NISO'S IOTA Initiative:
Fixing OpenURL Links Using Data Analysis
Improving OpenURLs Through Analytics

Special Libraries Association, Annual Conference,
Chicago, IL, July 17, 2012
American Library Association, Annual Meeting,
Anaheim, CA, June 24, 2012

Rafal Kasprowski, Rice University
What is IOTA?

- Initiative that measures the relative importance of the elements that make up OpenURL links to help vendors improve their OpenURL strings so that the maximum number of OpenURL requests resolve to a correct record.

Elements:
- journal title
- book title
- ISBN
- ISSN
- start page
- DOI
- volume
- author
- PMID
- issue
- date
- ...
Presentation

• Part I (Rafal Kasprowski, Electronic Resources Librarian, Rice University)
  – History of OpenURL and IOTA
  – IOTA’s Deliverables:
    • OpenURL Reports (comparing vendors’ OpenURL links)
    • OpenURL Quality Index (preliminary version)

• Part II (Susan Marcin, Licensed Electronic Resources Librarian, Columbia University)
  – Usefulness of IOTA’s OpenURL Reports in improving OpenURL links for e-books

• Part III (Oliver Pesch, Chief Strategist, EBSCO Information Services)
  – Improvements to IOTA’s OpenURL Quality Index and its limitations
Part I - Agenda

• Full-text linking: from proprietary linking to OpenURL

• From Cornell to NISO: IOTA created in response to OpenURL linking problems

• IOTA in the context of ERM best practices: IOTA and KBART

• IOTA’s analytical approach

• Reports comparing vendor OpenURLs

• Weighting OpenURL elements

• Concept of the OpenURL Quality Index and preliminary version.
Before OpenURL: Proprietary Linking

• Certain A&I database providers (e.g., CSA, PubMed) offered full-text linking options for a select number of content providers.

• Libraries manually activated full-text linking with providers they had subscriptions with.

• A&I --> Full Text
Proprietary Linking: Pros and Cons

- Linking had to be activated manually by libraries for each full-text provider.
- A&I providers offering this option were few.
- Selection of full-text providers was limited.

But...
- Once set up, the static links to full texts were accurate.
- Problem source pinpointed easily: A&I --&gt; Full Text
Advent of OpenURL

• Objective: Deliver full texts unrestrained by proprietary silos.

• Open standard generating dynamic links at time of request.

• A-Z list (e.g., e-journals, e-books):
  o Knowledge base (KB) with library's holdings.
  o Replaces librarian as intermediary in linking.
  o Indicates provider of "appropriate copy"

• A&I ("Source") --> A-Z list ("KB") --> Full Text ("Target")
Source Citation


Target Link (example using OpenURL syntax, similar to Source OpenURL)

Example of Resolver Menu Page

Pros & Cons of OpenURL

Pros:
• KB/Resolver vendors took over most of the linking setup: Less work for libraries and providers.
• Participation by A&I platforms and full-text providers exceeded proprietary linking: OpenURL scales better.

Cons:
• Dynamic linking less predictable than static linking: more difficult to pinpoint cause of link failures.
• OpenURL linking not improved significantly in last 10 years.
• No systematic method exists to benchmark OpenURLs.
Identifying source of problem…

"72% of respondents to the online survey either agreed or strongly agreed that a significant problem for link resolvers is the generation of incomplete or inaccurate OpenURLs by databases (for example, A&I products)."


Defining methodology for approaching problem

Recently, researchers have indicated the need for metadata quality metrics, including:

- completeness;
- accuracy;
- conformance to expectations;
- logical consistency and coherence.

Année philologique OpenURL Study

2008 Cornell study led by Adam Chandler*
• Problem: Too often links sent from Aph did not successfully resolve to requested resource.
• Objective: Examine quality of OpenURLs offered to users by Aph in order to improve the linking.

Aph Study investigated:
• Faulty citation metadata from source database.
• Method to evaluate the OpenURLs.

Concept of scoring in Aph study (based on B. Hughes study)*

- establish a baseline for comparison;
- results to be shared with data providers;
- develop a best practice.

Problem analysis in Aph study limited to:

- source link
- presence/absence of citation metadata elements

Results:

- OpenURL quality model: compares elements in Aph OpenURLs to those of other providers.
- No scoring was achieved for Aph, but model is first step towards scoring system.

IOTA is formed in January 2010

- NISO accepts proposal to take *Aph* Study to wider community.
- openurlquality.org

**what is IOTA?**

IOTA is an initiative that makes use of log files from various institutions and vendors to analyze element frequency and patterns contained within OpenURL strings.

The reports created from this analysis inform vendors about where to make improvements to their OpenURL strings so that the maximum number of OpenURL requests resolve to a correct record.

**run report**

Number of openuris analyzed

20,206,538
IOTA & KBART: complementary NISO working groups

IOTA
• Deals with issues specific to OpenURL linking;
• Seeks improvements in OpenURL elements used by:
  – OpenURL providers.

KBART
• “Knowledge Bases And Related Tools”
• Deals with data issues at the KB level
• Seeks improvements in data exchange practices between:
  – content providers (e.g. OpenURL providers);
  – product vendors (e.g. link resolver vendors).
  – subscription agents;
IOTA & KBART: related through OpenURL

- **IOTA:**
  - analyzing data sent from OpenURL source to link resolver.

- **KBART:**
  - creating best practices for data formats sent from content providers to knowledge base (and link resolver) vendors.
IOTA’s Basic Assumptions

• Results are achieved through an analytical investigation of how OpenURL links work.

• Practical: Not the OpenURL standard is addressed, but links (OpenURLs) generated by standard.

• Selective changes to OpenURLs will lead to significant improvement in linking success rate.
  - Motto: "small changes. big improvements"
IOTA’s Desired Outcomes…
…a continuation of *Aph* Study

A. Produce qualitative reports that will help OpenURL providers quickly compare their OpenURL quality to that of their peers.

B. Develop community-recognized index for measuring the quality of OpenURL links generated by content providers:

   - scalable across all OpenURLs and their providers
(A) Usefulness of comparing OpenURLs

- **Content providers** that generate OpenURLs can:
  - compare their OpenURLs with other providers;
  - make improvements to their OpenURLs.

- **Institutions** can:
  - compare OpenURL providers;
  - make local adjustments to OpenURL setup.

- **Resolver vendors** can:
  - compare OpenURL providers;
  - Change their settings for OpenURL providers:
    - Link resolvers;
    - Web-scale discovery products.
OpenURL Reports

Number of openuris analyzed
20,206,538

Select Report Type

1. Choose log file provider
   - American Institute of Physics
   - The Claremont Colleges
   - Cornell University
   - EBSCO Information Services
   - EDINA
   - Georgia Institute of Technology
   - Hispanic American Periodicals Index (HAPI)
   - Kansas State University
   - McGill University
   - WorldCat Link Manager
   - Serials Solutions
   - Thomson Reuters

2. Choose report type
   - completeness
   - metric
   - source
Report types

• **Source reports**
  – Viewing how a particular (1) vendor or (2) database
    • A. uses OpenURL elements (element frequency)
    • B. formats OpenURL elements (pattern frequency)

• **Element / Pattern reports**
  – Viewing how a particular (1) element or format
    • A. is used across vendors
    • B. is used across databases

• **Vendor Completeness Report**
  – Viewing vendors’ OpenURL quality score
(B) OpenURL Quality Index:
Rating vendors by their OpenURLs

1. Core Elements:
   • Any element contained in IOTA's OpenURL reporting system;
   • 20M OpenURLs obtained from libraries & content providers.

2. Scoring System:
   • Assumption: Correlation exists between
     o # of core elements ("OpenURL completeness") &
     o ability of OpenURLs to link to specific content.

3. Element Weighting:
   • Assigned based on their relative importance:
     o spage vs atitle
     o issn vs jtitle
     o doi/pmid vs date, etc.
OpenURL Quality Index
preliminary version

<table>
<thead>
<tr>
<th>element</th>
<th>count (# of OpenURLs in sample that contain this element)</th>
<th>weighted points</th>
<th>count * weighted points</th>
</tr>
</thead>
<tbody>
<tr>
<td>artnum</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>atitle</td>
<td>72</td>
<td>1</td>
<td>72</td>
</tr>
<tr>
<td>aulast</td>
<td>88</td>
<td>1</td>
<td>88</td>
</tr>
<tr>
<td>btitle</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>date</td>
<td>107</td>
<td>5</td>
<td>535</td>
</tr>
<tr>
<td>doi</td>
<td>87</td>
<td>8</td>
<td>696</td>
</tr>
<tr>
<td>eissn</td>
<td>23</td>
<td>3</td>
<td>69</td>
</tr>
<tr>
<td>isbn</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>issn</td>
<td>78</td>
<td>3</td>
<td>234</td>
</tr>
<tr>
<td>issue</td>
<td>69</td>
<td>3</td>
<td>207</td>
</tr>
<tr>
<td>jtitle</td>
<td>96</td>
<td>1</td>
<td>96</td>
</tr>
<tr>
<td>pmid</td>
<td>80</td>
<td>8</td>
<td>640</td>
</tr>
<tr>
<td>spage</td>
<td>107</td>
<td>3</td>
<td>321</td>
</tr>
<tr>
<td>title</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>volume</td>
<td>107</td>
<td>3</td>
<td>321</td>
</tr>
</tbody>
</table>

Total possible points = OpenURLs in sample * total weighted points for all elements
Rating = points for provider during period / total possible points for provider during period

Total possible points = 5040
Rating = 0.65
Further investigation was needed

• Element weighting needed to be adjusted in a more systematic way:
  o Importance of identifiers (doi, pmid) vs bibliographic data (issn, volume, spage, etc.)
  o Relative importance of bib. data (issn vs volume vs spage, etc.)

• IOTA focused on OpenURLs from citation sources only. How is OpenURL linking impacted by other factors?
  o knowledge base,
  o resolver,
  o full-text provider (target).

• High "completeness" score of OpenURLs not always indicative of "success" in linking to full texts
  o Combination of indexes, incl. “success index”, developed by IOTA and/or other groups may lead to more precise metrics.
Presentation: Parts II and III

• Part II (Susan Marcin, Columbia University)
  – Usefulness of IOTA’s OpenURL Reports in improving OpenURL links for e-books

• Part III (Oliver Pesch, EBSCO Information Services)
  – Improvements to IOTA’s OpenURL Quality Index and its limitations