IOTA: Improving OpenURLs Through Analytics

Elizabeth Winter | Adam Chandler
October 13, 2010

Where we’re headed

- NISO IOTA project
- Background
- Problem / Goals
- Benefits
- Current work
- Issues
- Upcoming events / Ways to follow IOTA

Background

- OpenURL standard
- IOTA = Improving OpenURLs Through Analytics Working Group
- History:
  - Adam Chandler @ Cornell
  - L’Année Philologique
  - Mellon Grant
  - NISO proposal

It's Only as Good as the Metadata: Improving OpenURL and Knowledge Base Quality

- Introduction
  - Todd Carpenter, NISO
- IOTA: Improving OpenURLs Through Analytics
  - Elizabeth Winter, Georgia Tech Library
  - Adam Chandler, Cornell University, will join Elizabeth during the Q&A portion
- KBART: Knowledge Bases and Related Tools
  - Andreas Biedenbach, Springer Science+Business Media
  - Sarah Pearson, University of Birmingham, will join Andreas during the Q&A portion
- OpenURLs and Knowledge Base Quality
  - Maria Stanton, Serials Solutions
Poll #1

The Problem

- 72% of respondents to an online survey agreed/strongly agreed that a significant problem for link resolvers is the generation of incomplete or inaccurate OpenURLs by databases.

- Why don’t OpenURLs always work?
Goals

• Two-year research project
• Produce a technical report
• Develop a qualitative report to help OpenURL provider compare their OpenURL quality to that of their peers

IOTA Desired Outcomes

• Recommendations for source vendors
• Qualitative report that will help OpenURL providers
• Recommendations for link resolver vendors?
• Better OpenURLS across the industry
• More success for users

Current Work

• OpenURL reporting system: http://openurlquality.niso.org/
• Data (~ 9 million OpenURLs) from
  – The Claremont Colleges
  – Cornell University
  – Georgia Institute of Technology
  – Kansas State University
  – EBSCO Information Services
  – WorldCat Link Manager
  – Serials Solutions
  – Thomson Reuters

Analytics / Reports

• Produce reports showing:
  – Elements that are showing up in OpenURLs
  – Patterns of elements that are showing up in OpenURLs
  – Frequency with which OpenURL elements (e.g., article title, ISSN, DOI, author last name, startpage, endpage, etc.) show up in OpenURLs from source databases and source vendors
Current Work

• Vendor completeness index:
  – Testing the assumption that more information makes for a better OpenURL
  – Can we create a relative ranking of source vendors based on how complete their OpenURLs are?

• Element weighting:
  – Which elements are present?
  – Are some more important than others?
  – How “good” are a particular vendor’s OpenURLs?

Issues

• Gathering and examining a lot of data
• Do the data accurately represent the user experience?
• Which OpenURLs (ejournals? ebooks? proceedings? everything?)
• Weighting the elements
• Role of link resolvers/resolver settings
• Case studies

Upcoming Events

• Charleston Conference:
  Date: Friday, Nov. 5 @ 2:00 PM
  Location: Francis Marion, Parkview Room
  Speakers: Rafal Kasprowski - Rice University & Susan Marcin - Columbia University

• … & hopefully other conferences soon. Stay tuned.
For more info

Reports:
http://openurlquality.niso.org/
NISO page:
http://www.niso.org/workrooms/openurlquality
Blog:
http://openurlquality.blogspot.com/

On Twitter

Follow IOTA on Twitter:
http://twitter.com/nisoiota
or @nisoiota

To post comments on the IOTA page via Twitter,
use
#nisoiota

Agenda

– Who – Publishers, aggregators, KB vendors, libraries
– What – a universally acceptable holdings data format
– Where – throughout the supply chain & at the UKSG info hub http://www.uksg.org/kbart
– When – Now
  • Phase 2 in progress
– Why – Better access for users through accurate holdings data
– How can you help?
  • For librarians
  • For publishers

Who is behind KBART?

• Standards organizations
  – UKSG and NISO
• The working group members, representing:
  – Knowledge Base vendors
    • Ex Libris, Serials Solutions, EBSCO
  – Content aggregators
  – Publishers
  – Subscription agents
  – Libraries & consortia
• Full list -- http://www.uksg.org/kbart/members

Phase II Working Group

• Jason Price – Claremont Colleges / CDL
• Gary Pollack – Cengage Learning/Gale
• Elizabeth Stevenson – Edinburgh University
• Chad Hutchens – University of Wyoming
• Sarah Pearson – University of Birmingham
• Paul Moss – OCLC
• Sheri Meares – EBSCO
• Christine Stohn – Ex Libris
• Sherrard Ewing – Serials Solutions
• Matthew Llewellyn – Royal Society
• Andreas Biedenbach – Springer
• Mariete Heins – Swets
• Ruth Wells – Taylor & Francis
• Rose Robinson – Publishing Technology
What is KBART?

- **Knowledge Bases And Related Tools**
- UKSG and NISO collaborative project
- UKSG 2007 research report: *Link Resolvers and the Serials Supply Chain*
- “A set of practical recommendations for the timely exchange of accurate metadata between content providers and knowledge base developers”

### A simple metadata exchange format...

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>publication_title</td>
<td>Publication title</td>
</tr>
<tr>
<td>print_identifier</td>
<td>Print format identifier (e.g., ISSN, ISBN, etc.)</td>
</tr>
<tr>
<td>online_identifier</td>
<td>Online format identifier (e.g., eISSN, eISBN, etc.)</td>
</tr>
<tr>
<td>date_first_issue_online</td>
<td>Date of first issue available online</td>
</tr>
<tr>
<td>num_first_vol_online</td>
<td>Number of first volume available online</td>
</tr>
<tr>
<td>date_first_issue_online</td>
<td>Number of first issue available online (or blank, if coverage is in progress)</td>
</tr>
<tr>
<td>date_last_issue_online</td>
<td>Date of last issue available online (or blank, if coverage is in progress)</td>
</tr>
<tr>
<td>num_last_vol_online</td>
<td>Number of last volume available online (or blank, if coverage is in progress)</td>
</tr>
<tr>
<td>title_url</td>
<td>Title-level URL</td>
</tr>
<tr>
<td>title_id</td>
<td>Title ID</td>
</tr>
<tr>
<td>embargo_info</td>
<td>Embargo information</td>
</tr>
<tr>
<td>coverage_depth</td>
<td>Coverage depth (e.g., abstracts or full-text)</td>
</tr>
<tr>
<td>coverage_notes</td>
<td>Coverage notes</td>
</tr>
<tr>
<td>publisher_name</td>
<td>Publisher name (if not given in the title)</td>
</tr>
</tbody>
</table>

### Where does KBART apply?

- **Representing years of thinking...**
- **What’s in your knowledgebase...the hard way**
- **Where does KBART apply?**
- **Why KBART?**

### Why KBART?

- Maintenance of accurate package content coverage data
  - Supports OpenURL Link Resolvers
  - Supports ejournal MARC record delivery services
  - Enables automated updating by KB providers
- Addresses common holding list inadequacies
  - Re-use of ISSNs
  - Embargo period ambiguities
  - Inconsistent date/enumeration formats
**How Librarians can help**

- Lobby publishers to adopt the KBART practices
  - LEARN about what KBART is and what it does
  - INSIST on “knowing” what you are buying!
    - Require delivery of a usable holdings list
      - Initially & annually going forward
    - When the list is inadequate, point them to KBART
    - Only your insistence will make it happen
  - ENABLE publisher sales staff to make the case to their company
  - FOLLOW UP with continued requests as necessary

**Endorsers**

- American Institute of Physics, Ex Libris, OCLC, and Serials Solutions
  - Announced by mid-September 2010: Alexander Street Press, Annual Reviews, EBSCO Information Services, Innovative Interfaces, and Royal Society Publishing

**What to do to be endorsed?**

- Submit a sample file to the KBART working group at kbart@niso.org
- Endorsement is finalized once the file’s format and content have been reviewed and approved and the provider has made it publicly available (in line with the recommendations).

**Providing KBART formatted data**

**What do publishers need to do to adopt the KBART best practices?**

1. Review the requirements that are accessible via http://www.uksg.org/kbart/s5/transition.
2. Format ejournal and ebook content availability data to meet the requirements.
3. Self check your datasheet(s) on the KBART website to ensure that they conform to the recommended practice and make any necessary corrections.
4. Ensure that you have a process in place for regular data exchange as outlined in section 5.2 of the KBART report.
5. Register your organization on the KBART registry website (http://bit.ly/kbartregistry) and provide a link to download the newly KBART formatted dataset(s)

**The Registry -- a contact and metadata content clearinghouse**

Phase II Action Items

- Endorsement / Compliance / Engagement
- Definitions for global vs local updates
- Consortia-specific metadata transfer
- Institution-specific metadata transfer
- Review of metadata transfer for e-books
- Open access material

Knowledgebase Management Before KBART

Proactive reconciliation of an ejournal package list

- General Process – library, consortium or KB vendor
  - Request updated access list from publisher
  - Sample publisher list for accuracy
  - Translate publisher list to match KB list
    - Number of titles never matches
    - Perform ISSN match with MS Access
    - Watch for & integrate title changes, mergers, acquisitions and losses
    - Watch for publisher-reuse of ISSN/title combinations
    - Identify date discrepancies manually (inconsistent formats)
  - Decide when it is "good enough" and go live/distribute new list
  - Lather, Rinse, Repeat

Our vision...after KBART

Phase 1 – Universally accepted standardized publisher metadata, regularly distributed AND available on demand
Phase 2 – Broad adoption, more content type coverage
Phase 3? – Consortia & institution level holdings metadata distribution based on what is actually accessible from a particular IP

Learn more

- [http://www.niso.org/workrooms/kbart](http://www.niso.org/workrooms/kbart)
- [www.uksg.org/kbart](http://www.uksg.org/kbart)
- Andreas Biedenbach – KBART co-chair
  andreas.biedenbach@springer.com
- KBART interest group
  [http://www.niso.org/lists/kbart_interest/](http://www.niso.org/lists/kbart_interest/)

OpenURLs & Knowledge Base Quality:

Collecting & Improving Knowledge Base Data

Maria Stanton
Director, Content Operations
October 13, 2010
Agenda

- Patron Perspective
- Referring Source Data & the Knowledge Base
- Target Source Data & the Knowledge Base
- “And the world would be a better place ... ”

Patron Perspective ... link leads to an article.

Click or two away ...
Link should get them to the full text they seek ...

Click or two away ...
Link should get them to the full text they seek ...

Patron perspective ...

When the link doesn’t lead to the treasured resource ... they have no idea why ... all they know is, it didn’t work ...
Data management and OpenURL

1 What does the referring source provide to support article level links?
2 What data does a typical target source provide to knowledge base providers?
3 Knowledge base provider – our efforts to support the patron?

OpenURL Standard – article level

Key parameters from the referring source ...

<table>
<thead>
<tr>
<th>Key</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>rft.title</td>
<td>Article title</td>
</tr>
<tr>
<td>rft.title or rft.jtitle</td>
<td>Journal title (v 1.0 prefer title)</td>
</tr>
<tr>
<td>rft.date</td>
<td>Date of publication</td>
</tr>
<tr>
<td>rft.volume</td>
<td>Volume designation (number, roman numerals or non-numeric)</td>
</tr>
<tr>
<td>rft.issue</td>
<td>Issue designation (number or non-numeric)</td>
</tr>
<tr>
<td>rft.spage</td>
<td>Start page</td>
</tr>
<tr>
<td>rft.ssn or rft.eissn</td>
<td>Standard identifier</td>
</tr>
<tr>
<td>rft.id</td>
<td>Identifiers unambiguously specific the referent article.</td>
</tr>
<tr>
<td>Parameters</td>
<td>DOI (Digital Object identifiers) or PMID (PubMed identifiers)</td>
</tr>
</tbody>
</table>

What does a typical knowledge base provider track

Full text coverage at the journal level

Title
Issn or elssn
Full text start date
Full text end date
Journal-level URL
Embargo information

Knowledge base – Sources of holdings data

Electronic files:
% are provided in a standard format (e.g., KBART)
% are content provider specific
Manually harvest: % of providers do not offer electronic files
Limitations of the files

Most files don’t include the full title history ...

EXAMPLE
A system that doesn’t identify all three titles won’t link correctly.

Provider Title
Freeman / 1559-1638, with coverage from June 1998 to present
Actual Title Split
Freeman / 0016-6852, coverage from June 1998 to Dec 1999
Ideas on Liberty / 1543-9888, coverage from Jan 2000 to Nov 2003
Freeman (2003) / 1559-1638, coverage from Dec 2003 to present

Authority data management ...

Titles & the accuracy of ISSN varies from provider to provider ...

Title: Journal of the American Society of Nephrology

<table>
<thead>
<tr>
<th>Provider List</th>
<th>Corrections</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. Am. Soc. Nephrol</td>
<td>1046-6673</td>
<td>✔️ Manual Title</td>
</tr>
<tr>
<td>JASN</td>
<td>1046-6673</td>
<td>✔️ Manual Title</td>
</tr>
<tr>
<td>Journal of the American Society of Nephrology</td>
<td>Missing ISSN</td>
<td></td>
</tr>
<tr>
<td>Journal of the American Society of Nephrology</td>
<td>1046-6673</td>
<td>✔️ DK</td>
</tr>
<tr>
<td>Journal of the American Society of Nephrology</td>
<td>1400-5665</td>
<td>❌ Error</td>
</tr>
<tr>
<td>Journal of the American Society of Nephrology</td>
<td>1553-3450</td>
<td>✔️ eISSN</td>
</tr>
<tr>
<td>Journal of the American Society of Nephrology: JASN</td>
<td>1046-6673</td>
<td>✔️ Manual Title</td>
</tr>
</tbody>
</table>

Results of the authority data management

Library’s profile is accurate
Authoritative title and ISSN passed to target source

Target source ...

After referring source and knowledge base hand shakes, need to call the target source
Knowledge base management focuses on journal title & coverage date issues
Article title is typically problematic

Other examples ...

Title Histories – Provider reports following coverage:
- R&D Magazine 1994 to the present
- R&D Magazine was Research & Development until 2002

Coverage dates – Provider reports the following:
- Public Relations Quarterly – full text 1965 to present
- Full text coverage ended in 2007

Target source – still resolve to correct item

Link resolvers rely on IF/THEN scenarios
- If a DOI is provided, use DOI to create article link
- If DOI is not provided, but ISSN, VOLUME, ISSUE, and SPAGE are, use that metadata to provide an article level link.
Question

Can we come together to do more?

Helping Your Library
Be The Best Partner For Research

Thank You

Contact us:
info@serialssolutions.com

Visit us:
www.serialssolutions.com

SerialsSolutions
A ProQuest Company

It's Only as Good as the Metadata:
Improving OpenURL and Knowledge Base Quality

Questions?

All questions will be posted with presenter answers on the NISO website following the webinar:

Thank You!

Thank you for joining us today. Please take a moment to fill out the brief online survey.

We look forward to hearing from you!