Update on NISO's Open Discovery Initiative

Nettie Lagace, Associate Director for Programs, NISO
David Lindahl, Director of Strategic Initiatives, UMKC Libraries
Roger Schonfeld, Program Director, Ithaka S+R

CNI Spring 2013 Member Meeting
San Antonio, TX – April 4, 2013

NISO

- Non-profit industry trade association accredited by ANSI with 150+ members
- Mission of developing and maintaining standards related to information, documentation, discovery and distribution of published materials and media
- Represent US interests to ISO TC46 (Information and Documentation) and also serve as Secretariat for ISO TC46/SC 9 (Identification and Description)
- Responsible for standards like ISSN, DOI, Dublin Core metadata, DAISY digital talking books, OpenURL, SIP, NCIP, MARC records and ISBN (indirectly)
- Volunteer driven organization: 400+ spread out across the world

NISO
How the information world CONNECTS

NISO Standards and Recommended Practices

- The point of a standard is to provide a reliable basis for people to share the same expectations about a product or service.
  - NISO standards creation governed by ANSI Essential Requirements
  - Standards generally contain requirements that the user “shall” complete.
- NISO Recommended Practices are "best practices" or "guidelines" for methods, materials, or practices in order to give guidance to the user.
  - RPs often more appropriate in new/emerging areas
  - Recommended practices use “should” rather than “shall.”

NISO
How the information world CONNECTS

The context for ODI

- Emergence of Library Discovery Services solutions
  - Based on index of a wide range of content
  - Commercial and open access
  - Primary journal literature, ebooks, and more
- Adopted by thousands of libraries around the world, and impact millions of users

The problem

- Governed by proprietary agreements between content providers and discovery providers
  - Issues: content scope, coverage, level of metadata/data
  - Varying requirements regarding stats/reports, ranking, linking,..
- Libraries are caught in the middle
  - Hard to evaluate discovery services and content products
  - Under-serving their users
- Resembles pre-OpenURL days
- The result: incomplete and inconsistent “eco system”

Initial discussions

- Round table discussion (@ALA annual, New Orleans 2011) – initiated by Oren Beit-Arie, Jenny Walker and Marshall Breeding
  - representation: libraries, consortia, content vendors, content organizations, discovery services providers
  - discussed different stakeholders views
  - (surprisingly) easy consensus

General Goals

- Define ways for libraries to assess the level of content providers' participation in discovery services
- Help streamline the process by which content providers work with discovery service vendors
- Define models for "fair" linking from discovery services to publishers' content
- Determine what usage statistics should be collected for libraries and for content providers

Approach

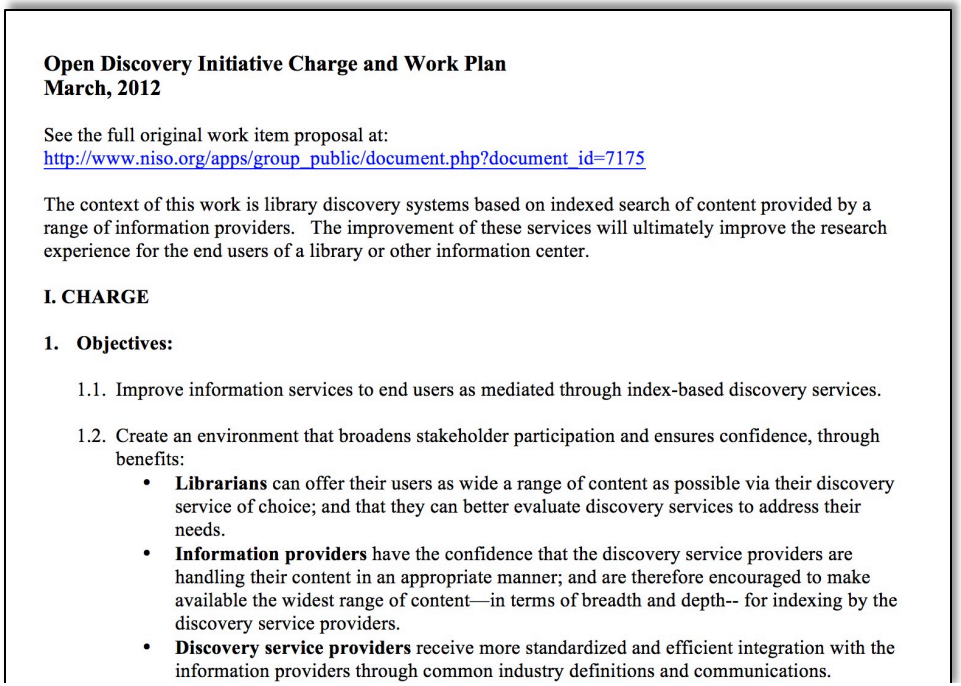
- Interest in a quick process of defining ‘ground rules’/best practices recommendations
- Agreed to approach NISO

NISO Undertaking

- 'Work item' drafted by ALA group
- Considered, discussed, approved by NISO Discovery to Delivery Topic Committee
- Circulated to NISO Voting Members for a 30-day ballot
- After approval, press release and other public announcements
- Working Group formed

ODI Charge and Work Plan

- Charge
  - Objectives
  - Goals
  - Deliverables



Open Discovery Initiative Charge and Work Plan
March, 2012

See the full original work item proposal at:
http://www.niso.org/apps/group_public/document.php?document_id=7175

The context of this work is library discovery systems based on indexed search of content provided by a range of information providers. The improvement of these services will ultimately improve the research experience for the end users of a library or other information center.

I. CHARGE

1. Objectives:

1.1. Improve information services to end users as mediated through index-based discovery services.

1.2. Create an environment that broadens stakeholder participation and ensures confidence, through benefits:

- **Librarians** can offer their users as wide a range of content as possible via their discovery service of choice; and that they can better evaluate discovery services to address their needs.
- **Information providers** have the confidence that the discovery service providers are handling their content in an appropriate manner; and are therefore encouraged to make available the widest range of content—in terms of breadth and depth-- for indexing by the discovery service providers.
- **Discovery service providers** receive more standardized and efficient integration with the information providers through common industry definitions and communications.

Objectives

- Focus on
  - End users
  - Librarians
  - Information providers
  - Discovery service providers
- Foster development of best practices and means of assessment

Goals

- Identify needs and requirements
- Create recommendations and tools for working together
- Enable librarians to assess offerings

Deliverables

- Vocabulary
- NISO Recommended Practice
  - Data format and data transfer
  - Library rights to specific content
  - Level of indexing
  - Fair linking
  - Usage statistics
- Mechanisms to evaluate conformance with recommended practice

Vocabulary

1. Actors

1.1. End User	The final consumer in an information retrieval session.
1.2. Licensor	The institution or individual who has acquired rights to or otherwise obtained access others content or services. Focuses on addressing the legal requirements.
1.3. Licensee (aka Rights Owner)	The rights holder granting search and/or access rights to others.
1.4. Publisher	The organization providing dissemination of literature or information. The publisher may also be the author or creator of the information, but may be an independent organization separate from the creator.
1.5. Content Aggregator	The organization that collects information from varied sources and provides consistent search, presentation and/or access

Vocabulary

2. Methods of Searching

2.1. Central Index (aka Indexed Search)	Method for storing and indexing content in a central location. Disparate content sources are aggregated with consistent formatting, indexing and ranking algorithms.
2.2. Federated Search	Method for searching multiple disparate content sources with one query. Results are coordinated and displayed to the user. For the purpose of this discussion Federated Search and MetaSearch will be used interchangeably.
2.3. Metasearch	Method for searching multiple disparate content sources with one query. Results are coordinated and displayed to the user. For the purpose of this discussion Federated Search and MetaSearch will be used interchangeably.

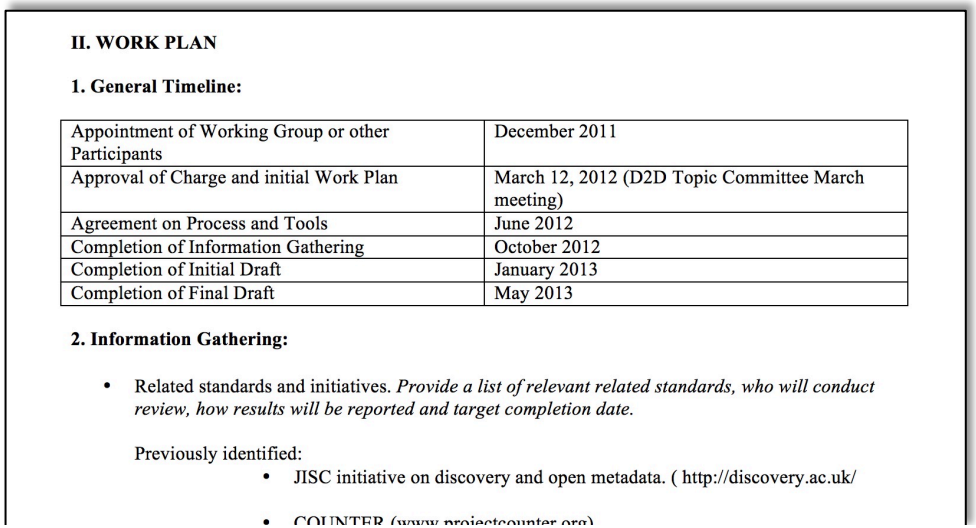
Vocabulary

4. Methods of Data Exchange

4.1. Harvest	Method of extracting indexing and/or fulltext from remote web accessible sites for the purpose of providing search and/or display from a central location. Includes methods like OAI/PMH.
4.2. Syndication	Method of pushing content to remote indexing/abstracting/display services. Examples include feeds from Publishers and access via FTP.
4.3. RSS (Really Simple Syndication)	Standard method for advertising the availability of frequently published content that includes metadata, publication date and authorship information.
4.4. AtomPub	Standard method for publishing syndication feeds similar to RSS. AtomPub provides a more robust method than RSS for publishing beyond blog and web page entries.
4.5. Screen scraped	Simplest method of harvesting content that places no technical burden

ODI Charge and Work plan

- Work Plan
  - Timeline
  - Information gathering
  - Document drafting



II. WORK PLAN

1. General Timeline:

Appointment of Working Group or other Participants	December 2011
Approval of Charge and initial Work Plan	March 12, 2012 (D2D Topic Committee March meeting)
Agreement on Process and Tools	June 2012
Completion of Information Gathering	October 2012
Completion of Initial Draft	January 2013
Completion of Final Draft	May 2013

2. Information Gathering:

- Related standards and initiatives. *Provide a list of relevant related standards, who will conduct review, how results will be reported and target completion date.*

  Previously identified:
  - JISC initiative on discovery and open metadata. (http://discovery.ac.uk/
  - COUNTER (www.projectcounter.org)

Subgroups

- Technical recommendations for data format and data transfer
- Communication of library's rights/Descriptors regarding level of indexing
- Definition of fair linking
- Exchange of usage data

Technical Formats Subgroup

- Chair: Mike Gorrell, EBSCO Publishing
- Members:
  - Marshall Breeding, Independent Consultant
  - Jeff Lang, Thomson Reuters
  - Dave Lindahl, University of Missouri Kansas City
  - Aaron Woods, Alexander Street Press

Technical Formats Subgroup

- Problem
- Stakeholders
- Related Initiatives
- Survey
- Recommended Practices

Libraries' rights/Level of indexing

- Co-chairs: Laura Morse, Harvard University, Ken Varnum, University of Michigan
- Members:
  - Marshall Breeding, Independent Consultant
  - Sara Brownmiller, University of Oregon
  - Mike Gorrell, EBSCO Publishing
  - Jeff Lang, Thomson Reuters
  - Bonnie Lawlor, NFAIS
  - Jenny Walker, Ex Libris
  - Aaron Woods, Alexander Street Press

Libraries' rights/Level of indexing

- Libraries unable to adequately evaluate discovery services and content provider participation because of lack of clarity regarding sources and types of content/variations in content made available
- Created set of data elements useful for all parties – improve transparency
- Survey questions focused on: needs of libraries in decision-making; data available in flow of metadata from content provider to discovery service; better understand barriers to participation for content providers

Libraries' rights/Level of indexing

- Recommendations often follow KBART (metadata fields, text, tab delimited format)
- Content providers make a basic set of metadata elements available for each item they submit for indexing (enhanced metadata options available)
- Discovery providers make available to prospective and current customers sufficient information about content to ensure adequate evaluation -> metadata format prescribed
- Standing Committee should be created to help with education & support
- Future: APIs, 'restricted' content, collection-level reporting

NISO
How the information world CONNECTS

Definition of Fair linking

- Chair: Roger Schonfeld, Ithaka S+R
- Members:
  - Sara Brownmiller, University of Oregon
  - John Law, Serials Solutions
  - Bonnie Lawlor, NFAIS
  - Dave Lindahl, University of Missouri Kansas City
  - Laura Morse, Harvard University

Exchange of Usage data

- Chair: John Law, Serials Solutions
- Members:
  - Jamene Brooks-Kieffer, Kansas State University
  - Lettie Conrad, SAGE Publications
  - Mike Gorrell, EBSCO Publishing
  - Jeff Lang, Thomson Reuters
  - Bonnie Lawlor, NFAIS
  - Jenny Walker, Ex Libris

Exchange of Usage data

- 2 audiences: Content Providers, Libraries
- COUNTER already well-established but COP 4 doesn't include discovery services (but does include metasearch)
- Conducted interviews with stakeholders to develop survey questions – understand motivations on what data would be used and how it would be used
- Survey results
- Recommendations:
  - Terminology
  - Data points
  - Distribution

Next steps

- Consolidation of subgroup reports into one ODI Recommended Practice; polishing
- **30-day Public comment period**
- Working group evaluation of comments, edits to RP, responses
- Discovery to Delivery Topic Committee approval
- NISO Publication

Questions ?

http://www.niso.org/workrooms/odi/

nlagace@niso.org / @abugseye
lindahld@umkc.edu
Roger.Schonfeld@ithaka.org / @rschon

NISO
How the information world CONNECTS