

ARTICLE
EXCERPTED
FROM:

ISQ

INFORMATION STANDARDS QUARTERLY

WINTER 2010 | VOL 22 | ISSUE 1 | ISSN 1041-0031

SPECIAL ANNIVERSARY EDITION: PART FOUR

PLANNED OBSOLESCENCE:
A NEW MODEL FOR
ACADEMIC PUBLISHING

DCMI: BEYOND THE
ELEMENT SET

E-BOOKS: THE ETERNAL
NEXT BIG THING

Z39.7 LIBRARY METRICS &
STATISTICS DATA DICTIONARY

OPENURL KNOWLEDGE BASES
RECOMMENDED PRACTICE

QA

[QUESTION & ANSWER]



John A. Kunze
Associate Director
UC Curation Center



Patricia Martin
Director, Discovery & Delivery
California Digital Library

MEMBER SPOTLIGHT:

California Digital Library: Standardizing Digital Practices Across the University of California System

The California Digital Library (CDL) was founded by the University of California in 1997 to take advantage of emerging technologies that have transformed the way digital information is published and accessed. Since its inception, in collaboration with the UC libraries and other partners, the CDL has assembled one of the world's largest digital research libraries and changed the way that faculty, students, and researchers discover and use information at the University of California and beyond.

The CDL is organized into five distinctive programs emphasizing the development and management of digital collections, tools and systems for online discovery and delivery, innovation in scholarly publishing, and digital curation and long-term preservation, which together provide a wide array of services on behalf of the University of California, its libraries, its pursuit of scholarship, and its public service mission.

Notable CDL initiatives include the Melvyl shared online catalog, the Online Archive of California (OAC), Calisphere, the CDL Web Archiving Service (WAS), eScholarship publishing services, and the UC Curation Center (UC3). CDL also operates an extensive licensing program on behalf of the UC campuses and organizes University of California participation in large-scale digitization initiatives with Google and the Internet Archive, including founding participation in the HathiTrust shared digital repository. With more than 220,000 students, 170,000 faculty and staff, and more than 35 million volumes in its combined library collections, the University of California Libraries together comprise the largest single university library system in the world.

NISO asked CDL to respond to the following questions regarding our use of standards and involvement in standards development.

Q What standards are most important to your organization and why?

John: CDL implements its services using a variety of specifications, from formal international standards in well-understood domains, to proposed standards and locally defined methods in domains where no suitable standards yet exist. Within this mix are mature, widely adopted standards that we and every online organization rely on implicitly and absolutely, such as TCP/IP and SMTP (e-mail), as well as core web-facing service standards such as DNS, URI, HTTP, and HTML. The nature of our enterprise requires that we actually consult the texts of these last four standards on a regular basis. Also, a large number of CDL services model information in XML and generate web pages (in HTML) using locally developed

CONTINUED »

XSLT (XML style sheet transformations). For metadata, Dublin Core has been broadly influential at CDL.

METS (Metadata Coding and Transmission Standard) has traditionally been a required wrapper for objects deposited to CDL special collections, although we will be modifying this requirement for ingest into our preservation repository. METS is also a key component of the architecture used to store digital objects in HathiTrust. Outside of CDL's bibliographic systems, ARK (Archival Resource Key) identifiers are used for most digital assets. (ARK is an actionable identifier optimized for persistence that the CDL was instrumental in developing.) Support for the Semantic Web's "Linked Data" is also being built into our new curation services; it is early days still for this sort of application, but the idea is to permit automated processes, not just people, to one day make inferences about relationship types that we are recording with our digital assets today.

Patti: For the Discovery and Delivery team at CDL, the most important standards are those related to metadata, both the descriptive methods—MARC, Dublin Core, and ERMI (Electronic Resources Management Initiative)—and the access methods—including OpenURL and Z39.50. We focus on providing a comprehensive set of discovery services, along with delivery mechanisms, so these standards are the workhorses of our team's services.

Q How has your organization incorporated standards into its products / services?

Patti: A good example is the ERMI standard. While looking for an Electronic Resource Management system, we required potential vendors to support this emerging standard. While implementing our ERM service, we relied on the standard to guide our implementation.

We provide a lot of service "glue" aiming to incorporate different services into our suite so that we make life easier for our end users. For example, our union catalog relies on MARC, OpenURL, ERMI, and Z39.50 to allow end users and library staff to find what we hold, to link to electronic copies, to look up our licenses for troubleshooting purposes, and to provide an integrated borrowing or document delivery solution. Because there's so much interpretation while implementing a standard, we tend to favor vendors or solutions that provide the most open, least proprietary solutions.

John: Agreed. And sometimes it is a challenge to deal with the eccentric application of standards by others. For example, while the XML we create may be rigorously correct, most of the HTML that our web harvesters gather is technically invalid. At the same time, that HTML content is too expensive for the sources to correct, too strategically

important for us to reject, and too comfortably rendered without complaint by current web browsers that compensate and fix errors (in ad hoc ways).

Q What benefits has your organization gained from utilizing standards and incorporating them into its products / services?

Patti: We use standards to find common ground with folks external to our organization and to our community. We recognize that we will need to broaden the communities that we usually talk with (i.e. the library community, vendors, and developers) to include others, such as publishers and content providers.

John: The BagIt and ARC/WARC (Web ARChive) container formats have given us preservation confidence by helping us exchange large amounts of content with the Library of Congress and with our university, non-profit, and national library partners (e.g., Stanford and the Internet Archive). Both of these standards provide methods for packaging multiple, related information objects with relevant metadata.

Q What standards development has your organization been actively involved in?

Patti: Three projects that come to mind are ERMI, KBART, and ILS-DI. ERMI was the Digital Library Federation (DLF) project that specified the functional requirements and data elements for electronic resource management systems. Another CDL colleague, Ivy Anderson, has been involved with that project since its inception and currently co-chairs a NISO working group to chart next steps for ERMI. KBART (Knowledge Base and Related Tools), the joint NISO/UKSG initiative, just issued a recommended practice on how to improve the quality of the metadata in OpenURL knowledge bases. ILS-DI (Discovery Interfaces) was another DLF project to specify an API for interoperability between integrated library systems and external discovery applications.

John: Other NISO efforts that CDL staff have worked on include Dublin Core metadata [ANSI/NISO Z39.85], SUSHI protocol for usage statistics [ANSI/NISO Z39.88], and the current effort on Institutional Identifiers. We were also involved with the international ISO standard for the PDF/A archival document format [ISO 19005-1].

CDL staff have been centrally involved in the standardization and/or specification of URL and ARK identifiers, the Z39.50 protocol, the WARC container for web archiving (now ISO 28500), BagIt for generic content exchange, and Dublin Core Kernel metadata. We've also participated actively in the development of the SERU (Shared E-Resource Understanding) Best Practice and on the METS advisory board.

Q What benefits does your organization gain from active involvement in standards development?

Patti: We help shape the conversation and bring in practical use cases. We tend to push the envelope on issues related to scale and complexity and we seek to avoid the balkanization of standards efforts, and try to cross fertilize whenever possible. We don't know what standards will be needed in the future, but we hope to be part of the community uncovering them, and helping to shape the conversations.

Q What problem areas have you encountered that would benefit from further standards or best practices development?

Patti: Best practices would include a profile, or an exemplar, for interpretation and implementation of a particular standard. A good example where profiles are useful is Z39.50, which is sort of like a family of standards. It's not hard for two technically compliant vendor implementations, each with a different application in mind, to be non-interoperable at the level of either search attributes or returned record syntaxes. Profiles help create common ground and also filter the many elements that a standard generically has to provide into those subsets that fit particular applications. We think of standards as similar to a human language: standards set the rules of a language, profiles set the rules of language dialects.

We also tend to favor more open implementations and interpretations of standards over closed and more proprietary ones.

John: In standards it seems there is a classic tension between a need for stability and a need for flexibility. Particularly in digital information services, organizations want a stable specification to maximize interoperability and minimize development costs. At the same time, it is only after several years of deployment that we learn what we should have standardized on! An ideal standards process would be solid enough to help us start building services in areas where our understanding is pretty good and be nimble enough to help us keep pace with the rapid evolution of our understanding and the breathtaking pace of technological change.

Q What else would you like NISO ISQ readers to know about your organization?

John: In its role as service provider to the ten campuses of the University of California and its many libraries, the CDL is committed to a high degree of interoperation within the university and beyond. Best practices and standards are critical in fulfilling our mission. | QA | doi: 10.3789/isqv22n1.201007

JOHN A. KUNZE <jak@ucop.edu> is Associate Director, UC Curation Center and **PATRICIA MARTIN** <patricia.martin@ucop.edu> is Director, Discovery & Delivery, at the California Digital Library. John is also the alternate voting representative to NISO for CDL.



In standards it seems there is a classic tension between a need for stability and a need for flexibility. Particularly in digital information services.

ARK (Archival Resource Key)

www.cdlib.org/uc3/ark

BagIt

www.cdlib.org/uc3/bagit

California Digital Library

www.cdlib.org

Dublin Core

dublincore.org

Dublin Core Kernel

dublincore.org/groups/kernel/

ERMI (Electronic Resources Management Initiative)

www.diglib.org/pubs/dlf102/

ERM Data Standards & Best Practices Review

www.niso.org/workrooms/ermreview

ILS-DI (Integrated Library Systems-Discovery Interface)

www.diglib.org/architectures/ilsdi/

KBART (Knowledge Base and Related Tools)

www.niso.org/workrooms/kbart

METS (Metadata Encoding and Transmission Standard)

www.loc.gov/standards/mets/

SUSHI (Standardized Usage Statistics Harvesting Initiative)

www.niso.org/workrooms/sushi

WARC (WebARChive) Format

www.digitalpreservation.gov/formats/fdd/fdd000236.shtml



RELEVANT
LINKS