Dissertations and OpenURL
State of the art / future state

SFX Track session S4

ELUNA, May 6, 2009

Geoff Sinclair
Education Centre Library
Nipissing University / Canadore College
goeffs@nipissingu.ca
Slides: http://www.eclibrary.ca/eluna/


**Background**

1 **Problem**

OpenURL links from many sources form links to dissertations as if they were links to a fulltext article in a journal *Dissertation Abstracts*, one of its variants, or a similar paper publication that abstracts dissertations. Thus standard targets, such as PROQUEST_DISSERTATION_AND_THESES, fail, because their services are only exposed if the genre is “dissertation”.

2 **Workaround**

We created a local target, called PROQUEST_DISSERTATION_AND_THESES_LOCAL, and a getFullTxt service under the target. We checked off Object Lookup in the service and added object portfolios for many of the variant ISSNs for Dissertation Abstracts. So, when an incoming request for a dissertation seems to be for an article in a journal with one of these ISSNs, the getFullText service is exposed in the SFX menu. We used a custom parser to process requests, which sends a title search out to the database ProQuest Dissertations and Theses (PQDT).

In 2005, we implemented this solution, which was based on a parser written by Andrew Forman, who was then at the University of Iowa. His parser used object portfolios get the parser to expose a getAbstract service. Andrew’s parser combined the first nine words to pass along as the title search. In Perl terms, it used the first nine clusters of non-whitespace characters. So, if parentheses were contained within the first nine words, PQDT had problems processing the search.

In 2007, we revised the parser for several reasons:

- Ex Libris had released a new parser for ProQuest databases.
- We changed the service from getAbstract to getFullTxt.

By this time, ProQuest had added so many PDFs to the database that it had tipped the balance from a database that was mainly abstracts with a bit of full text, to a database that was mainly full text, especially for recent years, with occasionally only the abstract. So at this point, we changed the service that was exposed from getAbstract to getFullTxt. Today, it would probably be better to call it a “get selective full text” service.

As in the 2005 version, the service would appear whenever one of the many Dissertation Abstracts ISSNs was passed to SFX. Unlike the 2005 version, the 2007 parser would submit the entire title as a title search for ProQuest Dissertations and Theses. We did remove parentheses, since this caused problems with the search, but the search failed with very long titles. A combined solution would work best: one that removes problematic punctuation and truncates the title search submitted to PQDT.
Current investigations

1 Workarounds reconsidered: “The workaround has become the work”
Has the workaround become the work? By pursuing this approach, were we missing a more effective solution that would address the problem at a more fundamental level? Was the workaround successful on its own terms? What could vendors do to improve linking to dissertations?

2 Alternative to a title search: Shared IDs
Since many of the issues with the workaround centre on failure in the title search, could we link using a unique identifier? Dependable OpenURL linking requires shared, mostly unique keys. These keys can be a unique ID, such as a DOI or ISBN. We normally use a compound key for OpenURL (ISSN/year/volume/issue/start page), but this isn’t necessarily unique because we can have more than one article on a page. I would call this a “mostly unique” key.

To dependably get from the source to the target, we need to have a shared key in both the source and the target. There’s many different sources for OpenURLs, but our current investigations examined only one target: ProQuest Dissertations and Theses (or PQDT). In this target we see three candidates for unique keys:

- the ProQuest Document ID
- Publication/Order No. [UMI number]
- ISBN

We did not find DOIs in PQDT. Perhaps this is because DOIs are not typically assigned to dissertations. In the sources examined, we did not find the ProQuest Document ID, occasionally saw ISBNs, and frequently found we found the UMI number in one form or another.

For now, the best choice is the UMI number, for two reasons:

1. The PDF is produced as an outcome of the microfilming process, so the UMI number appears on the document itself.
2. The UMI number was most frequently found shared key.

However, perhaps DOI or ISBN is where we should be heading. Shouldn't we, in the 21st century, consider dissertations and theses as “born-digital” objects. Should we still be using a microfilm accession number as the primary identifier? Perhaps what's really needed is an overhaul of the dissertation preservation workflow. While this subject merits further discussion, it is outside the scope of our current investigations.
Targets investigated

1 EBSCO’s Library, Information Science & Technology Abstracts

Unique keys
There are a number of dissertations indexed here, and in each record has an accession number. Unfortunately the accession number is not in PQDT. This leaves the combined key of ISSN/year/volume/issue/start page. But, while the journal name is shown, there is no ISSN.

Linking problems
- No ISSN
- genre=bookitem

I’m not sure why LISTA thinks Dissertations Abstracts International is a book. Maybe they were looking at a bound back issue. Since there’s no ISSN, the custom parser would never even see this OpenURL. We could do a workaround and create another target with a parser that searches the titles of all bookitems for the title Dissertation Abstracts International. It’s a particularly ugly workaround, but it does work.

LISTA Example

OpenURL: http://sfx.scholarsportal.info/nipissing?genre=bookitem&isbn=&issn=&title=Dissertation+Abstracts+International&volume=48&issue=4&date=19871001&atitle=An+evaluation+of+the+organization+of+some+European+business+school+libraries,+and+the+services+they+supply%2c+and+the+services+they+supply&aulast=Oort%2c+B&page=633&sid=EBSCO:Library%2c+Information+Science+%26+Technology+Abstracts&pid=%3Ca+authors%3EOort%2c+B%3C/authors%3E%3Cui%3ESTA2302861%3C/db%3E
Illustration 2: LISTA SFX Menu

Illustration 3: Add a local object (identified by Title only)

Illustration 4: Add the local object to the list of Portfolios
Illustration 5: SFX Menu, Local target now appears

Illustration 6: LISTA: Success! [But no fulltext]

2 MathSciNet

Unique keys

Each dissertation record has an accession number. As in LISTA, the accession number is not in PQDT. This leaves the combined key of ISSN/year/volume/issue/start page. But, while the journal name is shown, there is no ISSN.

Linking problems

Mathematics dissertations are listed in Dissertation summaries in mathematics [1086-962X]. The local target didn't appear on the SFX menu because I had only added variants of Dissertation Abstracts. Adding a portfolio for Dissertation summaries in mathematics solved the problem.
MathSciNet Example

A homogenization technique for the development of mesoscopic scale models for chemical vapor deposition. (English summary)


Summary: This dissertation presents a problem arising from the simulation of gas flow over microstructured surfaces. For the industrial application under consideration, the problem is appropriately given as a time-dependent nonlinear reaction-diffusion equation on a domain, which includes a flux condition on a boundary surface consisting of a microscopic fine structure. An equivalent problem for the bulk solution is derived, which incorporates all physical quantities of interest and which is accessible to efficient numerical simulation at the same time. This is achieved by applying a homogenization technique.

Illustration 7: MathSciNet, example dissertation entry


Illustration 8: MathSciNet, SFX menu

3 EBSCO America History and Life

Unique keys
Each dissertation record has an accession number and an entry number. Neither of these are in PQDT. There are two unique keys in America History and Life that are also found in PQDT.

- The combined key of ISSN/year/volume/issue/start page.
- A variant of the UMI number.

Linking problems
The SFX button does not appear in the source record. The Custom Link settings must be changed in the EBSCOhost admin interface. In the Mandatory Fields value: the SFX link will only appear if the JOURNAL field is in the EBSCO record. What is the JOURNAL field? Possibly the Journal Title field within EBSCO, but that doesn’t explain why the first example worked. However, when we remove this Mandatory Field condition, the SFX menu will appear.
**CustomLinks**

EBSCO has created a highly customizable feature called CustomLinks. This could be one of the ways our efforts should be spent to take OpenURL linking to the next level: source database vendors could give us the tools to help customize when OpenURL links should appear, and also which data are passed through to the OpenURL resolver.

It does answer, though, why Dissertations Abstracts International was given the genre “bookitem” in the first example: anything without an ISSN is a bookitem.

**America History and Life Example**

Illustration 12: AHL: Example Search Results Screen

Illustration 13: AHL: Example Detail Screen. Note the UMI number appears within the Source field
**Illustration 14: EBSCOhost admin interface: CustomLinks setup/maintainance**

<table>
<thead>
<tr>
<th>Name</th>
<th>Category</th>
<th>Owned By</th>
<th>Original Name</th>
<th>Originated By</th>
<th>Remove</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprint</td>
<td>Full Text</td>
<td>EBSCO</td>
<td>Preprint (journal content only)</td>
<td>EBSCO</td>
<td></td>
</tr>
<tr>
<td>resolver.scholarsportalinfo</td>
<td>Other</td>
<td>Self</td>
<td>resolver.scholarsportalinfo</td>
<td>Self</td>
<td></td>
</tr>
<tr>
<td>SFX</td>
<td>Full Text</td>
<td>Self</td>
<td>SFX</td>
<td>Self</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.freemedicaljournals.com">www.freemedicaljournals.com</a></td>
<td>Other</td>
<td>Self</td>
<td><a href="http://www.freemedicaljournals.com">www.freemedicaljournals.com</a></td>
<td>Self</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.ihealthreve.com">www.ihealthreve.com</a></td>
<td>Other</td>
<td>Self</td>
<td><a href="http://www.ihealthreve.com">www.ihealthreve.com</a></td>
<td>Self</td>
<td></td>
</tr>
</tbody>
</table>

**Illustration 15: EBSCOhost admin interface: CustomLinks detail screen. Remove the value in the Mandatory Fields field.**

**Illustration 16: AHL: the SFX link now appears**

**OpenURL:** [http://sfx.scholarsportal.info/nipissing?genre=article&isbn=&issn=04194209&title=&volume=68&issue=8&date=20080101&atitle="Perfecting+Nature%27s+Food%3a+A+Cultural+and+Environmental+History+of+Milk+in+the+United+States,+1900-1970%27.&aulast=Smith-Howard%2c+Kendra+D.&spage=&sid=EBSCO:America%3a+History+%26+Life&pid=3Cauthors%3ESmith-Howard%2c+Kendra+D.%3Cdb%3E America%3a+History+%26+Life%3Cdb%3E](http://sfx.scholarsportal.info/nipissing?genre=article&isbn=&issn=04194209&title=&volume=68&issue=8&date=20080101&atitle="Perfecting+Nature%27s+Food%3a+A+Cultural+and+Environmental+History+of+Milk+in+the+United+States,+1900-1970%27.&aulast=Smith-Howard%2c+Kendra+D.&spage=&sid=EBSCO:America%3a+History+%26+Life&pid=3Cauthors%3ESmith-Howard%2c+Kendra+D.%3Cdb%3E America%3a+History+%26+Life%3Cdb%3E)
Illustration 17: PQDT appears on SFX Menu

Illustration 18: PQDT: title search fails (hyphen in between 1900-1970 is indexed as a dash)
CustomLinks setup and syntax

Values can dynamically be inserted into the CustomLinks query string. Here's the full example, which is only partially visible in the above illustration:

```
genre={IfNotEmpty({ISSN},article,bookitem)}&isbn={ISBN}&issn={ISSN}&title={IfNotEmpty({journal},{journal},{IfNotEmpty({booktitle},{booktitle},{IfNotEmpty({book},{book},{seriestitle})})})&volume={VOLUME}&issue={ISSUE}&date={DATE}&atitle={TITLE}&aulast={IfNotEmpty({bookauthor},{bookauthor},{author})}&spage={strtpage}&sid=EBSCO:{database}&pid=<authors>{IfNotEmpty({bookauthor},{bookauthor},{author})}</authors><ui>{AN}</ui><date>{DATE}</date><db>{database}</db>
```

Here's the same example, with whitespace and comments to clarify:

```
isbn= {ISBN}  # Straightforward value assignment
genre = (IfNotEmpty((ISSN), article, bookitem)  # Conditional assignment

          , bookitem
          )  # Condition met
          )  # Condition not met

title= (IfNotEmpty((journal),  # Nested conditional tests
          {journal},
          (IfNotEmpty((booktitle),
            {booktitle},
            (IfNotEmpty((book),
              {book},
              {seriestitle}
            )
            )
          )
        )
      )
```

Illustration 19: CustomLink setup in EBSCO
4 CSA PsycINFO

Unique keys

There are a large number of dissertations indexed here. This leaves the combined key of ISSN/year/volume/issue/start page. But, while the journal name is shown, there is no ISSN.

Linking problems

CSA PsycINFO Example

Illustration 20: CSA PsycINFO

OpenURL: http://sfx.scholarsportal.info/nipissing?sid=CSA:psycinfo-set-c&pid=%3CAN%3E2009-99011-001%3C%2FAN%3E%26%3CPB%3EUS%3A%20ProQuest%20Information%26%20Learning%3C%2FPB %3E%26%3CDA%3EAAI3322819%3C%2FDA%3E%26%3CPY%3E2009%3C%2FPA%3E%26%3CAU %3EMarcum%2C%20Catherine%20Davis%20(1)%3C%2FAU%3E&issn=0419-4209&isbn=9780549741572&spage=2886&date=2009&genre=book&aulast=Marcum&aufirst=Catherine&auinitm =D&title=Adolescent%20online%20victimization%3A%20A%20test%20of%20Routine%20Activities%20Theory

Note in the OpenURL that genre=book, and an ISBN is present. With these values present, the PQDT full text service does not appear. However, if these values are excised from the OpenURL, the menu does appear. I.e.,

http://sfx.scholarsportal.info/nipissing?sid=CSA:psycinfo-set-c&pid=%3CAN%3E2009-99011-001%3C%2FAN%3E%26%3CPB%3EUS%3A%20ProQuest%20Information%26%20Learning%3C%2FPB %3E%26%3CDA%3EAAI3322819%3C%2FDA%3E%26%3CPY%3E2009%3C%2FPA%3E%26%3CAU %3EMarcum%2C%20Catherine%20Davis%20(1)%3C%2FAU%3E&issn=0419-4209&spage=2886&date=2009&aulast=Marcum&aufirst=Catherine&auinitm=D&title=Adolescent%20online%20victimization%3A%20A%20test%20of%20Routine%20Activities%20Theory

It is unclear why this happens. However, if more sources pass ISBNs in the OpenURL, this would open the door to using ISBNs as a unique identifier for dissertations in the future, using a portfolio-based target service where the dissertations, rather than Dissertation Abstracts, are the objects.
Illustration 21: CSA PsycINFO: SFX menu, no PQDT full text service appears