Knowledge Bases and Related Tools (KBART)

October 2013

Magaly Bascones
Jisc Collections – KB+
Co-Chair KBART Working Group
Metadata Supply Chain
Where the chain breaks: Data

• Wrong data.

• Outdated data.
Where the chain breaks: Workflows

- Need to contact providers again and again.
- Receiving title lists form different sources on different formats and sometimes with contradictory data.
- Invest a lot of time formatting and correcting data problems.
- Update procedures vary by provider.
Knowledge Bases and Related Tools

UKSG and NISO collaborative project
Who is on KBART?

• Working group members (stakeholders):
  • Knowledge base vendors
    ExLibris, Serials Solutions, EBSCO, OCLC
  • Content Providers (Publisher & Aggregators)
    Taylor & Francis, Springer, etc
  • Libraries & Consortia
• Full list -- http://www.niso.org/workrooms/kbart
KBART

- **Phase 1** - Universally accepted standardized publisher metadata, regularly distributed AND available on demand (2010)

- **Phase 2** - Broad adoption, Consortia, More content type coverage (eBooks, conference proceedings), Open Access (2013)
<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 (i.e., ISSN, ISBN, etc.)</td>
</tr>
<tr>
<td>online_identifier</td>
<td>Online-format identifier (i.e., 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>num_first_issue_online</td>
<td>Number of first issue available online</td>
</tr>
<tr>
<td>date_last_issue_online</td>
<td>Date of last issue available online (or blank, if coverage is to present)</td>
</tr>
<tr>
<td>num_last_vol_online</td>
<td>Number of last volume available online (or blank, if coverage is to present)</td>
</tr>
<tr>
<td>num_last_issue_online</td>
<td>Number of last issue available online (or blank, if coverage is to present)</td>
</tr>
<tr>
<td>title_url</td>
<td>Title-level URL</td>
</tr>
<tr>
<td>first_author</td>
<td>First author (for monographs)</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 file’s title)</td>
</tr>
<tr>
<td>publication_title</td>
<td>print_identifier</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Accounting and the Public Interest</td>
<td>1530-9320</td>
</tr>
<tr>
<td>Accounting Horizons 0888-7993</td>
<td>1558-7975</td>
</tr>
<tr>
<td>Acoustics Research Letters Online</td>
<td>1529-7853</td>
</tr>
<tr>
<td>Acoustics Today 1557-0223</td>
<td>1557-0223</td>
</tr>
<tr>
<td>AIP Advances 2158-3226</td>
<td>2011-03</td>
</tr>
<tr>
<td>American Journal of Physics 0002-950X</td>
<td>1933-02</td>
</tr>
<tr>
<td>APL: Organic Electronics and Photonics</td>
<td>1941-420X</td>
</tr>
<tr>
<td>Applied Mechanics Reviews</td>
<td>0003-6900</td>
</tr>
<tr>
<td>Applied Physics Reviews</td>
<td>1931-9401</td>
</tr>
<tr>
<td>ASCE Conference Proceedings</td>
<td>1996</td>
</tr>
<tr>
<td>ASME Conference Proceedings</td>
<td>2002</td>
</tr>
<tr>
<td>Astronomy Education Review</td>
<td>1539-1515</td>
</tr>
<tr>
<td>Behavior Research in Accounting</td>
<td>1050-4753</td>
</tr>
<tr>
<td>Biointerphases 1559-4106</td>
<td>2006-03</td>
</tr>
<tr>
<td>Biointerphases 1934-8630</td>
<td>2007-03</td>
</tr>
<tr>
<td>Biomicrofluidics 1932-1058</td>
<td>2007-01</td>
</tr>
<tr>
<td>Communications Engineer 1479-8352</td>
<td>1741-0452</td>
</tr>
<tr>
<td>Computers in Physics 0894-1866</td>
<td>1997-01</td>
</tr>
<tr>
<td>Computing and Control Engineering</td>
<td>0200-8014</td>
</tr>
</tbody>
</table>
Why KBART has been adopted by publishers?

- It is an accepted data format.
- It exposes content to greater usage.
- It increases quality and customer satisfaction.
The KBART adoption process

1. Expression of Interest
2. Read the requirements
3. Create a sample file
4. Test of sample file
5. Endorsement
KBART Registry: Endorsement framework for content providers

Where are we now?

- Phase I KBART Recommended Practice released Jan 2010.

- Endorsement framework for content providers.

- Phase II started in 2010 and draft out for consultation until 4th October 2013 on NISO website. Currently draft under review.
Phase II

- Open Access
- Consortia Metadata
- Ebooks

As today, the final document hasn’t been published yet
Majority agreed that accurate “standardized” holding lists were very important to members of their consortium.

- Perceived their knowledge base accuracy to be about 80% although they strive for greater than 95% accuracy.

- Almost 70% of respondents indicated that it would be useful to identify Open Access and free content.

- 55% would like to see Open Access and free content grouped together as a collection and 45% would like to see OA and free content interspersed amongst vendor and publisher packages.
Open Access

The decision was made not to differentiate between Free and OA for KBART.

“Needed to strike a balance between noting significant OA content and making the file understandable”
Phase II: OA Recommendations

Free-text **coverage_notes** field suggested to explain subtleties of OA availability for a particular title.

New field - **access_type**

“F” - title is mostly fee-based (subscription/purchase)

“OA” - 50% or more of the title is OA/freely accessible.
Consortia Metadata

- Consortium-specific files should be created when:
  - A unique set of titles has been packaged for the consortium, different than the Content Provider’s standard packages.
  - A package contains unique dates of coverage.
Consortia – New File Name Examples

• Title list is not region or consortium-specific, includes all titles from the content provider:
  – JSTOR_Global_AllTitles_2008-12-01.txt
  – Taylor & Francis_Global_AllTitles_2012-08-30

• Title list is consortium-specific, for a specific package:
  – IOP_NESLi2_Option 1 (2011)_2012-05-31.txt (includes a year as part of the package name)
  – Oxford_SCELC_AllTitles_2012-01-09.txt (contains all titles that the consortium has subscribed to)

• Title list is region-specific, for a specific package:
  – Springer_Asia-Pacific_Medicine_2012-08-03.txt
Ebooks

• Phase I
  • Recommendations were serial-centric. Only one Ebooks specific field: first_author

  • Some fields were dual-purpose
    • date_first_issue_online
    • Identifiers

  • Holding’s content type was ambiguous

• Phase II
  • 5 new monographic fields added
Phase II: New Ebooks fields

- date_monograph_published_print
- date_monograph_published_online
- monograph_volume
- monograph_edition
- first_editor
Book Series / Proceedings - Phase II

• Challenges
  – Both serial and monograph
  – Users search for both titles

• New fields
  – parent_publication_title_id
  – preceding_publication_title_id
Shared Knowledge Base Initiatives

- **Knowledge Base+** ([www.kbplus.ac.uk/kbplus/](http://www.kbplus.ac.uk/kbplus/)) - Run by JISC Collections (UK), KB+ is a shared community service knowledge base for UK academic libraries to support the management of electronic resources.

- **GOKB**, the Global Open Knowledgebase ([gokb.org/](http://gokb.org/)) (USA) - an open, community-based, international data repository that will provide libraries with publication information about electronic resources.
Phase 1 – Universally accepted standardized publisher metadata, regularly distributed AND available on demand

Phase 2 – Broad adoption, Consortia, More content type coverage (eBooks, conference proceedings), Open Access

Phase 3? – Even more content types, automated delivery, institutional metadata????
Thank You!

Magaly Bascones
Co-Chair KBART
Data Manager, Knowledge Base +,
Jisc Collections
m.bascones@jisc-collections.ac.uk