



June 23, 2008

NISO Thought Leader Meeting on Digital Libraries & Digital Collections

Background

The National Information Standards Organization (NISO) received a Mellon Grant that included support for four Thought Leader meetings designed to engage the broader community in recommending topics for standards development. The Digital Libraries and Digital Collections Meeting was held in Baltimore on Tuesday, June 17th. Participants identified problems that users encounter in accessing content and explored issues that affect the exchange of data for an improved user experience.

Process

The one day meeting brought together 12 participants representing diverse segments within the community: librarians (Cornell University, Columbia University), commercial publishers (Nature, Sage), educators (University of Illinois Urbana Champaign), content providers (ProQuest), policy agencies (Institute of Museum and Library Services, National Science Foundation), standards (BISG), book vendors (Ingram), preservation agencies (Portico), and software developers (Microsoft). Prior to the meeting participants were involved in one of two conference calls which also included those who were unable to attend the onsite: librarians (University of Washington, New York University, University of Pennsylvania), integrated library system vendors (ExLibris), and organizations (OCLC, Digital Library Federation).

Each participant submitted ideas that were collated and distributed to the group for review and discussion. Conversations focused on obstacles for users in working with digital content and opportunities to improve delivery and use of content in an environment that relies increasingly on machine to machine interfaces. The ideas that emerged were ranked in following order.

Participants

<i>Participant</i>	<i>Organization</i>	<i>Title</i>
Adam Chandler	Cornell	Coordinator Serv Design, Digital Tech
Alex Wade	Microsoft	Sr. Program Manager, Scholarly Communication
Amy Miller	Ingram Digital	VP Ingram Digital
Evan Owens	Portico	Chief Technology Officer
Howard Ratner	Nature	Chief Technology Officer
John Shaw	Sage	Asst. Managing Editor
John Unsworth	UIUC	Dean of Grad School Library & Info Sci
Rachel Frick	IMLS	Senior Program Officer
Mary Chute	IMLS	Deputy Dir for Libraries
Kishor Patel	ProQuest CSA	President
Lee Zia	NSF	Lead Program Director, NSDL
Michael Healy	BISG	Executive Director

<i>Participant</i>	<i>Organization</i>	<i>Title</i>
Patricia Renfro	Columbia Univ	Digital Programs & Services
Jim Michalko	OCLC	President RLG
Michael Stoller	New York Univ	Director of Collections & Research Services
John Ockerbloom	Univ of Penn	IT Sr. Project Manager
Oren Beit-Arie	ExLibris	Chief Strategy Officer
Tim Jewell	Univ of Washington	Dir Info Res, Collections, Schol Com
Peter Brantley	DLF	Executive Director

Recommendations

#1 Evaluate publisher metadata for accuracy

Issue: Poorly formed OpenURLs and DOIs prevent users from linking to content.

Publishers generate metadata that may work for their applications however they are often unaware of problems that consumers and content owners experience when the same metadata is used downstream. Clean metadata is essential for linking. Metadata appears to be worse on born print than born digital products.

Suggestions:

- Create a set of rules for the creation of good quality metadata.
- Identify the business case for the publisher by describing the impact of poor quality metadata elsewhere in the system and the value to them of creating good quality metadata.
- Develop a test suite that allows publishers to test their metadata against a set of guidelines and provides feedback in the form of a red/green/yellow rating for Open URL and CrossRef metadata.
- Consider BISG's reverse certification system as a model that provides for both the developers and users of metadata to critique both the metadata and how it is used.

#2 Establish guidelines for aggregating content

Issue: Humanities researchers face hurdles in text mining.

There is growing demand by students and faculty in the humanities to aggregate content as part of their research. For example, students may be required to write scripts for text mining. Currently they cannot access content as their activity is likely to trigger the system by shutting down a port. They may have to request publisher permission to conduct a legitimate search and then may be uncertain about rights in sharing the results of their analysis. (This recommendation is related to [#3 Authentication Methods](#).)

Suggestions:

- Create "rules for engagement" that specify what is allowed and how it is supported.
- Establish protocols for aggregating content that would be standard practice across databases.
- Devise a method that would enable a machine to distinguish at a basic level whether the content has rights restrictions.
- Identify the rights of the researcher as aggregator for redistribution and repurposing (such as Creative Commons).
- Provide a machine based method to identify:
 - The Object - is it an abstract, an article, references?
 - Available formats - image, text or OCR (#6 - Identifying format Properties of PDFs could be included in this recommendation.)
 - Rights - what are the access rights of the user? (May be handled by RightsLink, RightsSphere)

#3 Review and recommend authentication methods

Issue: Users encounter access walls when discovering content through Google

Users frequently begin research with search engines such as Google and then encounter a barrier when they try to access licensed content from home or when traveling. How can users be provided with a better experience? How can the users' affiliation be connected to their right to access specific content? (#3 Authentication is also related to [#2 Guidelines for Aggregating Content](#).)

Suggestions:

- Assess the current state of IP based authentication systems, the points of failure and make recommendations for satisfactory performance.
- Establish recommended practice for publishers to provide users with graceful error routing.
- Determine how to identify content that is freely available to users.

#4 Monitor and assess identifier initiatives

Issue: What are the interoperability issues with new identifiers?

Work is being done throughout the community on identifiers for people (as authors or readers) and for institutions. For example: WorldCat Identity Project, International Standard Name Identifier. Each community is defining a standard based on their own perspective and applications rather than having all industry participants who would use it engaged in the development.

Suggestions:

- Create a working group to monitor the initiatives and organizations developing identifiers for individuals and for institutions.
- Assess the need for interoperability among these identifiers as they develop so that the broader needs of the community are well served.

#5 Incorporate packaging in eJournal standards

Issue: eJournal standards lack specs included in newer eBook standards

The ePubs standard for books describes protocols for packaging content that would be useful to incorporate into the eJournal standards (NLM DTD). This capability would allow for content to be moved onto new systems in a standard way across publishers and platforms.

Suggestions:

- Review ePub standard to determine a suitable provision for journal packaging.
- Recommend the inclusion of packaging provision into the current eJournal standard.

#6 Identify format properties of PDF

Issue: Users cannot easily determine what can be done with a PDF

Knowing how a document can be processed is essential in a networked environment. This issue is considered sufficiently important to stand alone. (#6 could also be addressed as part of [#2 Guidelines for Aggregating Content](#).)

Suggestion:

- Create specifications that will provide both a visual code for users and metadata for documents to determine whether they are: 1 - text, 2 - OCR, 3 - image.

#7 Establish a protocol to communicate system downtime

Issue: Users raise questions about the status of systems

Library staff are involved in addressing user questions about system downtime. Having a standard set of communications about system downtime that would display in the library's OPAC would minimize the impact on library staff and provide a better user experience.

Suggestion:

- Establish a recommended practice for reporting and posting standard notifications by content providers.

#8 Address issues of title changes in journals

Issue: Journal title changes create user confusion

As more journal archives are digitized, the titles changes that occurred over time present many problems for users, libraries and publishers. Articles published under the prior journal title may only be searched under the current journal title. Users are confused when presented with a journal title that is different than what they were expecting. Providing access to prior subscriptions can be problematic and affect perpetual access rights.

Suggestions:

- Appoint a Working Group to review the pending NFAIS report and recommend next steps to address the many issues caused by journal title changes over time.