Evaluating and Implementing Web Scale Discovery Services: Part Two

Jason Vaughan  
*University of Nevada, Las Vegas, jason.vaughan@unlv.edu*

Tamera Hanken  
*University of Nevada, Las Vegas, tamera.hanken@unlv.edu*

Follow this and additional works at: [http://digitalscholarship.unlv.edu/libfacpresentation](http://digitalscholarship.unlv.edu/libfacpresentation)

Part of the [Cataloging and Metadata Commons](http://digitalscholarship.unlv.edu/cataloging/), [Databases and Information Systems Commons](http://digitalscholarship.unlv.edu/database/), [Science and Technology Studies Commons](http://digitalscholarship.unlv.edu/technology/), and the [Strategic Management Policy Commons](http://digitalscholarship.unlv.edu/policy/)

Repository Citation


Available at: [http://digitalscholarship.unlv.edu/libfacpresentation/89](http://digitalscholarship.unlv.edu/libfacpresentation/89)
Evaluating and Implementing Web Scale Discovery Services in Your Library

ALA TechSource Workshops, July 13 & 20, 2011

Jason Vaughan
Director, Library Technologies
UNLV Libraries

Tamera Hanken
Director, Technical Services
UNLV Libraries
General Outline (July 20)

Part 4: Quick Tour of the Current Marketplace
- “The Big 5”
- Similarities and differences

Part 5: It’s Not All Sliced Bread
- Shortcomings of web scale discovery
General Outline (July 20)

- Part 6: Implementation (pre launch steps)
  - Selecting and preparing implementation staff
  - Preparing and communicating process/decisions with all staff
  - Working with the vendor (roles, expectations, timeline)
  - Workflow changes and implications (technical services)

- A Brief Pause for Initial Q&A
General Outline (July 20)

- Part 7: Specific implementation tasks, issues, considerations
  - Record loading and mapping (catalog content)
  - Harvesting and mapping digital/local content
  - Working with central index data (internal & external content)
  - Web integration and customization
  - Assessment and continuous improvement

- Session Wrapup: Q & A
Part 4

A Quick Tour of the Marketplace

(Similarities and differences between some of the services)
Similarities and Differences

- Content (scope/volume, level of metadata/indexing)
- User Look & Feel (and functionality)
- Level of customization the library can do to “make it their own” (branding, etc.)
- Other goodies
- Pricing models
Similarities and Differences

In GENERAL, there are more similarities then differences, but the devil is in the details. Kind of like integrated library systems and the front end web catalogs those systems offer – platforms are a lot alike.
Similarities (Content Scope)

- Vendor publisher agreements
  - Existing
  - On the horizon

- All vendors already have a huge amount of indexed content (hundreds of millions of indexed items . . . at least two vendors indicate they have already surpassed a half billion indexed items).
Similarities and Differences: (Content - Metadata & Indexing)

- What’s being indexed?
- Level / amount / source of metadata
- Vendors each have an opinion on the strength of their metadata and their competitor’s metadata. You’ll have to talk to them.
Similarities (User Look & Feel / Functionality)

- Each platform offers a modern interface with design elements expected by today’s students.
  - A single search box (but with a link to advanced search modes)
  - Faceted navigation (subject, content type, publication date range, etc.) to help users drill down a large set of results
  - Inclusion of enriched content such as book cover images
  - Shopping carts to easily mark items and later export the materials (email, print, save)
Similarities (User Look & Feel / Functionality)

- Realtime status calls to the underlying ILS to provide call number, location, and status information for library hardcopy materials

- “Did you mean?” spell checkers

- User configurable RSS feeds to easily re-run searches later
Differences (User Look & Feel / Functionality)

- Some (not all) systems offer user accounts with certain abilities

- Some systems offer tighter integration to the full text.

- Some systems may offer more full text content as “native” PDFs, which are of higher quality and searchable (as opposed to scanned image pdfs).
Differences (User Look & Feel / Functionality)

- Some systems “take you out of the discovery interface” for the full record or full text.
  - Remotely hosted content
  - Locally hosted content
  - Some of this integration may depend on whether you have an ILS from the same vendor; or if the journal content you’re looking at is sourced from the discovery platform vendor
Differences (User Look & Feel / Functionality)

- Variations -- Advanced search capabilities

- Variations -- Faceted navigation

- Some (not all) products offer Web 2.0 social community features
Similarities and Differences: Level of Local Library Customization

- All products allow for some level of look and feel interface, but it varies from product to product
  - One extreme: Library Logo, Colors
  - Other extreme: Create Your Own Interface

- Algorithm Tweaking
Similarities and Differences: Other Goodies

- Widgets
- Mobile Interface
- Recommender Services
Similarities and Differences: Pricing Models

Pricing models can vary among vendors.

- Subscription Model

- Hosting Options
  - Central Index
  - Application / User Interface
Similarities and Differences: Pricing Models

- What Determines the Pricing?
  - The amount (item count) of local library content harvested
  - Whether you host the application or they host it
  - University FTE count and/or degree granting status.
  - “Size of your user community”
Similarities and Differences: Pricing Models

Other factors:

- Additional services you may choose, a la carte
  - Vendor brokered content enrichment services
  - Article recommender services
  - Optional federated search components offered by the vendor
  - Consulting / development of custom ingestors to harvest unique, non mainstream local library databases

- Multi-year and consortial discounts are often available
Part 5

It’s Not All Sliced Bread
(real and/or perceived shortcomings of web scale discovery)
Web Scale Discovery: Benefits

- It’s very fast. Google fast. You are searching a single index, and not lots of individual database indexes, your own catalog, your digital collections, etc.

- A single central index lends itself to data normalization and relevancy ranking.

- It offers a streamlined interface, some features of which are really made possible due to the preindexed nature of these services.

- Can aid interdisciplinary research, by putting lots of content from multiple disciplines into one index, one search interface.

- Can be seen as generally aligning with information literacy efforts.

- Can help foster a more user self-sufficient environment.
Things to Be Aware Of

- Does not cover 100% of your resources
- Potential role of federated search
- Can lose the unique interface / functionality of specialized subject databases
Some Other Concerns You May Hear

- Occasional Display Issues
- Broken / Dead End Links
- Known Item Searches may be tricky
Some Other Concerns You May Hear

- Google-Think: Students may not think (or even be aware of) other databases, with other, additional content.

- Interface is not perfect

- Won’t deliver full text 100% of the time, and students may want (expect) this

Some Other Concerns You May Hear

- How will we pay for this?
- Tons of results can be returned
  - Separating wheat from chaff
  - Post-search refinement vs. pre-search refinement
- Algorithms are not necessarily understood and are proprietary
- What’s included in the index?
Some Other Concerns You May Hear

- **Content Neutrality**
  
  - In the sense that the discovery vendor is owned by a parent company whose business is content . . . Is that parent’s company content promoted or weighted more heavily in search results?

  - In the sense that some vendors may be inking exclusive agreements with publishers whereby only that vendor’s discovery tool can index that publisher’s content
Some Things to Be Aware Of

Web scale discovery systems are NOT the last evolutionary step for information discovery related to libraries

• They still can’t “read” the user’s mind and know precisely what it is the user is searching for.

• Future systems will likely take the search features and functionality even further
Positive Forward Thinking

- Marketplace adoption rate
- Library usability studies & student adoption
- Ever increasing amount of indexed content
- Early research appears to indicate an increase in the usage of the library’s licensed e-content (full text downloads)
Part 6

Implementation
(pre launch steps)
Implementation (pre launch steps)

- Implementation (pre launch steps)
  - Selecting and preparing implementation staff
  - Preparing and communicating process/decisions with all staff
  - Working with the vendor (roles, expectations, timeline)
  - Workflow changes and implications (technical services)
Implementation: a change management issue

- Comprehensive evaluation involving wide range of stakeholders (2009-2011)
  - As Jason described in last week’s session
  - Full disclosure and opportunity to participate in decision
  - We all recognized the Web Scale Discovery platform represented a fundamental shift in the discovery and delivery of content
    - Before--multiple silos of content
    - After--a single interface

Easily discover the world of library content
Implementation: A change-management issue

Culture Shock: Librarians’ Response to Web Scale Search by David Howard & Constance Wiebrands @ Edith Cowan University

“The biggest risk we faced was negative librarian experiences and views flowing through to our users”

"...these concerns (regarding the Summon single search box) reflected librarians’ desire to find information in the “right way”. For our users the “right way” is the quick, simple and effective way - something that does not always sit comfortably with librarians as a profession.“

This Conference Proceeding is posted at Research Online. http://ro.ecu.edu.au/ecuworks/6206
... a change management issue

“I was responsible for helping libraries make the transition from their old systems to Innovative’s. However, as these librarians implemented their new automation systems, they often adhered to their established work-flow processes based on staffing and their old systems. Their attachment to existing methodologies was not out of nostalgia, but out of practice and habit...”

Seeman, C. (2002) ... Computers in Libraries
... a change management issue

- For technical services staff there were also implications in going from maintaining multiple knowledge databases-- to one more?

**Research Tools**
- Library Catalog EJ
- Research Gateway EJ (articles & databases)
- Journal List EJ (electronic & paper journals)

- Questions to the vendors at the beginning of the evaluation process:
  - Based on your experience with existing customers, have you found or are you aware that they have had to change some technical services staff ongoing patterns of work?
  - What new duties have been started?
  - What duties have been modified or dropped?
  - Have customers reported excessive work related to record “clean up” activities?
Our strategy to manage change

• Parallel to Evaluation were changes in Technical Services (2009-2011)
  • Implementing continuous improvement
    • Focus on value from the customers perspective
    • Process reviews, cross training

• Technical Services staff, ready for change (2011)
  • Increased capacity to absorb and modify new workflows with the same amount of resources
Web Scale Discovery: Technology

- **Scalable Index**
  - Content from various sources is normalized into a common schema or record type.
  - To some degree, content is deduplicated
  - Automated transfer routines, load tables, and indexing steps are in place to add newly published content and to keep the index up to date.
  - The index is hosted (and backed up) in a cloud environment.
  - Relevancy algorithms have been developed and tweaked
Ready or Not: The Charge

- Project managers appointed to implement
  - Kee Choi and myself
  - A distributed team

To designate and coordinate the working groups according to the specific actions and functions necessary to implement a fully functioning discovery service/platform. The co-managers will share responsibility in prioritizing tasks, managing workflow, working across and communicating with the many library departments and divisions involved in implementing the discovery service/platform.
The Charge

Working Groups (according to function) and Key Contact

MARC/catalog export and mapping
  Cataloger and Systems staff
Electronic resources management
  Electronic Resources staff (A-Z and SFX)
Digital (non-MARC) mapping and integration (CONTENTdm)
  Digital Collections staff
Institutional Repository (BePress)
  IR staff
Hidden collections integration
  Special Collections Staff
Campus Web services integration
  Campus Services/Distance Learning liaison
Liaison/Instruction services
  Instruction and Liaison
Assessment
  Assessment, Usability, Continuous Improvement
Working with the Vendor

- Initial conference call with vendor project manager--introduction of team members, roles and expectations, timeline
- Checklist and Implementation documents provided
  - Location codes, mapping templates, record upload instructions, linking and authentication forms, etc.
- Training (Admin module) scheduled and provided
- Access to client listservs, Wiki sites, documentation
- Ingestion of MARC and harvesting of content
Working with technical services staff/implementation team

- The message... reminding ourselves of the principles of continuous improvement... reminding ourselves of ongoing process reviews... considerations of workflow implications

- The goal would be to improve services with existing resources (at minimum)
  - Management of knowledge databases (A-Z/SFX)
  - Cataloging options
Questions?
Part 7

Specific implementation tasks, issues, considerations
Specific Implementation Tasks, Issues, Considerations

- Specific implementation tasks, issues, considerations
  - Record loading and mapping (catalog content)
  - Harvesting and mapping digital/local content
  - Working with central index data (internal & external content)
  - Web integration and customization
  - Assessment and continuous improvement
From last week: Considerations of moving from JIC to JIT?

- **Just in Case vs. Just in Time**
- **Within the context of Web Scale Discovery?**
  - **Just in Time**
    - The workflow model strongly supports especially in the realm of electronic resources.
    - Access to a collection (e.g., an ebook collection) can be immediate not need to wait to batch load the MARC records and deal with the related workflows (authority control)
  - **PDA**
  - **Just in Case**
    - The Hathi Trust public domain provides ‘new opportunities’ for our Just in Case collections...
    - There is the potential to provide full text search to lesser used/known collections ‘hidden’ in the stacks--when there is overlap
Working with staff

- Record loading
- MARC metadata mapping and indexing
- Cataloging practices
- Electronic resource management
- Harvesting content and mapping
- Website interface, architecture customization
- Assessment and continuous improvement
Working with staff: record loading

- Our Systems Librarian is responsible for exporting the bibliographic records from our ILS for inclusion in the central index.
  - Initial batch export and import followed by periodic (quarterly?) refreshes
  - Two scripts run daily:
    - The first script handles the routine ILS updates/additions, changes—records still viewable via the central index
    - The second script handles bibliographic records marked for deletion, suppression or withdrawal – records no longer viewable
Record loading: initial challenges

- Setting up the initial/complete export of bibliographic records took some time. With 2 million+ records the Systems Librarian would need to divide the file into 3 separate files to accommodate export and import.
- The scripts were also a challenge that required the assistance of IT staff [Expect scripts with cron jobs]
- Once the process for export/import was set up and the scripts were in place there is no additional work load.
Record loading: considerations

- A technical services workflow that is affected by the daily scripts is the catalogers’ routine of deleting, suppressing, or withdrawing records from the ILS
  - While the ILS is dynamically indexed, the discovery platform (ours) is not
    - Availability of items in the catalog are reflected in the discovery service
    - But, if a bib record has been deleted in the catalog, the discovery displays the message, “No such record exists”
Record loading: a new routine

- Catalogers mark records for deletion, suppression, or withdrawal throughout the day, as needed (coding the record)
- The following morning [after ILS routines/updates] the scripts run to update holdings in the central index
- After holdings are updated in the central index, permanent changes can be made to the ILS
- A minor change to the routine but something to be aware of if removing large batches of bib records (i.e., an ebook collection) Timing is everything!
Working with staff: Public side

- Discovery vs. Known item searching
  - The Discovery platform is great for discovery
  - But many are surprised at how difficult it can be to find known items

- Catalog vs. Discovery Service
  - The Catalog is great for known item searching because of the metadata flexibility
  - Within many of the discovery platforms the metadata is simplified to accommodate the central index
Working with staff: mapping

- Our Senior Cataloger accepted responsibility for mapping the bibliographic data from the ILS database structure to the central index database structure
  - Similar to the process of migrating MARC data from one ILS platform to another—in many ways, much ‘simpler’
  - The design of the central index is less complex than the ILS in order to accommodate various sources of MARC, Dublin Core, etc... the normalization Jason referred to in last weeks presentation
# MARC Mapping example

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Display</th>
<th>Display Facet Loc</th>
<th>Default MARC</th>
<th>Changes/Corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Author</td>
<td>Display/Brief &amp; Preview</td>
<td>100 !a</td>
<td>b</td>
</tr>
<tr>
<td>Corporate Author</td>
<td>Corporate Author</td>
<td>Display/ Brief &amp; Preview</td>
<td>110</td>
<td>a</td>
</tr>
<tr>
<td>Title : Subtitle</td>
<td>Maps to title field</td>
<td>Display--Preview</td>
<td>245</td>
<td>a</td>
</tr>
<tr>
<td>Genre</td>
<td>Genre</td>
<td>Display--Preview</td>
<td>600</td>
<td>v; 610</td>
</tr>
</tbody>
</table>
Mapping: consideration

With the Beatles

Author: Lennon, John, 1940-1980; McCartney, Paul; Harrison, George, 1943-2001; Starr, Martin, George, 1926 Corporate Author: Beatles
Call Num: MUS CD 10973 Date: 2009
Subjects: Rock music
Mapping: considerations

- Like the record loading process most of the work is during the implementation
  - The cataloger continues to make mapping adjustments
  - Currently (in our case) only the vendor can make mapping changes, in the near future the client will have control
  - Currently (again in our case) the mapping allows for a one to one relationship, e.g., a government document is a government document (not also an electronic resource)
    - Part of what made mapping difficult--choose one!
    - This is a near future enhancement
Mapping: considerations

- The mapping template can be an indicator of:
  - A view of sorts to the database structure
  - Which fields will display and how
  - Content labels
- The template may indicate what can and cannot be changed (indexing, labels, display)
- To peek or not to peek (whether to review the mapping document during product evaluation)
Catalog: implications of past practices

- Loading ebook vendor records into the catalog when also available, and harvested, from the knowledge base of the central index
  - Quality of vendor records (brief metadata or coding errors)
  - Duplicate records
- Practice of the one record approach for books and journals that have an electronic counterpart
- Obsolete or wrong MARC coding
Cataloging staff: new or changes to routines

- An additional knowledge base was added to the mix (the catalog, the A-Z list, SFX, Discovery Platform)
- Beyond the initial mapping projects there are minor workflow adjustments (dealing with deletions)
- Decisions to be made on whether to continue to add records to the catalog instead of or in addition to reliance on the knowledge database
  - Duplicate records
- Ongoing cleanup (would happen anyway)
Working with staff: Electronic Resources

- Familiarity with the A-Z knowledge database
- Opportunity to change link resolver products and ‘consolidate’ knowledge databases
  - Less staff maintenance
- Opportunity to re-educate staff about the role of the E-resources staff—and the workings of the knowledge database
Technical Services workflow changes/implications

- Management of knowledge databases (A-Z – SFX)
  - Before and after (gaining one FTE)
  - Cataloging opportunities/consequences
  - Systems librarian (only real increase in workload)
Working with staff: Digital content

- The most challenging aspect of the evaluation
  - Platforms primarily MARC based products
  - Platforms were new (as Jason discussed)
  - Integration of Dublin Core or other schemas was new
- We were specifically interested in CONTENTdm, BePress, and local databases with very limited metadata
- Lack of documentation and no real examples to preview
  - What would a CONTENTdm collection look like in the search results?
Map of Tonopah Mining District, Nevada
Mapping harvested digital content

- Record harvesting is handled by the vendor... it’s a matter of providing the OAI url for services to harvest content.
- Software, like CONTENTdm or BePress have the ability to suppress content and not allow harvest of areas/directories... a pleasant surprise
Mapping harvested digital content

<table>
<thead>
<tr>
<th>CONTENT Item Local Field Names</th>
<th>Current QDC Mapping</th>
<th>Desired Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of Birth</td>
<td>Date</td>
<td>None</td>
</tr>
<tr>
<td>Place of Birth</td>
<td>Coverage</td>
<td>Coverage-Spatial</td>
</tr>
<tr>
<td>Affiliation</td>
<td>Relation</td>
<td>Subject or None</td>
</tr>
<tr>
<td>Length</td>
<td>Coverage-Temporal</td>
<td>Format-Extent</td>
</tr>
<tr>
<td>Format</td>
<td>Format-Medium</td>
<td>Format</td>
</tr>
<tr>
<td>Digitization Specifications</td>
<td>Format</td>
<td>None</td>
</tr>
<tr>
<td>Photo Credit</td>
<td>Publisher</td>
<td>Contributor</td>
</tr>
<tr>
<td>Digital ID</td>
<td>Identifier</td>
<td>None</td>
</tr>
<tr>
<td>Original Item ID</td>
<td>Identifier</td>
<td>None</td>
</tr>
</tbody>
</table>
Harvested digital content

- It was a shock to see the consequences of the object removed from the collection
- Staff are considering how the Dublin Core metadata can be enhanced to build back in the context...
- Would have been nice to get a preview of what search results would look like
- The metadata exposure reveals issues
Mapping harvested digital content

- Advice is to not do anything specific with the mapping/metadata
- Conform to standards because collections will be added to various systems (GWLA, WC, Catalog, ContentDM.... Don’t customize for a specific system
- The digital object won’t look good outside of it’s collection...
- Level of partnership with vendor is important... some will customize radically, others cannot...
- WC provides a self service model for mapping and indexing-- a consideration when evaluation... in WC you can see results immediately...
Auditing Digital Collections Metadata

UALC Digitization Committee Webinar
Joint presentation by:
Best Practices and Standards Task Force
Training Task Force
12/16/2010
Working with staff: Web interface

- Website interface, architecture customization
  - Out of the box vs. full customization ability
- Jason?
Continuous improvement

• The Assessment, Usability and Continuous Improvement Team
  – Head of collections
  – Head of Assessment
  – E-resources Librarian (driver of Admin module)
  – Systems librarian
    • Ongoing contact w/ vendor
    • Making informed changes according to a schedule and user feedback
    • Informing marketing of product AND collections
Do we have any regrets?

- No!
- It is better than it was before--discovery is now possible
- The catalog is still available for known item searching
- We’ll continue to learn
- Discovery platforms will continue to evolve
Evaluating and Implementing Web Scale Discovery Services in Your Library

ALA TechSource Workshops, July 13 & 20, 2011

Jason Vaughan
Director, Library Technologies
UNLV Libraries
jason.vaughan@unlv.edu

Tamera Hanken
Director, Technical Services
UNLV Libraries
tamera.hanken@unlv.edu