the use of URIs, in addition to or instead of relying solely upon strings of text, may connect information from different sources, support distribution to indexing and research services, and increase resource visibility on the web.

Feedback: The DPLA, almost by definition is a shared data network and this application profile meets this principle. No specific feedback provided here is specifically related to the 4.1 update. The fact that DPLA relies on existing resources (e.g. the EDM, MODS and other schemas and vocabularies) is viewed as a positive trend.

2. Metadata standards should be open and reusable

Principle context: Open metadata is a foundational building block of information systems and computationally-enhanced research. Metadata standards and associated vocabularies need to be open for use and re-use, free of charge, and managed with openness and transparency. Metadata standards governance and maintenance bodies should consider explicit licenses that promote reusability of standards and vocabularies.

Feedback: The DPLA's use of namespaced vocabularies promote reusability and computational analysis. The fact that the MAP is published along with supporting documents with alternative

descriptive information are viewed as a good mechanism to promote this make it easier to re-use. We note that in some cases, the MAP collects data as literal values without a specific definition of the source vocabulary (e.g. the hierarchicalGeographic). We wondered if there were cases where content providers would provide URIs rather than literal values and if this would influence the behavior of the schema (e.g. would URIs be de-referenced before inclusion). Likewise, although not necessary, we wondered if some statement around the ability to re-use the metadata standard itself would be helpful.

3. Metadata standards and creation guidelines should benefit user communities

Principle context: The volume of information in a web-enabled world, along with the ability of information systems to analyze and index digital objects themselves, changes the value of traditional metadata. Metadata standards should be created with the intended audience, its related communities, and the public in mind. The depth and granularity defined by metadata standards should prioritize data elements that support clear use cases defined and contributed by their user communities and allow implementers to use only features relevant to their needs.

Feedback: The focus of the MAP on consolidating multiple metadata schemas into a single shared platform is perceived as a positive step towards sharing and community use. By using standards and platforms that LAM communities are familiar with (e.g. DC, DCTERMS, MODS, EDM) we believe that the MAP is positioned to be more successful than it would with an independent schema or vocabulary. Specifically separating licenses from rights statements is viewed as a positive development in the MAP. We are unsure if there are types of objects that are more likely to have licensing rather than rights however and collectively had difficulty identifying clear use cases for this more granular metadata. As such more definition on the intended use or role of the license and rights fields could be useful. The group debated for example if there were cases where a rights and license statement would be used together rather than in some hierarchical fashion. Perhaps one argument against this approach is that having separate fields in this case could introduce complications around derivative uses.

4 . Metadata standards should support creative applications

Principle context: Up to the 21st Century, bibliographic metadata has supported a relatively narrow vein of research, involving reading, viewing, listening to, or otherwise engaging with individual resources by humans. As new research methods emerge (e.g., computational linguistics, computational bibliometrics, linguistic analysis, network analysis), and as new data modeling methods geared toward machine actionability mature, metadata standards and exchange/access methods should easily support evolving modeling and use practices. Encoding of metadata standards in machine-processable formats such as RelaxNG, XML Schema, or RDF Schema can help to promote computational use of metadata in emerging fields.

Feedback: The 4.1 MAP version appears to extend and improve creative application and extended uses. We note that the standard uses RDF but could not readily find information on how the standard is made available (e.g. we could only find the PDF and excel version). A published schema in RDF, XML

or RelaxNG might be helpful. Likewise ensuring that there is a URI that points back to the full metadata is helpful for creative applications.

5. Metadata standards should have an active maintenance and governance community

Principle context: Metadata standards are only as valuable and current as their communities of practice, which are changing with increasing rapidity. The needs of implementers should be a driving force for the development of standards over time. Governance bodies should support transparent, timely development and revision processes by publishing standards development and revision protocols. Such protocols should include mechanisms for community contributions to draft standards. Governance bodies should change and develop over time to reflect the diversity of their current and potential user communities.

Feedback: The DPLA is not only active but appears to maintain ties to other active communities. The MSC notes that the call for review was widely distributed and that the standard itself had received feedback in the form of comments. It would appear from these activities that DPLA is not just active but actively contributing to the standards.

6. Metadata standards should be extensible, embeddable, and interoperable

Principle context: Central to the successful deployment of a metadata standard is its ability to be used effectively alongside other standards. Standards should be designed in a modular way to allow relevant parts to be incorporated into local systems together with parts of standards from other sources, with each given equal preference. Relatively complex standards should provide "lite" style implementation options, to allow for wider adoption and utility, even if this means lossy data transfer in some cases. "Lite" style implementations are helpful in enhancing interoperability of metadata standards that have differing underlying intellectual models.

Feedback: We note that the MAP itself is a "lite" version of the metadata that DPLA might have to work with and, from the committee's read, appeared to strike a good balance between complexity and adaptability. The fact that a large amount of the terms map to some external vocabulary is viewed as positive and the reliance on optional metadata we believe is a good mechanism to spur adoption and participation. It could be that establishing a rights vocabulary (if possible) would enable higher scale derivative uses or simplify computational decision making around derivative use. One potential method would be to list potential vocabularies to encourage adoption (e.g. rightsstatement.org)

7. Metadata standards should follow the rules of "graceful degradation" and "progressive enhancement"

Principle context: Metadata standards should be designed such that, as the complexity of the standard is stripped away, the metadata degrades gracefully. Removing complexity, for example to make

metadata that is encoded according to a robust standard useful to a less specialized audience or to remove sensitive information, should result in metadata that is still useful. Metadata standards that are designed from a [progressive enhancement](#) approach prioritize simplicity while being flexible enough to support enrichment of the standard to accommodate specialized metadata use cases.

Feedback: The MAP appears to follow this guideline.

8. Metadata standards should be documented

Principle context: Alongside the metadata standard, information about its use in practice, examples of the standard as applied to a real-life resource, its governance and management structure, rationale behind design decisions, and its connections to metadata standards from other domains should be published. Efforts should be made to allow easy adaptation of documentation for multilingual environments and to make the documentation understandable to implementers from other fields. Applying a license to the documentation will clarify its use and reuse.

Feedback: We believe that it would be useful to make recommendations for vocabularies, even if they are not required (e.g. we note that AAT is cited as one vocabulary but many other elements would benefit from vocabulary recommendations as well). This might create an opportunity to support LAMS and encourage overall higher quality metadata by including these recommendations, even if not as a specific part of the MAP.

9. Metadata standards should be inclusive and transparent about historical and cultural biases

Principle context: Metadata standards development is not neutral; human beings unavoidably assign value judgments when making (and not making) assertions about a resource, and in defining the assertions that can be made about a resource. Metadata standards developers should be aware of these value judgments, make them explicit to the degree possible, and take as a guiding principle not neutrality, but rather inclusivity of worldviews. A diversified team approach can be considered in the creation, implementation, and further enhancement of the metadata standards. Metadata standards and vocabularies should reflect changes in language.

Feedback: We note that as a consolidated application profile the risk of loss of context is a key issue around inclusivity and transparency. Providing access back to the object and source metadata is a good mechanism to mitigate this risk. We also note that a lack of vocabulary recommendations or requirements is useful in encouraging and supporting a wide range of adoption and actually may help address this challenge. Finally, by having so few requirements, this MAP opens the number of communities who can participate in the DPLA. We believe that this could further encourage participation in metadata aggregation and collaboration across libraries, especially for LAM institutions of multiple sizes and resource scales. Given that the focus of the DPLA is America, some Western-centric orientation in the geographic portions is sensible, but the committee wondered if there

were alternative viewpoints (Native American or Inuit, for example) that have been overlooked in forming the structure of the metadata schema or vocabularies.

Metadata priorities and focus profile: National Library Service (December 13, 2017)

BY [ERIK](#) · DECEMBER 13, 2017

Introduction

The Metadata standards committee spoke with [Karen Keninger](#), Director of the National Library Service in October 2017 with the goal of learning more about the current focus and interest areas around metadata. A critical area of exploration included questions around how the areas of diversity, inclusivity and accessibility were factoring into metadata decisions.

In her role as the Director of the NLS, Kenninger oversees and directs programs that provide library services to blind or otherwise disabled. The service involves approximately 104 libraries around the country that collaborate in the provision of this service.

About the NLS

The NLS is a free public library service for residents of the United States who are not able to read standard print. Primarily, the NLS provides talking books and braille books for readers with a general demand that often focuses on fiction and general interest non-fiction with special collections around accessibility. One unique area of focus addresses a need for braille music. Kenninger notes that this an area where the NLS is seeking to provide some type of repository to facilitate access to this content.

The users of the NLS are quite often blind or visually impaired (approximately 86% of users). Users are often over 60 (approximately 60%). In part this is driven by the fact that there are many age-related vision loss. Critical interest areas include general reading, civic engagement and historical information. There is a growing K-12 users and organizations such as Bookshare also do a good job of serving the needs of these users.

Metadata background for the NLS

The NLS primarily uses the catalog of the Library of Congress to manage and provide access to the collection. As a result the collection is cataloged using MARC and MARC-related standards. A critical need that has emerged in the past few years has been the need to be able to share more information about accessible resources internationally. In relation specifically to the work being inspired from the Marrakech treaty, building additional support for cross-country discovery of and access to shared collections is increasingly important. Specifically being able to improve discovery of and access to accessible content as well as enabling more detailed discovery through more detailed metadata. Critical challenges here include metadata interoperability and the ability to work with metadata and materials in multiple languages.

Critical work in diversity, inclusion and accessibility

By definition, the NLS focuses on accessibility issues for a significant community. A major metadata need in this area is more granular need in the description of types of accessibility. The [Accessible Books consortium](#) of the World Intellectual Property Organization (WIPO) is building a shared catalog of all member materials for accessible content. Such platforms would be greatly improved with additional format, narration and other types of content messages.

NLS focus and future work

Greater collaboration with other similarly focused organizations is on the radar for NLS as the landscape of information evolves and as new formats emerge. One area of emphasis is the exploration of an authentication service that would lower barriers of access to eligible users. In this distributed network users would be able to use a single authentication service to sign into a variety of information resource services so that a user, once validated as an eligible user, would not need to re-establish that relationship with each information provider. Such a service would significantly reduce the barriers that disabled users currently have to information.

Tags: [accessibilitydiversityinclusionNational Library Service](#)

Metadata priorities and focus profile: Program for Cooperative Cataloging (December 13, 2017)

BY [ERIK](#) · DECEMBER 13, 2017

Introduction

The Metadata standards committee spoke with Xiaoli Li (Chair Elect, PCC Policy Committee), Kate Harcourt (Former chair PCC Policy Committee), Lori Robare (Chair, PCC Policy Committee) and Matthew Haugen (Participant, Task Group on Gender in Name Authority Records). These members of the PCC spoke with MSC in October 2017 with the goal to share current work related to metadata. A critical area of exploration included questions around how the areas of diversity, inclusivity and accessibility were factoring into metadata decisions.

In their various roles in the PCC Policy Committee and in partnership with the many members of PCC, Li, Harcourt, Robare, and Haugen are responsible for advancing policy and practice of PCC.

About PCC

The Program for Cooperative Cataloging (PCC) is one of the nation's leading standards bodies related to the cooperative cataloging and sharing of MARC-related metadata standards. The PCC's mission is to create and refine metadata to meet specific user needs. According to their mission statement they accomplish this through advocacy, training, outreach, and best practice definition for the library community.   More information about specific work is available at the PCC website at: http://www.loc.gov/aba/pcc/about/.

The primary focus of the PCC is on establishing practice around cataloging standards and ensuring that the library community develops the expertise to implement these practices. As part of that, the PCC Policy Committee provides policy guidance for descriptive cataloging, especially as these policies related to metadata technologies. Working across a wide range of metadata, the PCC influences bibliographic, serial cataloging, name, authority and other similar vocabularies and classifications.

The PCC primarily works with catalogers and other organizations with an emphasis on metadata in the library community. This includes collaboration with OCLC, ALA and sub groups within both organizations. Somewhat recently PCC has explored collaboration with national projects such as LD4P, especially as these projects examine issues central to the PCC mission (e.g. a focus on standards setting and training).

PCC focus and future work

The PCC is engaged in a strategic planning process in the Fall 2017. Major issues under consideration include thinking about the "critical path" for metadata moving forward, how ontology, vocabulary and identities management is changing in a linked data world, how standards definition and management will change in an environment which is likely to be more distributed.

Two critical areas identified by the PCC include how to approach the training for and implementation of linked data and how to appropriately engage in identity management in this environment. A related issue is the natural transition from a "record-based" model under MARC to a "statement-based" model in linked data. A PCC task group was charged to study the issues surrounding work entities that emerge in various communities and to propose feasible options to advance the provision of work-level metadata:
http://www.loc.gov/aba/pcc/documents/PoCo-2017/WorkEntitity%20Preliminary%20White%20Paper-2017-09-27.pdf

For past few years PCC has had a group studying how to insert URIs into MARC records. This has been important work given the ongoing transition from MARC to linked data models and the expectation is that this work will facilitate migration and management of metadata across multiple schemas and encoding technologies.

Critical work in diversity, inclusion and accessibility

As the metadata world moves away from MARC standards, it is important to the PCC to explore how a broader array of vocabularies and ontologies create opportunities to represent communities and areas of emphasis in more granular ways. The PCC Task Group on Gender in Name Authorities for example found that the RDA standard originally specified three gender terms (male, female, not known). The RDA Steering Committee removed that pre-defined vocabulary in 2015, the result of an attempt to add transgender to the list of "male, female, not known." In the absence of a set vocabulary, PCC found that there was some confusion about what catalogers should record (if anything). In response, PCC formed a task group which developed recommendations for best practices for recording gender, contained in the report https://www.loc.gov/aba/pcc/documents/Gender_375%20field_RecommendationReport.pdf.

In addition the report recommended other areas of focus, on guidance on how to approach the gender identification of authors including issues of perception and privacy. In the future PCC may extend this work to examine other vocabularies and how well they serve current needs.

As a standards body focused on metadata issues the PCC has invested fairly heavily in thinking around the detailed implications of shifts in vocabulary management. There are clear cases for example where community norms around the use of language to describe groups change, introducing a need for the library community to adjust its use of standards accordingly. In addition, the practice of identifying which attributes to gather about authors, people are topics itself requires consideration in relation to diversity, inclusion and accessibility. Addressing these issues is important to PCC given its focus as a group that helps metadata be more trustworthy and useful.

In recent years PCC has focused specifically on BIBFRAME with two working groups, one focused on how the BIBFRAME vocabulary could be used for cataloging monographs and another working group focused on serials. Thinking about the complex relationship between expertise, training and adoption is central to PCC's work in the coming years.

Tags: [accessibilitydiversityinclusionPCC](#)

Still on the MARC While Diving into Next Gen Cataloging and New Standards (January 31, 2018)

BY [LAUREN E. CORBETT](#) · JANUARY 31, 2018

During the fall of 2017, the Metadata Standards Committee surveyed a variety of organizations deeply immersed in the metadata needs of libraries, archives and museums (LAMs). Two vendors who responded, Backstage Library Works (specifically, Nate Cothran, President, and Kate Clayborne, Metadata Services Product Manager) and Casalini Libri (Michele Casalini, Managing Director), are meeting the needs of their clients by primarily working with MARC. As would be expected, they both pay close attention to the Library of Congress (LC) and the Program for Cooperative Cataloging when interpreting standards. Keeping up with the move from AACR2 to the new standard, Resource Description and Access (RDA), Backstage became involved as an official test partner in LC's RDA pilot program and Casalini participates in the European RDA Interest Group (EURIG). Both companies are highly interested in BIBFRAME.

With over 25 years of experience, Backstage Library Works specializes in authority control, cataloging, reclassification, retrospective conversion, digitization, preservation microfilm, and on-site services, and has served libraries of all sizes and types, across the globe. Backstage has also "branched out into supporting several different kinds of XML metadata, such as TEI, EAD, DublinCore, CONTENTdm, and a host of other local library-generated XML. Backstage is actively investigating BIBFRAME and has started to delve into JSON in order to simplify metadata ingestion and exports." Backstage also uses the METS standard with the Analyzed Layout and Text Object (ALTO) XML Schema in digitization services. Backstage currently provides linked data enrichment to MARC 21 records through means such as adding $0 to existing access points during authority control processing, and is discussing other ways to help libraries move into the use of linked data.

Casalini Libri was founded by Mario Casalini in the late 1950s, and provides full bibliographic and authority data, integrated services for the management of acquisitions, supply of books and periodicals, and full-text content for over 250 publishers, particularly in the Humanities and Social Sciences. Regarding metadata, Casalini Libri has a focus on "the production of metadata for all resource type publications in a standardized form, and the conversion of existing records to the same standardized metadata format. The company also works with defining and attributing identifiers, such as DOI and ISNI." Casalini Libri works with MODS for digital resources. As an active partner in the SHARE-VDE project (Share Virtual Discovery Environment in Linked Data, www.share-vde.org), Casalini Libri currently "is experimenting with enrichment procedures for a more precise conversion into Linked Open Data, reconciliation and conversion processes, full adherence of the RDA guidelines for the creation of 'well-done' data, and the publication in a three-layered platform as one of the first examples of the practical application of BIBFRAME." Casalini plans to offer a number of new services this year including URI registry, MARC enrichment with the preferred URIs for various entity types, and entity reconciliation as well as conversion into BIBFRAME and data publication applying the BIBFRAME model.

Tags: BIBFRAMElinked dataMARC

Metadata priorities and focus profile: VIAF (February 5, 2018)

BY NADINE ELLERO · FEBRUARY 5, 2018

The Metadata standards committee exchanged emails with Hayley Moreno (OCLC-VIAF Liaison, Database Specialist II), who in turn shared our Outreach Survey with her colleagues on the VIAF team. Hayley's role as VIAF Liaison exists to serve as support for VIAF and as the main contact person for the VIAF Council. She is also involved with maintenance of the VIAF database working directly with VIAF contributors and data providers regarding accuracy with authority clustering.

About VIAF

The VIAF® (Virtual International Authority File) service — an OCLC service since 2012 — has its origins as a collaborative activity of the United States Library of Congress, the German National Library (Deutsche Nationalbibliothek), the National Library of France (Bibliothèque nationale de France) and OCLC Research. This OCLC service provides access to a global set of library name authority files and has the distinction of being one of the most frequently used datasets in the library domain. It also serves as an excellent example of an important linked data hub.

The distinctive work of the VIAF is to cluster authority records from different countries for the same entity. In concert with OCLC's non-profit mission to make information more accessible and useful to people on a global scale, the VIAF aggregates and makes available valuable library authority work from national libraries and other agencies.

OCLC provides shared technology services, conducts original research, and hosts community programs to address the ever-evolving needs of library users, libraries, archives, and other interested communities. As the world's largest library cooperative, OCLC and its members continually improve metadata creation, interoperability, and indexing.

VIAF Council represents the VIAF Contributor agencies and advises OCLC on both strategic and tactical matters related to the VIAF. VIAF Council and the VIAF team at OCLC routinely monitor standards activity and other relevant work by the ALCTS MARC Advisory Committee (MAC), the DCMI (Dublin Core Metadata Initiative), Europeana, ISNI (International Standard Name Identifier), ORCID, the LITA /ALCTS Authority Control Interest Group, NISO, the PCC (Program for Cooperative Cataloging), the SAA (Society of American Archivists) Metadata and Digital Object Section, schema.org, and the W3C.

More information about VIAF is available on the OCLC website: https://www.oclc.org/en/viaf.html.

VIAF focus and future work

The VIAF database can ingest various versions of MARC and output data in a variety of linked data formats. The VIAF service utilizes RDF, an important metadata schema, since it is widely consumed on the Web and supports linked data. At the same time, one challenge for consumers of RDF is that external systems output RDF differently. The VIAF team would like to see more harmonization or a commitment to a common output form that is standard.

The VIAF team continually works to improve matching algorithms for clustering in the VIAF and to support FRBR-inspired clustering of WorldCat bibliographic data. Currently, the VIAF team is supporting internal discussion of OCLC's plans for linked data, and the team views linked data as critical to VIAF's future work and development.

Critical work in diversity, inclusion and accessibility

Aggregation is central to the work of the VIAF and as such they accept authority files in native form from a diverse domain of experts. VIAF has implemented Unicode support and can handle a wide variety of scripts presented by the files VIAF ingests. Aiming for inclusivity, the VIAF team has actively worked with users to translate the VIAF interface into several different languages.

On the "subjects" side of OCLC operations, the Dewey Decimal Classification (DDC) system's editorial team has been consulting with indigenous nations to create DDC numbers reflecting their geographic, political, etc. perspectives of these peoples. This work, led by Rebecca Green, has recently been presented at the SIG/CR workshop, October 27, 2017, "Providing for Indigenous Nations in the Dewey Decimal Classification." See: https://sigcr.files.wordpress.com/2017/10/2017sigcrslides-green.pdf.

Tags: [linked dataName Authority](#)

ALA Metadata Standards Committee Midwinter 2018 Minutes (February 11, 2018)

BY [ERIK](#) · FEBRUARY 11, 2018

The ALA MSC met on 2/11/2018 for ALA Midwinter. Thanks to everyone who came to the committee meeting. Minutes are available at http://connect.ala.org/node/273377.

Metadata priorities and focus profile: DPLA (February 28, 2018)

BY [MIKE BOLAM](#) · FEBRUARY 28, 2018

The Metadata Standards Committee spoke with Gretchen Gueguen, Data Services Coordinator, [Digital Public Library of America (DPLA)](#). Gretchen spoke with MSC in October 2017 with the goal to share current work related to metadata. A critical area of exploration included questions around how the areas of diversity, inclusivity, and accessibility were factoring into metadata decisions.

Gretchen is responsible for overseeing the process of harvesting and mapping all of the data that DPLA partners contribute. This includes communicating with partners about DPLA metadata application profiles, reviewing their data to be sure that it meets certain quality benchmarks, and then working with them to get the metadata to hit those targets. She also manages the data itself. She works with the development team to map data from the partner's original format and schema to the[DPLA Metadata Application Profile (MAP)](#). Gretchen analyzes the data for quality, and is working on initiatives to improve or enhance the data through outreach with partners as well other automated approaches.

About DPLA

Planning for DPLA began in 2010 as an initiative to create a national digital library. They launched the first website in 2012 as an aggregation of 2.4 million metadata records from 15 different institutions, available through an API and a searchable web site. DPLA has grown to more than 20 million records from 39 partners. Some of their partners are large institutions that provide a feed of records from their own holdings (NYPL, the Smithsonian, NARA, etc.). Others contribute to DPLA though [Service Hubs](#), which aggregate records from institutions all across their state or region. There are 22 state or regional-based Hubs (one of which represents a partnership between multiple states) and more forming in other states. Each record in the DPLA databases contains a link back to the original full record and digital object at the host institution. DPLA only collects metadata, not digital objects.

Metadata Background for DPLA

The DPLA MAP, derived from the [Europeana Data Model](#), relies heavily on [Dublin Core](#)for descriptive data, but includes additional properties relevant to the needs of aggregated metadata (for communicating original institution, reference links, thumbnail image location, etc.)

Harvesting metadata is crucially important for DPLA. Although most of their partners use the OAI PMH protocol to share records, others have developed APIs, perform large file transfers, or are working on ResourceSync implementations.

Metadata mapping is a resource-intensive activity. As part of the transformation, DPLA attempts to enrich and standardize the data. Currently, DPLA tries to normalize dates and match place names to geographic coordinates and clean up errant HTML or white spaces. They are working towards more robust enhancements in the future, including authority reconciliation and interoperability with other linked data services. DPLA does lightweight quality assurance (QA) and analysis, mostly during the initial harvest to ensure that the mappings and enrichments work as planned. They have been planning for a more sophisticated analytics project in the future.

DPLA focus and future work

DPLA is redeveloping their metadata ingest, mapping, and QA application and hopes that it will be a lot simpler and give them the ability to do more robust data analysis and QA, provide more useful information to partners to improve their metadata, and eventually improve the enrichments DPLA provides. They are also currently doing a website redesign and it has given them a chance to re-evaluate where better metadata could improve the search experience. They are hoping to work with their partners to improve data around collections and formats to provide a better front-end experience.

Critical work in diversity, inclusion and accessibility

DPLA has recently worked on updating their [values statements](#) and has begun to examine their hiring and outreach processes to become more inclusive. So far, they have not done much to ensure their metadata is inclusive. Because DPLA is an aggregator, they do not directly control any metadata creation. For technological, financial, and philosophical reasons, they do not make many modifications to the data values in the metadata. As they move along with their analytics project, Gretchen hopes to get a better handle on how inclusive their metadata is. She hopes to use this information to generate discussions or initiatives among the DPLA partners to make changes.

Tags: [aggregationdiversitydpladpla mapDublin Coreedm](#)

Metadata priorities and focus profile: ALCTS Standards Committee (March 3, 2018)

BY [MIKE BOLAM](#) · MARCH 3, 2018

The Metadata Standards Committee exchanged emails with Miranda Nixon, Chair of the [Association of Library Collections and Technical Services](#) (ALCTS) [Standards Committee](#). Miranda spoke with MSC in October 2017 with the goal to share current work related to metadata. A critical area of exploration included questions around how the areas of diversity, inclusivity, and accessibility were factoring into metadata decisions.

Miranda has been a member of the Standards Committee since 2015, and is serving as chair from July 1, 2017, to June 30, 2018. She is currently the Preservation Coordinator at the University Library System at the University of Pittsburgh.

About the ALCTS Standards Committee

The ALCTS Standards Committee's mission is to provide ongoing education about relevant standards to ALCTS members and other interested individuals in the information industry, and collect and share information regarding current and developing standards within the scope for ALCTS. They actively promote member involvement in the standards development process by maintaining close relationships with the ALCTS Section Executive Committees. The committee provides support for the ALCTS NISO representative.

Metadata Background for the ALCTS Standards Committee

As a committee interested in standards related to libraries, the committee attempts to track any standards that may be relevant to their community, which includes metadata standards. The committee work is focused on the dissemination of information and getting members involved in the standards review work, but does not do much work on metadata standards themselves.

ALCTS Standards Committee focus and future work

The ALCTS membership it the primary audience of the committee's work, but it's possible they may include more of the general public in the future. Beyond internal contacts who may pass on information to our committee regarding relevant standards' work, the committee follows the large organizations of standards development (ISO/NISO, ANSI, BISG, EDItEUR, IETF, IFLA, JSC International, LOC, NIST, PCC International, W3C). We then collaborate further down into the ALCTS division to collaborate with committees at the section level whose work may be relevant to standards (such as the PARS Preservation Standards & Practices Committee).

The committee publishes a list of standards related programming for the ALA Annual Conference and Midwinter Meeting on their ALA Connect page.

Critical work in diversity, inclusion and accessibility

The committee is not currently having conversations related to diversity, inclusion, and accessibility. Miranda feels that this work is potentially relevant to the committee's mission, and is interested in incorporating these topics into future discussion. The committee could potentially help provide input on standards when the issues of diversity, inclusion, and equity need to be better addressed.

Tags: alaalctscollectionsnisopreservationstandardstechnical services

Metadata priorities and focus profile: Federal Geographic Data Committee (March 13, 2018)

BY <u>ERIK</u> · PUBLISHED MARCH 13, 2018 · UPDATED MARCH 21, 2018

Introduction

The Metadata standards committee spoke with Jen Carlino and Lorna Schmid, Metadata Coordinators for the Federal Geographic Data Committee (FGDC), a 32-federal agency member body, in October 2017 with the goal of learning more about their current metadata interests and focus, including questions around how interoperability and inter-agency communication factor into metadata decisions.

In their role as Metadata Coordinator, Jen Carlino chaired, and succeeding her Lorna Schmid now chairs, the FGDC Metadata Working Group. Geospatial metadata issues and implementation, their primary responsibility, cuts across numerous working groups and organizations both federal and non-federal. With input from across this Metadata Community, the Metadata Coordinator assesses available tools and schema associated with national and international standards. She uses this information to propose endorsement by the FGDC of those standards recommended for adoption and implementation by federal agencies. A member of the FGDC's Office of the Secretariat, she also facilitates metadata documentation best practices to enable access-for-use of geospatial data and services.

About the FGDC

Formed in 1990 during a revision of OMB Circular A-16, the FGDC's role is to act as an inter-agency committee enabling and supporting improved use of spatial data. OMB Circular A-16 developed the National Spatial Data Infrastructure (NSDI), defined as "the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data." NSDI provides the framework through which FGDC fulfills their mission.

Under NSDI, FGDC metadata is applied to all spatial data collected or derived using federal funds and to many State, local, Tribal and non-governmental organizations, making these data discoverable online through the integrated Geospatial Platform and Data.gov. The Geospatial Platform provides electronic access to distributed spatial data sources through a single portal. Universal metadata standardization is therefore key to preserving spatial data and serving the diverse users of the same.

Links

https://www.fgdc.gov/policyandplanning/a-16/index_html

https://www.geoplatform.gov/

Metadata background for the FGDC

Geospatial metadata describes many of the same attributes as bibliographic metadata, such as creator and date of creation. The Content Standard for Digital Geospatial Metadata (CSDGM), a standard

unique to U.S. federal agencies, has been the default FGDC standard for years, and is still in wide use. Jen and Lorna explained the Metadata Working Group has recommended the voluntary transition for all agencies to the ISO (International Organization for Standardization) 191xx series of geospatial metadata standards.

Adopting these international standards will improve spatial data accessibility across borders and cultures, serving the committee's core mission, though as adoption remains on the basis of voluntary consensus, CSDGM will likely have a long-term presence in geospatial data description, cataloging and discovery.

Links

https://www.fgdc.gov/metadata/geospatial-metadata-standards

FGDC focus and future work

The FGDC's Metadata Working Group will continue to advocate for ISO standards adoption as the additional extensibility of the 191xx suite of standards moves geospatial data discovery into a machine-to-machine interpretable environment, increases interconnectivity of records, and improves more efficient discovery and access to live data services. Meanwhile, the FGDC continues to assist agencies in the management and best-practices for seventeen categories of geospatial data, called themes, into which data sets are organized and made available on the open web through the Geospatial Platform. Within themes the FGDC undertakes initiatives that serve the public interest. For instance, Jen and Lorna highlighted the National Geospatial Data Asset (NGDA) Address Theme and its supporting National Address Database, established in 2016 and managed by the Department of Transportation and U.S. Census Bureau as a publically available resource providing accurate address information sourced from State and local agencies. This database would not be possible without inter-agency consistency in the use of geospatial metadata.

Links

https://www.transportation.gov/nad

https://www.fgdc.gov/topics/national-address-database

Metadata priorities and focus profile: Canadian Committee on Metadata Exchange (CCM) (March 21, 2018)

BY NADINE ELLERO · MARCH 21, 2018

Introduction

The Metadata Standards Committee reached out to the Canadian Committee on Metadata Exchange (CCM) to learn about their organization and work with metadata. We communicated with them by email

which offers the advantage of having the CCM describe their organization and mission in their own words.

About the CCM

In essence, CCM is the Canadian advisory committee on MARC formats and related national and international standards for the representation in machine-readable form of bibliographic information. It acts as a Canadian Advisory Committee to Library and Archives Canada by examining the MARC 21 communication formats and related national and international standards for the representation in machine-readable form of bibliographic information, and making recommendations on these formats. CCM receives, evaluates and makes recommendations on proposed national and international schemas for the representation in machine-readable form of bibliographic information and other related standards. It also maintains liaison with its constituent organizations and relevant outside agencies. A CCM member from Library and Archives Canada (LAC) represents CCM on the MARC Advisory Committee (MAC). On occasion, as appropriate, CCM reviews and provides feedback on ISO standards, particularly those pertaining to identifiers.

In addition to the Secretariat (which is managed by Library and Archives Canada), CCM members represent stakeholder groups within the Canadian library community. This includes members from the Association pour l'avancement des sciences et des techniques de la documentation (ASTED), the Canadian Federation of Library Associations (CFLA) (formerly these members were from the now-defunct Canadian Library Association), and one member representing the Canadian archival community, from the Canadian Council of Archives' Canadian Committee for Archival Description (CCA-CCAD). CCM currently has four additional non-voting, co-opted members who represent other facets of the community and contribute additional expertise, including one member employed by a Canadian company that develops and sells library automation systems, to provide the perspective of an implementer of MARC 21 communication formats and related standards.

CCM focus and future work

CCM's focus is on encoding standards for transmission of bibliographic metadata, which directly relates to our role in MARC 21 governance. Regarding upcoming work, CCM reviews all papers and proposals being submitted to MAC in advance of its meetings and provides feedback on them. CCM also prepares and submits its own papers and proposals, on occasion. For instance, during the past few years, CCM has sponsored 3 or co-sponsored papers concerning coding of electronic resources of various content types (recorded sound, maps). CCM recently developed and sponsored a paper on recording accessibility information in MARC 21 bibliographic records (http://www.loc.gov/marc/mac/2017/2017-dp03.html). CCM will also have a role in reviewing upcoming metadata encoding standards (e.g. Bibframe) for their applicability in the Canadian context. It is CCM's role to ensure that the Canadian point of view is heard, whether formal channels for feedback (such as MAC) are established or not. Evaluating data value standards for cultural biases is outside the scope of CCM.

Metadata priorities and focus profile: NoveList (March 21, 2018)

BY ERIK · MARCH 21, 2018

The Metadata Standards Committee received a survey response from Renee Young, Technical Metadata Manager in NoveList's Book Discovery Department, in March, 2018. She detailed NoveList's metadata standards and current work, with a particular focus on how concern for greater diversity, inclusivity, and accessibility influenced metadata decisions.

In her role as Technical Metadata Manager, Renee coordinates between the department responsible for cataloging, authority control, and content creation (Book Discovery) and the Technology department, entering and tracking metadata projects in Tech's priorities.

About NoveList

NoveList is a readers' advisory database and related products, first launched in 1994. They provide metadata-driven reading recommendations and catalog enrichment as well as custom-written recommendations, articles, newsletters, and library marketing tools to, primarily, school and public libraries. They've been a division of EBSCO since 1999.

Metadata Background for NoveList

NoveList uses MARC records and adapted RDA rules, cataloged in the Polaris integrated library system. These records are enhanced with locally-created descriptive metadata as well as Library of Congress Subject Headings, diverging from standard LCSH as needed.

Vendor partnerships provide some NoveList data, such as book jacket images and annotations, but they rely primarily on internal resources for metadata creation.

For large scale metadata projects, they may consult with readers' advisory experts. Joyce Saricks, Neal Wyatt, and Barry Trott, all leaders in RA, have contributed their knowledge to the creation of local vocabularies.

NoveList focus and future work

NoveList's proprietary metadata vocabularies for both bibliographic and authority records are the focus of continuous discussion and innovation, aiming to increase discoverability and improve machine-matched recommendations. For instance, they will soon release a new vocabulary of themes and tropes found in fiction.

NoveList Select, which enhances the library catalog and other user points of access with NoveList recommendations, is their fastest-growing product. A vendor partnership with Zepheira in 2016 led to the creation of NoveList Select for Linked Data, which adds NoveList metadata and recommendations to linked data resources, like Linked Library Service. Other modes of NoveList content availability on the open web may be implemented in years to come.

Critical work in diversity, inclusion and accessibility

Members of the NoveList Book Discovery department are librarians, aware of the need for discoverability of more diverse books, especially in schools and the children's section of public libraries—their core market. Therefore, they created a vocabulary of headings for juvenile fiction which indicate whether a book features characters with disabilities (Ability diverse), characters of color (Culturally diverse), characters from religious backgrounds other than Christian (Religiously diverse), or LGBTQIA+ characters (LGBTQIA diverse). They are still discussing whether to adopt similar headings for their adult holdings.

Resources promoting diverse materials are found throughout the NoveList database and related newsletters. Book Squad sends an email on Diverse Reading to subscribers every month. With an author characteristic search, librarians can find books by, for example, African American or LGBT authors—who also have devoted carousel displays for browsing by genre.

NoveList has an established internal suggestion and review process for their LCSH heading use. This process has driven such projects as adopting people-first language across all subject headings describing people with physical, mental, or developmental disabilities. Although the suggestion process is internal, NoveList users can, and occasionally do, submit feedback requesting heading changes. These external suggestions always receive highest priority for review and action.

Metadata Priorities and Focus Profile: HathiTrust Digital Library (July 31, 2018)

BY AYLA STEIN · PUBLISHED JULY 31, 2018 · UPDATED APRIL 12, 2019

The Metadata Standards Committee spoke with Barbara Cormack, Metadata Analyst for Zephir at the California Digital Library (CDL), and Angelina Zaytsev, User Services Librarian at HathiTrust with the goal of learning more about HathiTrust's current activities related to metadata.

About HathiTrust

HathiTrust is a partnership-based collective of member organizations made up of research institutions and academic libraries from around the world. The primary project of HathiTrust is the stewardship and growth of the HathiTrust Digital Library, a digital repository that contains digitized materials from its partners' print collections. HathiTrust provides online access to resources in the public domain and makes their collections available for research through non-consumptive data mining and search within restricted texts.

Metadata Background for HathiTrust

HathiTrust deals with metadata in a variety of contexts. In the Digital Library itself, there is bibliographic, descriptive and preservation metadata for each work, as well as metadata about rights and access. Manual copyright review projects have resulted in data describing the copyright status of texts. Members are annually required to submit minimal data describing their print collections for the

purposes of determining fees. Below, we primarily focused on bibliographic metadata in the HathiTrust context.

Bibliographic metadata is submitted to HathiTrust via Zephir, the HathiTrust bibliographic metadata management system, by partner institutions for the digitized materials they are contributing to the Digital Library. Zephir validates submitted records for adherence to MARC binary and MARCXML standards, file structure, UTF-8 encoding, and HathiTrust item identifier requirements. In addition to validation, Zephir is also the storage and management system for HathiTrust metadata records and facilitates transmitting the records between HathiTrust workflows. Metadata records are made available in a variety formats and services, including: the HathiTrust Catalog, an OAI-PMH feed, the Bibliographic API feed, and text files called hathifiles.

HathiTrust focus and future work

Currently, development of a new metadata schema for recording information on the quality of HathiTrust materials is underway. Recording information about the quality would facilitate additional future uses of content, e.g., allowing print disabled users to identify works that have high-quality OCR or facilitating collection management work at member institutions. An ambitious project to collect bibliographic records for all known U.S. federal government documents is also ongoing.

Duplicate detection is an area of focus for HathiTrust, which contains 8.5 million records from 54 different libraries, and they are tackling the problem via time-honored ways, such as looking at ISBNs and OCLC numbers. They are also thinking about how to make it easier to rectify errors and enhance metadata, as well as developing ways to efficiently report errors for correction.

Although not included in the official development roadmap, beginning conversations about whether linked data has potential for HathiTrust have been broached.

Finally, the HathiTrust Research Center (HTRC) is developing cutting-edge tools for non-consumptive data-mining research, including generating collection-level metadata for user-created worksets; extracting additional feature data such as page numbers, genre, etc.; and generating other types of metadata, e.g., identification of parts-of-speech, page level metadata, among others.

Critical work in diversity, inclusion and accessibility

HathiTrust is not currently having conversations related to diversity and inclusion in their metadata work. These types of reviews and policies would be more likely held within their member institutions, who continue to retain custodial responsibility for their metadata after sending it to HathiTrust.