7-13-2011

Evaluating and Implementing Web Scale Discovery Services: Part One

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

Part of the Cataloging and Metadata Commons, Databases and Information Systems Commons, Science and Technology Studies Commons, and the Strategic Management Policy Commons

Repository Citation

Available at: http://digitalscholarship.unlv.edu/libfacpresentation/90

This Presentation is brought to you for free and open access by the Library Faculty/Staff Scholarship & Research at Digital Scholarship@UNLV. It has been accepted for inclusion in Presentations (Libraries) by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
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 13)

- Preface: Before Web Scale Discovery
  - A very brief overview
- Part 1: What is Web Scale Discovery
  - Content
  - Technology
- Part 2: Why is Web Scale Discovery important?
  - What’s the need?
  - How is it different from earlier attempts at broad discovery?
- A Brief Pause for Initial Q&A
General Outline (July 13)

- Part 3: A Framework for Evaluating Web Scale Discovery Services
  - What we did at UNLV
  - Other options
- 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
- Session Wrapup: Q & A

Note: If you have additional questions about the material presented in the first session, please send questions to jason.vaughan@unlv.edu, and we may have time to address them during the second session.
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

Note: If you would like to send specific questions about topics to be covered during the second session, please send questions to tamera.hanken@unlv.edu before July 19.
Preface

Before Web Scale Discovery . . .
A Quick History ( +/- 15 Years)

The Web Based Online Catalog
A Quick History (cont) . . .

Federated Search

ExLibris MetaLib

WebFeat

Research Pro
A Quick History (cont) . . .

“Next Generation”
Discovery Platforms
And Today . . .

“Web Scale” Discovery Services
Part 1

Web Scale Discovery, What is it?
What are “Web-Scale” Library Discovery Tools?

- “A combination of content (through provider agreements) & technology that greatly facilitates the discovery and delivery of a tremendous amount of purchased, licensed, and free information.”

- “[A service] that is delivered on demand to library users via the browser, with infrastructure, processing and indexing provided and maintained remotely by the vendor.”
Web Scale Discovery: What’s Included?
Remotely Hosted Content
(Purchased/Licensed)

• Journal / magazine Articles
• Newspaper articles
• Conference proceedings
• Abstracts and indexes
• Increasingly, e-books
What’s Included: Open Access Content

- HathiTrust Digital Library -- 2.4 million digitized books from their Public Domain eBook Collection; digitized public domain journal issues, etc.
- DOAJ - Directory of Open Access Journals (over 3000 journals searchable at article level)
- Hindawi Publishing Corp. (200+ open access academic journals)
- arXiv e-prints (Cornell Univ hosted, over 685,000 items mostly in the sciences)
- Some can incorporate and expose the bibliographic and digital collections of other libraries – the collections that you yourselves have and have given permission to be harvested and discoverable.
Web Scale Discovery: Content

• In many cases, the publishers are providing the full text content for indexing purposes
• Agreements may be brokered whereby the publishers provide fielded metadata (title, author, publication info, etc) to the discovery service vendors
• Vendors can develop multiple content streams for the same, finite content. For any given article, there are lots of potential sources for that exact same article, not just the original primary publisher . . .
Web Scale Discovery: Content

- Important to understand: these systems are not providing a library “free access” to licensed, full text content.

- Access to “the final full result” – the licensed, full text content -- is still dependent on the publisher / aggregator content licenses the local library purchases / maintains. In some cases, the “final result” may be a citation/abstract information, such as is found in A&I indexes

- Still, you may have access to some citation level content which you otherwise wouldn’t have access to (and haven’t licensed) – even this is helpful for discovery.
Web Scale Discovery: Content

- Authentication Requirements

- These systems work with existing, common library tools to broker the access to the full text content
  - Link resolvers
  - Proxy servers
  - Other rights management knowledge databases associated with the discovery vendor
Web-Scale Discovery: What’s Included? Local Content

- Bibliographic records from your integrated library system (doesn’t matter which ILS you use).

- (Dublin core, etc.) metadata associated with your digital collections

- Content from other hosted repositories, such as institutional repositories and Libguide subject guides
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
Web Scale Discovery: Technology

- **Interface**
  - Vendors have each developed (and tweaked) end user interfaces to search the index and return results.
  - **Interface often includes**
    - A single search box
    - Faceted searching
    - Evaluative content (book covers, reviews, etc.)
    - Social networking tools, etc.
Interface is often hosted by the vendor, but some systems allow for local hosting of the interface (the content index is always remotely hosted in the cloud).

Discovery Services are quite “open” compared to old-school ILS platforms – with flexible APIs and customization capabilities allowing you to hack, repurpose, or customize the interface.
Part 2

Why is Web Scale Discovery Important?
Why Web Scale Discovery for Library Resources?

If new web scale discovery services are a solution, what’s the problem?

Three perspectives:
- The User
- The Library
- The Publishers
The User Perspective

The Principle of Least Effort . . .

“People do not just use information that is easy to find; they even use information that they know to be of poor quality and less reliable—so long as it requires little effort to find—rather than using information they know to be of high quality and reliable, though harder to find.”

The User Perspective

“End users’ expectations of data quality arise largely from their experiences of how information is organized on popular Web sites . . . (user) expectations are increasingly driven by their experiences with search engines like Google and online bookstores like Amazon. When end users conduct a search in a library catalog, they expect their searches to find materials on exactly what they are looking for; they want relevant results.”

Q: “If you could provide one piece of advice to your library, what would it be?”

A: “Just remember that students are less informed about the resources of the library than ever before because they are competing heavily with the Internet.”

The User Perspective

“The continuing proliferation of formats, tools, services, and technologies has upended how we arrange, retrieve, and present our holdings. Our users expect simplicity and immediate reward and Amazon, Google, and iTunes are the standards against which we are judged. Our current systems pale beside them.”

- The University of California Libraries. Rethinking How We Provide Bibliographic Services for the University of California: Final Report, 2005
The User Perspective

“Users don’t understand the difference in scope between the catalog and A&I services (or the catalog, databases, digitized collections, and free scholarly content).”

The Library Perspective

“Basic scholarly information use practices have shifted rapidly in recent years, and as a result the academic library is increasingly being disintermediated from the discovery process, risking irrelevance in one of its core functional areas (that of the library serving as a starting point or gateway for locating research information)”

“Today, there are numerous alternative avenues for discovery, and libraries are challenged to determine what role they should appropriately play. We have seen faculty members steadily shifting towards reliance on network-level electronic resources, and a corresponding decline in interest in using locally provided tools for discovery.”

“It is our responsibility to assist our users in finding what they need without demanding that they acquire specialized knowledge or select among an array of ‘silos’ systems whose distinctions seem arbitrary.”

- The University of California Libraries.  
Rethinking How We Provide Bibliographic Services for the University of California: Final Report, 2005
“By making metadata or full text available to the discovery service, a publisher of electronic content gains better exposure of their content, while retaining control over the display or delivery of that content (i.e. hit their server at the end)”

The Publisher Perspective

“It’s mutually advantageous to both publishers and discovery providers to cooperate, since it both increases the effectiveness of the discovery products and improves the value of the content for libraries as it makes that content more easily available to their users.”

The Publisher Perspective

- Libraries are interested in return on investment, and showing their value to their faculty colleagues and their provost / president

- Library budgets are tight

- Librarians like to look at usage statistics and conduct content overlap analyses to help determine what may go on the chopping block

- It’s in the publishers best interest to have their content exposed. Exposure can lead to usage and downloads.
But to emphasize, of all the perspectives, the user perspective should not be underestimated . . .
In Short, Before . . .
Before . . .
Before ...
And After . . . (ONE search box)
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
Current Platforms “The Big 5”

- OCLC WorldCat Local (late 2007)
- Serials Solutions Summon (mid 2009)
- EBSCO Discovery Service (early 2010)
- Innovative Interfaces Encore Synergy (mid 2010)
- Ex Libris Primo Central (mid 2010)
Initial Questions?
Part 3

A Framework for Evaluation
“As history has shown, multiple solutions arise to address real needs, and each solution has its own characteristics. In terms of discovery solutions, I'm confident that each library, after conducting a thorough evaluation of facts and features, will be able to determine which of the available products best fits the library's mission, needs, policies, and environment.”

Evaluation

Acknowledge that

- A new discovery service could be the primary entrée for a majority of your users – certainly your undergraduates – to both local library materials and your huge portfolio of licensed e-content.

- Whichever service you choose, while, not permanent, may be a relationship you have for several years if not longer.
Evaluation

Unless you live under a repressive and controlling dean/director, or have a very rapid timeline in which you need to spend a lot of money, you may want to:

- Research, in detail, the (changing) marketplace
- Be inclusive, and communicative, with your fellow library staff (and perhaps beyond)
- While your institution may be unique, it may not be as unique as you think, so don’t recreate the wheel
- Don’t rush to a selection, yet don’t get caught in indecision, which, of course, is a decision
Evaluation Models

Published research can help. Examples include:

- Oregon State University “Discovery Services Task Force Recommendation to University Librarian”
- University of Arizona “Implementing Web-Scale Discovery in an Academic Library”
- University of Michigan “Article Discovery Investigation”
- University of Minnesota “Discovery Phase 1 and 2 Reports”
- University of Nevada, Las Vegas “Investigations Into Library Web Scale Discovery Services”
Evaluation Models

- **CONTENT** (scope and depth, richness, update frequency, ease of incorporating local content)

- **SEARCH** (interface simplicity, quality of results, ability to customize relevancy, etc.)

- **FIT** (ease of implementation, compatibility with existing software/content environment, overall customer support, etc.)

- **COST** (as a new service to existing tools, instead of other finding tools, as justified vis a vis the libraries’ goals/objectives)

Sample Evaluation Model: UNLV

• 15 month evaluation period

• Internal Work (library staff education, surveys, etc.)

• External Work (questions to vendors, early adopter references, etc.)
UNLV Eval Process: Timeline

- Established a “Discovery Task Force” (Sept 2009)
- Background Research; Test Drives of Existing / Demo Implementations (late 2009 – mid 2010)
- Question list to Vendors (Fall 2009)
- Task Force Presentations to Library Staff (Fall 2009, Spring 2010, Fall 2010)
- Library Staff Surveys (April 2010; June 2010)
- Content Overlap Analysis (May 2010+)
UNLV Eval Process: Timeline

- Vendor Onsite Demos to all library staff (Spring - Summer 2010)
- Follow-up question lists, conference calls to vendors (Summer 2010)
- Detailed Q&A and conference calls with early adopters of these services (Fall 2010)
- Final Administrative Discussion & Recommendation (Winter 2010)
- Purchase (end of 2010)
Initial Background Research

- Literature Review

- We provided a context to the vendors of our particular environment by describing our various locally hosted and remote licensed content

- We developed and organized a list of 70+ questions which we sent to all five vendors (whether they had a released product, or just an announced product)
Organized questions into nine broad areas, some technical, some non-technical

Vendors were asked to respond in 2-3 weeks, and they all did
Vendor Question Categories

- Section 1 Background
- Section 2 Locally Hosted Systems & Associated Metadata
- Section 3 Publisher/Aggregator Coverage (Full Text and Citation Content)
- Section 4 Records Maintenance and Rights Management
Background Research

- Section 5  Seamlessness & Interoperability with Existing Content Repositories
- Section 6  Usability Philosophy
- Section 7  Local “Look and Feel” Customization Options Controllable by the Library
- Section 8  User Experience (Presentation, Search Functionality and What the User Can Do With Results)
- Section 9  Administration Module & Statistics
Staff Education

- Created staff website; library wide internal presentations

- First Presentation
  - Education

- Second Presentation:
  - Update on Work
  - Live Test Drives of Existing Implementations

- Third Presentation:
  - Overall findings and recommendation
Library Staff Surveys (2)

- Used the Survey Monkey Tool
- Surveys questions were a mix of “rank on a scale,” multiple choice, and free text response questions
- Respondents could skip any question they wished
- Higher response rate from first survey; lower response rate for second survey
Staff Survey 1

• First Survey: Conducted BEFORE Vendor Visits. We asked questions in three functional areas.
  - Local Library Customization Capabilities.
    - “Is it important for the library to be able to control/tweak/influence the following design elements . . .” (Strongly Agree < - > Strongly Disagree)
  - End-User Aspect: Features & Functionality
    - “The following functionality is important to have in the discovery service . . .” (Strongly Agree < - > Strongly Disagree)
  - Content
    - “Please rank on a 1-10 scale how vital it is that a discovery service accommodate records from these information repositories”
Staff Survey 2

- Conducted AFTER All 5 Vendor Visits/Demos
- Had questions along the same functional areas as first survey (local library customization features; end user features/functionality; content)
- For each question, respondents were asked to respond to the question for each of the five products.
  - e.g. “The Discovery Platform appears to ADEQUATELY cover a MAJORITY of the CRITICAL publisher titles (strongly agree, agree, disagree, etc.)”
More Work . . .

- Collection Overlap Analysis
- Consolidating Vendor Responses
- Vendor Onsite Visits
- More Questions for Vendors
- Reference Checks with Early Adopters
- More Research: Keeping Ahead of the Curve
- Vendor Quotes
- Final Recommendation to Library Administration
Early Adopter Questions

- **Background Questions**
  - “How long was the implementation period? Is it now your default search tool?”

- **Content Questions**
  - “Have you observed any particular strengths in terms of subject content in any of the three major overarching areas -- humanities, social sciences, sciences? Have you observed that the discovery service leans toward one or a few particular content types?”

- **Interface Satisfaction Questions**
  - Do you have any sense of how satisfied your (non-faculty) end-users are with the discovery service’s interface? Is there any particular feature or function that is missing or non-configurable within the discovery service that you wish were available?
Evaluation: Other Potential Avenues to Explore

- Usability testing of discovery services w/ your students
  - Using other sites’ implemented platform
  - Using a vendor test site
  - Using a custom test site the vendor may set up for you, which may include your own content / subscriptions
- Surveys of students and/or faculty
- Involvement of faculty senate, provost
- Discussions with potential consortial partners
- Request for information / bidding
Evaluation: Other Potential Avenues to Explore

- Creating a detailed matrix comparing the various discovery services, to the best of your understanding
  - Must have features / capabilities
  - Features nice to have
  - Can include topics such as
    - Interface design (real time status calls for ILS items; faceted navigation; advanced search; etc.)
    - Content inclusion, local and subscription
    - Customization capabilities, APIs, etc.
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
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)
Session Wrap Up

Part 2
“Evaluating and Implementing Web Scale Discovery Services in Your Library”

July 20, 2011, 2:30 PM Eastern / 11:30 AM Pacific

Questions . . .
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