Communities of Knowledge: Creating and Connecting Resource Metadata

Yvette Diven
Senior Product Manager
yvette.diven@serialssolutions.com

22nd Annual North Carolina Serials Conference
March 15, 2013

We help libraries work better.
Collaboration, Community, and Connection. . .

- The idea of a knowledge base
- Motivating factors
- Some projects underway
- Connectivity building blocks
- Opportunities and challenges
- Your observations
Situation analysis

- E-resources = ~70% U.S. academic library collection spend
- U.S. academic library collections\(^1\)
  - 158.7 million e-book holdings
  - 1.8 million e-reference sources and aggregation services
- The number of e-resources continues to grow exponentially, year over year.

The idea of a knowledge base

► Repository of (technical) answers
  ➔ “How can I get database access from off campus?”

► Address book for e-resource look-up
  ➔ At minimum: “Those titles appear in this database from that provider.”

► Metadata + content management system
  ➔ Data store, service platform, APIs, expert systems
What’s a knowledge base for?
What’s a knowledge base for?
Commercial knowledge bases

- For-profit and not-for-profit providers
  - E-resource metadata
  - Library holdings data
  - Software and services
    - e.g., discovery services, ERM, link resolvers
  - Varied levels of
    - Publisher relationship programs
    - Data updating and maintenance
    - Format and system interoperability
    - Data use policies
“Today, no freely available centralized database exists that comprehensively describes the e-holdings associated with library content packages.”

- Marshall Breeding
Knowledge Base and Link Resolver Study: General Findings (May 1, 2012)
Why start a knowledge base project?

Reasons expressed...

- Get information “as soon as it’s available”
- Direct “control of our own data”
- We can “save time and duplicative effort”
- We can “help vendors better track our consortium’s subscriptions”
- “Librarians are the best suited” to do this
- “Everyone can share in this information”
What are the goals of these projects?

- **Start local, go global**
  - Consortium, national, international initiatives

- **Community-driven**
  - Created, maintained, and supported

- **Open**
  - Reusable information, without restrictions
Projects underway: consortium: 2CUL

- Mellon-funded, Columbia and Cornell Universities, plus Ithaka

- Focus:
  - Collective e-resource management
  - Descriptive metadata stored in a single, shared KB by 2015

- Current status:
  - New round of funding, January 2013
  - “Phase 2” focus on combined technical services and additional cost savings
Projects underway: national: JUSTICE

- NII-backed, Japan Alliance of University Library Consortia for E-Resources (JANUL+PULC)
- **Focus:**
  - Consortium license agreements
  - Building a national e-resources collection
  - E-resource management and discovery
- **Current status:**
  - 500 potential participant institutions
  - Early discussions with publishers
  - Reaching out to JISC KB+
JISC-funded, “community-driven”, plus outside tech consultants

- Focus:
  - Subscriptions and renewals (UK NESLi2 packages)
  - Interoperability to improve commercial knowledge bases
  - Self-sustaining subscription basis later in 2013

- Current status:
  - Data model, provider files from 2011 and 2012
  - Partnering with Kuali OLE’s GOKb
Projects underway: international: GOKb

Mellon Foundation-funded, Kuali OLE “community-driven”

Focus:
- E-resource management
- Create trackable title-instance identifiers (TIPP)
- Data sourcing from publisher web sites and in partnership with CUFTS

Current status:
- Targeting e-serials from top providers
- Partnering with KB+ for data model
Building blocks of connectivity
Creating and connecting metadata

**Standards**
- Identifiers (e.g., ISSN, ISBN)
- Data delivery (e.g., ONIX, SUSHI)
- Industry-driven (e.g., KBART, COUNTER)

**Data Models**
- Proprietary relational (e.g., KB+/GOKb)
- FRBR (Work>Expression>Manifestation>Item)
- LoC’s BIBFRAME (Work>Instance, Authority, Annotation)
- Linked Data (RDF, ‘triples’, URIs)

**Open Systems**
- Creative Commons 0 (CC0) public domain license and others
- Open APIs
“The creation and maintenance of an e-content knowledge base requires an ongoing investment of both automated and manual processes to achieve reasonable levels of comprehensiveness, consistency, and correctness. It requires a continual flow of information from publishers and significant effort to enforce consistency, eliminate errors, and to ensure completeness.”

- Marshall Breeding
Knowledge Base and Link Resolver Study: General Findings (May 1, 2012)
Opportunities and challenges
Identifying the community

- U.S. academic library staffing
  - 88,943 full-time equivalent (FTE) staff
  - 26,706 FTE librarians

- A skilled community
  - E-resource librarians, serials specialists, catalogers, acquisitions librarians, bibliographers, selectors, systems librarians and others

---

Challenges of being “community-driven”

**Average Academic Library**

- Maintaining 60,497 holdings

**Time Spent**

- 190 hours per month
- $68,000 per year

Approximately $1 per holding per year*

---

*Serials Solutions calculation based on reported mid-sized academic libraries’ experiences*
## Opportunities and challenges

### Scope of work
- Determining how much data is enough
- ERM-only focus vs. ERM + discovery
- Provider outreach

### Nature of e-resource metadata
- Ambiguous terms (e.g., database, package, aggregation, collection)
- Controlled vocabulary maintenance and management
- Complexity of serials (e.g., title splits)
- Metadata change management

### Quality
- Publisher and aggregator data quality
- The 'good enough' trade-off: accuracy vs. timeliness
- Dealing with de/duplication
# Opportunities and challenges

<table>
<thead>
<tr>
<th>Sustainability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivating and empowering community participation</td>
</tr>
<tr>
<td>Staffing levels vs. current workload</td>
</tr>
<tr>
<td>Cultural change (e.g., trusting other librarians to create and maintain quality data)</td>
</tr>
<tr>
<td>Viability of the free-service business model</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels of compliance (e.g., assigning and using identifiers)</td>
</tr>
<tr>
<td>Role of newly-created, proprietary identifiers and data models</td>
</tr>
<tr>
<td>Selecting and implementing interoperability standards</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rights and governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consortium relationships and packages</td>
</tr>
<tr>
<td>Understanding limits of disclosure and sharing</td>
</tr>
<tr>
<td>Managing 'community-driven' work</td>
</tr>
</tbody>
</table>
"Collective problems require collective action, which requires a shared vision."

– Association of Research Libraries
ISSUE BRIEF: 21st-Century Collections: Calibration of Investment and Collaborative Action

Thanks for joining the discussion!
Monitoring project progress

- JISC KB+: [http://www.jisc-collections.ac.uk/knowledgebaseplus/](http://www.jisc-collections.ac.uk/knowledgebaseplus/)
- Kuali OLE® GOKb: [gokb.org](http://gokb.org)
- Linked Data: [http://www.w3.org/standards/semanticweb/data](http://www.w3.org/standards/semanticweb/data)
- NISO Bibliographic Roadmap Development Project: [http://www.niso.org/topics/tl/BibliographicRoadmap/](http://www.niso.org/topics/tl/BibliographicRoadmap/)