

2006

American Library Association (ALA) Midwinter Meeting Report

Christopher Cox

University of Wisconsin-Eau Claire, cnc2@clemsun.edu

Hui Hua Chua

Michigan State University, chua@msu.edu

M. Claire Stewart

Northwestern University, cstewart@unl.edu

S. G. Ranti Junus

Michigan State University, junus@msu.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/libraryscience>



Part of the [Library and Information Science Commons](#)

Cox, Christopher; Chua, Hui Hua; Stewart, M. Claire; and Junus, S. G. Ranti, "American Library Association (ALA) Midwinter Meeting Report" (2006). *Faculty Publications, UNL Libraries*. 388.

<https://digitalcommons.unl.edu/libraryscience/388>

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Published in *Library Hi Tech News* 23:3 (2006), pp. 5–11; doi: 10.1108/07419050610668115
Copyright © Emerald Group Publishing Limited. Used by permission.
First published March 1, 2006.

American Library Association (ALA) Midwinter Meeting Report

Christopher Cox,¹ Hui Hua Chua,² M. Claire Stewart,³
and S. G. Ranti Junus²

1. University of Wisconsin–Eau Claire Library, Eau Claire, Wisconsin, USA
2. Michigan State University, East Lansing, Michigan, USA
3. Northwestern University Library, Northwestern University, Evanston, Illinois, USA

The American Library Association Midwinter Meeting was held in San Antonio, Texas, January 20–25, 2006. The following are highlights from the meeting. We have a range of coverage and reviewers. Christopher Cox gives us an update on groups discussing digital media such as HD-DVD and Blu-ray, electronic reference books, and RFID issues. Hui Hua Chua writes about a discussion on federated searching. M. Claire Stewart reports on the inaugural meeting of the ALCTS Preservation and Reformatting Section's Digital Preservation Discussion Group and a whirlwind meeting on standards, which included the discussion of five standards and the overall National Information Standards Organization (NISO) structure. We conclude with a review of the day-long program on OCLC Digitization Standards sponsored by OCLC Digital Collection Services & Preservation Service Center by S.G. Ranti Junus. The ALA is holding the ALA Annual Conference in New Orleans, Louisiana, June 22–28, 2006.

The Library and Information Technology Association (LITA) Digital Media Discussion Group met on Saturday, January 21, to share their thoughts on digital streaming, HD-DVD vs Blu-ray, and multimedia production in libraries. Approximately 40 people attended, the majority university librarians. Meghann Matuwichuk, Assistant Librarian at the University of Delaware, led the discussion.

The digital streaming conversation began with an announcement from a member of the Virtual Library of Virginia (VIVA) that the organization had recently purchased PBS on Demand. The product, the cost of which was not divulged, provides VIVA libraries with

200–300 popular PBS titles, including the *American Experience*, Ken Burns’s documentaries, and *Scientific American Frontiers*. Content will be provided in the form of MPEGs on DVDs which VIVA can copy and/or digitize and then send to its member libraries throughout the state. The contract restricts access to students, faculty, and staff of VIVA member institutions only, something which the VIVA representative felt might be difficult for smaller member libraries to enforce. Access could be restricted via IP address or login authentication.

There are a number of problems with this product and digital media streaming in general. First of all, some filmmakers are nervous about PBS’s new product, concerned that licensing and copyright might not be enforced. Thus, VIVA owns the content for only three to five years. At that time, some of the content may be recalled and replaced. This differs from Annenberg’s video on demand product (www.learner.org/view_programs/view_programs.html), which offers lifetime ownership of downloaded titles. Annenberg is also free. VIVA would also have to purchase new titles as they became available. Another concern involved files from PBS which included more than one episode. Could these be broken up? How would they be cataloged if they couldn’t be? Good examples exist for cataloging digital media content—the University of Maryland’s Films@UM being one. Another complaint from the group was that much streamed video content cannot be viewed full screen, except on a Mac, and often at a loss of picture quality.

The next topic was HD-DVD vs. Blu-ray. The advantage of these new technologies is that they have far greater storage capacity than the current DVD: three times more with HD-DVD and five times more with Blu-ray. The extra space comes about due to the shorter wavelength blue laser (a blue ray) which is used to read them, allowing smaller pits of data and thus more data to be stored on the disc. This will also increase the streaming rate—up to 36Mbps from the current 11Mbps—resulting in an even clearer picture. The discs also have other benefits, including being double-sided and hard coated, thus less susceptible to damage.

Just like the old VHA/Beta debate, there is no clear sense which technology will win out, especially since neither is available yet. Many experts, including discussion leader Matwchuk, put their money on Sony’s Blu-ray, which will be included in its April PlayStation 3 release. Sony has connections in Hollywood and has already convinced a number of big-name studios to release their upcoming movies in Blu-ray.

These technologies could have a significant impact on libraries. First of all, more available space will mean more information stored on each disc. This information could include the usual extras DVDs currently come bundled with but could also include related television shows or other movies, even companion games. Whole TV series could be included on one disc. This would greatly increase the prices of these discs and make cataloging significantly more frustrating. To learn more about these new technologies, see “Inside Tech: Blu-ray & HD-DVD” (*Sound and Vision Magazine*, May 2005—www.soundandvisionmag.com/article.asp?section_id=2&article_id=826&page_number=1).

The final topic of discussion was multimedia production in libraries. Lately more and more professors have been requiring their students to create multimedia presentations including video. Libraries are starting to circulate digital and video cameras and are making room for editing stations. Syracuse University, for example, has a self-service lab in its library which is the only access point for digital editing hardware and software. This brings

with it a number of challenges. Staff training becomes paramount, as students expect librarians to know about video editing and to be able to assist them. Librarians do not always have the benefit of knowing the assignment students are working on, resulting in significant miscommunication. There are hardware concerns since files are large and discs are rarely big enough to hold them. Student time constraints round out the list. Many students believe video editing can be done in real-time, being unfamiliar with the time it takes to render the video. One librarian in the group shared that her library renders student projects overnight and checks on them in the morning. More and more libraries are turning to partnerships with their IT departments in an effort to alleviate these problems. This is a trend we will no doubt see continue.

The Reference and User Services Association (RUSA) Reference Services in Large Research Libraries Discussion Group met on Sunday, January 22. The topic of discussion was electronic reference books: who is using them and how do they change reference collection development? It was an outgrowth of Mirela Roncevic's November 15, 2005, *Library Journal* article "The E-Ref Invasion: Reference 2006." The topic must have hit a nerve because over sixty people showed up to listen and/or participate.

Carol Tobin, Head of Reference at Davis Library at the University of North Carolina at Chapel Hill, set the stage by outlining the issues of most concern to her. First, was cost—sometimes you want something electronically and you cannot afford it. Budgets have been flat in recent years and costs for resources keep climbing. Second, is how you get it. Is it a subscription, a one-time purchase? Is there content in addition to what was included in the print? Third, Tobin expressed concern that, if we buy these things, will people be able to find them? Reference materials have always been difficult to find in the online catalog. This becomes even dodgier when vendors offer collections of materials in a variety of formats (books, electronic journals, primary sources, free stuff) on one topic. Tobin calls these "playpens"—how does a user discover what content is in these?

Other considerations brought up included how choices are made, inclusion in federated search engines, and competing with Google. When it comes to choice, there are hundreds of resources out there. How do you learn about them? How do you evaluate them? If you purchase them, how do you assess usage? Is usage restricted to one course or instructor? When is it better to purchase electronic vs. print? Does one do so to improve accessibility to students outside of the library or in distance learning programs? The University of Southern California explained how they embed links to electronic reference content in Blackboard courses, making the library and its resources more visible to students. Then there is the search interface. Is it superior or inferior to the print? At times there may be products, like the *Dictionary of Literary Biography Online*, which leave out information included in the print—images, for example. Is easier access worth the loss of some content? Speaking of accessibility, is searchability by federated search engines a factor in choosing what to purchase? These tools may be a partial solution to the dilemma of student discovery of reference materials. Finally, how do we convince students that what we are spending money on is better than Google at least for some things? Studies like the one published in *Nature* comparing *Wikipedia* and *Encyclopedia Britannica* do not do anything to help us with this.

The University of Arizona shared a cost/benefit analysis that they conducted. Owing to space constraints and a renovation of the reference area to make room for an information commons, the library aggressively pursued electronic alternatives and kept only 10 percent of its paper reference collection. In their assessment of the project, they noted that they do not use 10 percent of the 10 percent they kept. The University of South Florida will be reducing the footprint of its reference collection from 14,000 to 7,000 square feet. How much reference is enough to do our jobs and to serve patrons? Is an information commons more important to users?

This begs the question: Who is the audience for the reference collection? Is it students? Is it faculty? Is it us? If it's students, how do we promote the collection to them? Will they use it without mediation? If it's faculty, what can we afford if it's only one faculty member who needs it? If it's us, how large of a collection do we need to answer users' questions?

Current electronic reference products were discussed. Xreferplus (www.xrefer.com/) offers online, full-text access to over 200 reference works, including the recently signed *Encyclopedia Britannica*. Reference Universe (<http://refuniv.odyssey.com/>) takes a different approach, indexing over 6,000 print reference works. In effect, it is sort of a link resolver for reference collections. Along these same lines, David Tyckoson suggested that he believes the two most innovative reference tools of the last few years are not even reference tools: Serial Solutions and SFX. How does one even define an electronic reference source? Isn't Google one?

Sarah Wenzel from Columbia University had an interesting closing comment. We have to remember in pursuing electronic over print or in reducing our reference collections that we cannot get rid of all of it. We risk losing the justification for our future. We are the place of last resort when Google and Wikipedia do not answer people's questions. Despite the lure of information commons, we should not forget our purpose: to assist users in meeting their information needs. We should have whatever tools are needed to achieve that goal.

The LITA RFID Interest Group held its meeting on Sunday, January 22. The agenda included discussion of plans for future programs and possible partnerships to offer those programs as well as a panel discussion of RFID issues. Panelists included Pat Stevens from NISO, vendor representatives from VTLS, Integrated Technology Group, 3M, Light Speed, Tech Logic, Bibliotheca, and others as well as libraries that had implemented or were in the process of implementing RFID. About 35 people attended the meeting, led by group officers Qing Haley from Chicago State University and Lynne Jacobsen from Warren-Newport Public Library.

The discussion got off to a rousing start with an explanation of the differences between RFID applications in retail and in libraries. There is confusion within the library community concerning Wal-Mart vs. library applications of RFID. One guest noted that tags put on shipping crates were literally falling off in transit, so how effective could RFID in libraries be? In point of fact, the tags used in industrial applications are different from those used in libraries. They are ultra-high-frequency tags, for one thing. They also hold different information. The thought that RFID might really catch on if book jobbers start putting tags in books might not hold water if there is no interoperability or unless libraries decide among themselves what they really want included on the tags themselves. Regardless, retail centers and libraries have very different uses for the tags, so the chance of there being

in the future one tag used for both applications is very slim. There was some debate within the group about offering a program detailing the differences, but the consensus was that such a program would not have a broad enough audience.

The focus next turned to privacy issues and what data should be stored on the chips. A task force of the Netherlands Association of Public Libraries has come up with standards regarding this for its members. Compulsory data include barcode ID, library ID, and item ID if different from the barcode. All other data is optional. The reason this is so important in the Netherlands is that its interlibrary loan system is countrywide. A NISO working group is currently crafting similar standards. In the experts' minds, ISBNs should not be included, something retail desperately wants.

One of the challenges is weighing privacy against interoperability. Many fields on the tags are locked after coding, and encryption of this nature inhibits interoperability. There is also the cost of the tag. The more info that is stored on the tag, the more memory it must have, and the higher the cost per tag may be. Do we envision each tag as carrying a tiny database of information or the bare bones necessary?

The discussion then moved to the relationship between security and tags. The panel seemed to favor application family identifier (AFI) tags, already approved by the ISO, over electronic article surveillance (EAS) tags. AFI tags are interoperable and do more than just security. Tangentially, a question was posed about retagging collections if NISO says a certain tag or standard is better. In general, many libraries work in a mixed environment. If necessary, retagging could take place as materials are returned, so those that circulate the most get tagged first. Some tags can even be reprogrammed—Bibliotheca allows this.

The meeting ended with a discussion of libraries that were currently implementing RFID or had already implemented it. Libraries currently piloting RFID included Queens, New York, and Fayetteville, North Carolina. Lynn Jacobsen from Warren Newport Public Library described her implementation as "tremendously successful." At her library, 50 percent of users use self-checkout, and materials are automatically checked in when they are returned, saving both time and labor. Unfortunately, if the representation at this meeting is any indication, not enough libraries are currently utilizing RFID, something that will need to change before costs come down even further.

The RUSA Machine-Assisted Reference Section discussion forum titled "Federated Systems in the Age of Google Scholar: Pros and Cons" drew a crowd of over 80 participants. The majority of the attendees represented libraries with a federated search product in place, with a smaller number also having either enabled OpenURL links to subscription resources through Google Scholar and/or planned on featuring Google Scholar prominently on their library websites.

The discussion was wide-ranging, with some participants expressing dissatisfaction with the complexity of customizing the functionality and "look and feel" of specific federated search products. Various methods of organizing and presenting federated search on library websites was discussed, such as offering broad subject categories versus narrower discipline-based searches. Another method of grouping resources was based on the amount of time a user had to complete their research, with only full-text resources presented if material was required immediately. Positive comments regarding federated search centered on its ability to group resources more effectively for those unfamiliar with

the vast array of resources available. This included undergraduates and remote users without access to guidance from library staff.

The relationship between Google Scholar and federated search was not directly addressed. Librarians from smaller libraries generally preferred federated search to Google Scholar because of their institution's lack of access to the full-text resources represented in Google Scholar. Many directed their comments toward Google and the direct impact its "one-box" search had on their library's decision to offer federated search. One striking comment was that while federated search products had successfully imitated Google's simple search interface, presentation and relevance of results from federated search engines was lagging behind Google's. The disservice this provided to library users accustomed to Google was a cause of concern. In total, the discussion forum raised more questions than it answered, with many agreeing that federated search continues to remain a work in progress.

The Association for Library Collections and Technical Services (ALCTS) Preservation and Reformatting Section held its inaugural Digital Preservation Discussion Group meeting at the 2006 Midwinter Meeting. Robin Dale of the Research Libraries Group and Lars Meyers from Emory University co-chaired the heavily attended discussion.

Dale started things off by offering some background for the discussion and suggesting three topics: defining digital preservation, differences between creating and preserving digital objects, and the role of metadata in digital preservation. The group quickly agreed with the statement that digital preservation is preservation of objects, not digital reformatting. They were not able to reach firm consensus on a more expansive definition of a digital preservation program, but a number of promising directions were suggested, including the oft-repeated "preservation is not an activity, it is a process."

It has been suggested that the ALA Preservation Policy (www.ala.org/ala/alctscontent/alctspubsbucket/webpublications/alctspreservation/alapreservationp/alapreservation.htm) be revised to more accurately reflect the real issues of working with digital content. Currently, the discussion of digital content in the policy makes preservation the responsibility of commercial information providers. The group was generally supportive of the idea to revise the policy, articulating a shared responsibility for preserving content, and emphasizing the imperative to also preserve locally created digital content. Drafts will be shared via the ALCTS preservation administrators email list.

Shared responsibility was a theme throughout the discussion, as was collaboration with other groups, both to devise local solutions and to solve broader issues. Digital preservation activities rely heavily on information technology infrastructure, and technology staff are often critical to a repository or digital preservation program implementation. Groups such as SPARC and the Society of American Archivists have been facing licensing and electronic retention issues for some time and are natural partners.

Discussion turned to intellectual property and whether there ought to be a difference between access copies and those retained only in dark archives. While dark archives might be safer, in a legal sense, than open, it is difficult to justify the significant expenditure of funds to reformat and store things that are not being made publicly available. Several participants commented on the work of the Section 108 Study Group and hoped that rights denied for digital reformatting might be restored through the 108 revisions.

Near the end of the session, discussion turned to preservation metadata, and the difficulty with which it is generated. Community-generated content is a particular challenge; this metadata is expensive and difficult to produce. Faculty, and perhaps even publishers and commercial providers, are not going to create it. There will continue to be a gap between the data that we want and the data we get. On the other hand, we have been in this situation before with descriptive metadata. AACR2 has been set forth as the model and the ideal but is rarely implemented fully. The same may be true for PREMIS and other preservation metadata. At this early stage, we describe everything we may want, and over time, levels of metadata will emerge to fit our economic capabilities.

At the conclusion of the session, a number of topics for future discussion meetings were suggested. Many attendees are interested in the actual tools of digital preservation, and how repository systems like dSPACE, Fedora, and DLXS were really being used. It was also suggested that some examples be developed to help guide the discussion about digital preservation: what are examples or potential or actual losses? What are the doomsday scenarios for digital content? What constitutes a digital disaster?

The LITA Standards Interest Group meeting offered an update of the work ongoing in a number of NISO standards committees and working groups. Although just two hours were allocated for the whirlwind meeting, five standards and the overall NISO structure were reviewed. PowerPoint presentations from the speakers who used them have been collected on the LITA website: www.ala.org/ala/lita/litamembership/litaigs/igstandards/standards.htm.

Yan Han from the University of Arizona libraries, chair of the Interest Group, introduced the first speaker, Pat Stevens, Interim Executive Director of NISO. Pat offered an overview of NISO's structure and new mission, which is to "foster the development and maintenance of standards that facilitate the creation, persistent management, and effective interchange of information so that it can be trusted for use in research and learning." NISO is at something of a crossroads. The Board has been engaged in a strategic planning process and commissioned a blue-ribbon panel to offer guidance; the panel's report may be accessed here: www.niso.org/members/secure/BRPrpt05.pd. One of the panel's recommendations is that NISO clearly define its constituency and structure activities within a framework so that it functions in a proactive, rather than a reactive, mode when engaging in standards building activities.

Tim Jewell, University of Washington, introduced the work of the Standardized Usage Statistics Harvesting Initiative, or SUSHI. SUSHI will create standards for usage reports from Electronic Resource Management Systems (ERMS). Although the earlier COUNTER project sets standards for the types of usage data, generating useful reports often required a significant amount of data massaging and customization. The standards work surrounding ERMS, which has been supported both by NISO and the Digital Library Federation, seems to be on a relatively fast track, with the draft of the COUNTER standard released in January 2005, followed by the first SUSHI test reports in November.

Nathan Robertson, from the University of Maryland School of Law, followed to talk about the license expression working group, which is also focused on ERMS. Like usage data, interpreting licenses for these systems is currently a very manual, localized process. The working group seeks to establish some predictable language for authorized users,

sites, and privileges. Robertson hastened to emphasize that license expression is very different from rights expression languages, which enable actual enforcement through rights management technologies. Licenses, in contrast, are silent about a large number of potential uses, which cannot be interpreted either to allow or disallow them, and are therefore not sufficient for digital rights management. The working group expects to be testing with vendors in the fall.

Next on the agenda was the Radio Frequency Identifier, or RFID, standard, introduced by Dr. Vinod Chachra, CEO of VTLS and chairman of NISO's RFID working group. He outlined the four goals of the group: interoperability, isolation from other applications, privacy, and cost. Dr. Chachra feels that it is critical the group separate legitimate concerns, particularly relating to privacy, from some of the exaggeration, and that it look closely at RFID work already accomplished by the Danish, rather than reinvent the wheel.

Candy Zemon from Polaris Library Systems spoke about the Web Services and Practices Working Group, which has not yet decided whether a standard is warranted to address the proliferation of web-based transactions by commercial library products. The group is currently limiting its work to best practices definition.

The last set of speakers focused on the transition to the 13-digit ISBN. Ted Fons (Innovative Interfaces) spoke from the perspective of a vendor who must continue to support staff and public functions, including indexing, search, and matching, for the transitional period when both 11- and 13-digit ISBNs are in use. The Library of Congress, represented by David Williamson, has developed an implementation plan, new allowances for cataloging in publication (CIP) data, and is currently accepting some 2,000 CIP records with ISBN-13s each month. ISBN-13 arrived at a time of transition for OCLC, while they were in the midst of a systems migration that affects all of their projects. Glenn Patton explained that a record replacement plan will replace ISBN-13s mistakenly marked as EANs during OCLC's transitional period.

The OCLC Digitization Standards program was sponsored by OCLC Digital Collection Services and Preservation Service Center, which works in the area of digital collections preservation and provides services supporting the digital collections life cycle: planning, processing and conversion, e-content management, user discovery, and access. The following is the digital lifecycle in more detail:

- *Planning*: Assessing user needs and collection conditions as well as preparing digitization projects.
- *Processing and conversion*: Converting collections through the Preservation Service Center by digitizing, OCR, creating XML records, cataloging, and metadata for reformatting purposes.
- *E-content management*: Managing e-content over time; ongoing management perspectives in working with the collection; organizing collection using CONTENTdm; and preserving digital e-content using their digital preservation service.

- *User discovery and access*: Exposing the digital collection through WorldCat. The idea is to make sure that the metadata is organized in WorldCat so it can be exposed in places where users can access it for research purposes, through services such as Google, or learning management systems on campus.

Preservation metadata deals with the types of information we need to gather about e-content that help us support the curation process of digital content over time. Digital Preservation is a process or a life cycle that content goes through; we need to gather metadata about e-content and have it structured in a way so we can work with the digital objects over time.

Some examples of administrative issues surrounding preservation metadata:

- *Provenance*: Who has had the ownership/custody of the digital object before it comes to our collection?
- *Authenticity*: Is the digital object what it purports to be?
- *Preservation activity*: What sort of manipulation has been done to preserve the digital object?
- *Technical environment*: Have certain applications been used to create the digital object? What is needed to render and use the digital object?
- *Rights management*: What intellectual property rights must be observed when we work with the digital object?

Preservation metadata is important for digital preservation because:

- Digital objects depend on technology. The technological environment between content and users, such as the software and hardware needed to view or work on the digital objects, needs to be documented.
- Digital objects are mutable. Changing or altering digital objects needs to be documented and validated, which makes the provenance and authenticity of metadata especially important.
- Digital objects are also bound by intellectual property issues. Rights management metadata, especially if the copyright is still in effect, is especially important.

The idea behind preservation metadata is to make the digital object self-documenting as we continue working on it over time.

This is a good time for a conversation about the issues of preservation metadata. The basics of the digitization technology and processing are generally understood already. As we integrate our individual projects into the whole program of the library's regular work of our collection, we need to consider not only what we are doing today but also the long-term issues associated with our e-content work.

The following are brief notes on various strategies and standards discussed in the meeting:

PREMIS (PREservation Metadata Implementation Strategies)

Realizing the importance of preservation metadata in digital preservation, OCLC and RLG jointly sponsored an international working group on preservation metadata in March 2000. In June 2002 the group came out with a preservation metadata framework. This framework is a high-level description of the types of elements we need, drawing from the OAIS reference model for digital repositories. The group proposed a set of preservation metadata elements and in June 2003 came out with the PREMIS working group to develop the proposed elements down into practice.

In May 2005 the working group produced a Data Dictionary for Preservation Metadata final report, which is a comprehensive, practical resource for implementing preservation metadata in digital archiving systems. This 237-page report included PREMIS Data Dictionary 1.0, accompanying report such as context, data model, and assumptions. The report also contains special topics, glossary, and usage samples based on different types of repositories, to help institutions put preservation metadata theory into practice.

More information about PREMIS can be found at www.loc.gov/standards/premis.

METS (Metadata Encoding and Transmission Standard)

METS fills a different role in digital preservation than PREMIS. Currently, there is no MARC interchange format in broad use for digital libraries. This lack of standardization causes problems for cross-searching or moving data between repositories. Tying metadata to a proprietary package aggravates this issue and hinders us from getting the information for reuse in the future. This leads to the danger of data becoming obsolete, or the high cost of moving/converting the data to a different system.

To solve the problem we need a semantic consistency, or usage standard for metadata, which is analogous to AACR2. We also need a common container analogous to MARC syntax. METS is designed to act as a container of digital objects and metadata. METS was developed through collaboration in DLF, started at the University of California at Berkeley during their work on the "Making of the America II" project. METS is currently maintained by Library of Congress. It is an XML-based framework for gathering together all the pieces of a digital object, including all metadata associates for all of the pieces, and keeping track of them in one place. METS does not prescribe any particular metadata standard. It only endorses a number of schemes to enable flexibility for each institution to implement a particular scheme for a particular digital project.

Typically, a METS object file or package corresponds to one digital object. The digital object contents are either referenced by pointers in the XML file or can be embedded and encapsulated in the XML itself. Depending on the application, we can apply it either way. The metadata (descriptive, administrative, and structural) is also either referenced by pointers outside of the METS package or can be embedded in line.

METS package basically contains five different sections beyond the header:

- *File inventory*: An inventory of the files that make up the digital object.
- *Descriptive metadata*: An area to plug in metadata schema we want to use and fill out the metadata form for the descriptive metadata in that section.
- *Administrative metadata*: Similarly structured as descriptive metadata.
- *Behavior metadata*: Also similarly structured, and can be used in different applications for the expected or anticipated behavior the digital object should provide.
- *Structural metadata*: An important section and the biggest part of METS. It contains information on how the pieces of a digital object relate to each other.

To enable the METS package to exchange data with other repositories in other institutions, a METS profile is built to describe the standard applications of METS we use. For example, in the creation of a book profile, the profile contains information on a scanned image of a book with the number of pages, the use of MARC AACR2 for the descriptive metadata, and uses of PREMIS for the administrative metadata.

The recipient then can use that profile to drive their machine actions on the METS file we send to them. The profile is registered with the Library of Congress as the maintenance agency and is used to narrow the scope of METS so it can be applied to a certain situation consistently.

More information on METS can be found on the Library of Congress website at www.loc.gov/standards/mets.

Digital data curation

“Digital curation, broadly interpreted, is about maintaining and adding value to, a trusted body of digital information for current and future use.” Digital Curation Center (DCC) UK, approach to digital curation) (definition taken from Lorean Dempsey’s blog) <http://orweblog.oclc.org/cgi-bin/mtsearch.cgi?IncludeBlogs=1 &search=data+curation>

The standards that we have been discussing work exactly as defined above: they standardize a practice of data curation and to allow the following of good practices in maintaining all the correct information necessary to preserve our collections as we move forward. A quote from Digital Curation Center paper is particularly interesting:

Data’s future quality richness, trustworthiness is a function of investment in it (Lord, et al. “From Data Deluge to Data Curation”; JISC Joint Committee for the Support of Research, DCC approach to digital curation) <http://dev.dcc.ac.uk>

This paper concerns the science of data: the amount of data, data handling, data management, preservation, keeping it valid, keeping track of provenance, understanding data selection, and deciding what to save. This also includes investment, not only in data management but also in human time and resources in working with the digital objects.

Metadata and preservation metadata is one of the key components of maintaining and preserving our collection over time. However, there are a lot of other data that also have

to be maintained, updated, and put in our curation practices to support our preservation metadata and its integration into the technical infrastructure. We pick and choose the standards we use, and comply with and create best practices in implementing those standards. These best practices need to fit within the policy framework of our organization and any regulations in effect. Documenting metadata of all types, such as what kind of metadata schema we use for each collection material and how they have been applied, is also important. Technical infrastructure information such as software, hardware, software updates, and backup processes also need to be documented.

Other documentation is also important, such as legal documentation pertaining to copyright and ownership; intellectual property assets; risk management profile and policies; evaluation and assessment metrics and benchmarks; budget and financial data particularly cost accounting; and user data based on how they access, interact with and use our collections.

Standards exist and were created as reference models. An implementation work flow consists of adoption of standards; adoption of best practice models; implementation within the institutional policy framework and consistency through a local style guide and protocols; scheduled activities (data backup, QA/QC); documentation of decisions; and digital collection change management that we need to track.

In IT areas, some of the data sources that we need to track are: computers and the physical hardware used to connect them with users; transmission media and other devices that control transmission path; software used to send; receive and manage information transmitted; and everything that supports the flow of the information.

Long-term program policies and procedures are also important to keep our collections alive because technology and tools keep changing. Policies and procedures for digital programs need to be reviewed and updated regularly. Constant testing, review, and change management implementation also need to be recorded for our digital program. In short, data gathering and documentation are the best defenses for our digital collection program.

Data Curation Awareness is recognition of the need to be aware of data curation as a practice. Such practice includes maintaining documentation for all of our digital collection development. Education and training in data curation are also important. We also want to start early with documenting and tracking the data, revising it frequently, and making sure we are updating the information consistently. We also need to store the aggregation of policies, protocols, standards, and documentation in one place (intranet, etc.) for others to access and use. Finally, communication of responsibility—assigned documentation roles, maintaining, preserving, updating, and making sure that the documentation is accessible—is also a good practice.

Update on NDIIPP (National Digital Information Infrastructure and Preservation Program)

NDIIPP is a Library of Congress program, authorized by legislation in December 2000 with the aim of developing a national collaborative for significant, at-risk digital content, and also to work with other institutions to test, refine, and implement digital preservation.

NDIIP began the program by creating a general plan called “Preserving Our Digital Heritage,” which is available on their website. Owing to the sheer size of the project, a

program was announced looking for formal partnerships. In September 2004, eight partnerships were established, comprising 36 institutions, and were funded for three years to work on developing various types of infrastructure in preserving digital materials. The general goals of the partnership are to identify what type of information should be preserved; looking closely at the intellectual property issues; collaborate in developing a shared technical architecture; study the economic sustainability of maintaining the digital collection, especially after the grant money is no longer available; identify and share best practices; and learn how to build, and incrementally improve, a preservation network.

Another research area that they recently completed is the LC/NSF Collaboration on digital repository. They worked on cutting-edge research in the areas of digital preservation models, including the tools, technology, and processes as well as organization, economic and policy issues. In May 2005, NSF awarded \$3 million in research funding to ten institutions, which are required to produce results within one year (Summer 2006). The research efforts and results will be integrated with the larger NDIIPP efforts.

The NDIIPP project was originally envisioned to run until 2005. But Congress recently authorized an extension of the project to 2010. This allows them to document the project in more detail and to assess what they still need to do. These areas include more collaboration opportunities; developing e-deposit capacity to accept materials electronically; engaging private sector IT and content companies; and communicating emerging standards and best practices.

Something that they have learned so far is that there is no silver bullet solution to digital preservation. It may end up that digital preservation is a set of various methods for preserving materials that we hope would be interoperable. Another thing they are looking at is how to integrate various approaches to digital preservation into some kind of technical network. Budget is also a big concern, especially for preserving state records.

Another group that was recently established is the Section 108 Study Group. This is an independent group of experts sponsored by NDIIPP and the US Copyright Office. Section 108 is the section of the US copyright law that provides exceptions to libraries and archives, allowing them to provide services on copyrighted materials that are not available from other organizations. This group consists of individuals representing various aspects of copyright point: libraries, archives, and various copyright industries such as the film industry, publishers, and other content creators and companies. This group serves as experts providing advice to Library of Congress on making recommendations to Congress this summer on changes in the copyright law to cover the digital technology while still within fair use.

Detailed information on the NDIIPP program is at www.digitalpreservation.gov. Section 108 study group information can be found on their website at www.loc.gov/section108.

Author Affiliations

Christopher Cox (coxcn@uwec.edu) is the assistant director at the University of Wisconsin–Eau Claire Library in Eau Claire, Wisconsin.

Hui Hua Chua (chua@msu.edu) is a US documents librarian and journalism bibliographer at Michigan State University, East Lansing, Michigan.

M. Claire Stewart (claire-stewart@northwestern.edu) is head of Digital Media Services and coordinator of Digital Library Projects at the Northwestern University Library in Evanston, Illinois.

S. G. Ranti Junus (junus@mail.lib.msu.edu) is a systems librarian at Michigan State University Libraries, East Lansing, Michigan.