Request for Information – Web-Scale Discovery Tool

September 3, 2013

Introduction

On behalf of Harvard University, the Harvard Library is seeking responses for the attached Web-Scale Discovery Tool Request for Information (RFI). Information returned in RFI Responses, including pricing information for various solutions, will inform recommendations and budget planning. While this RFI is non-binding, Harvard Library is hopeful that the information received in the responses will assist the organization in deciding how best to proceed.

In addition to providing RFI Responses, Harvard invites vendors to participate in vendor product demonstrations and Q&A meetings, to be scheduled in November, 2013.

Questions related to the RFI must be sent to the Harvard contact and received by September 25, 2013 at 3:00 pm. Answers to questions will be made available by October 9, 2013.

RFI responses must be sent to the Harvard contact and received by October 30, 2013 at 5:00 pm.

Response Requirements

Please submit one original copy of the response on 8½x11 text weight paper, double-sided. Please provide an additional copy of all materials electronically, either through deposit on a Harvard University secured ftp site (details upon request) or on physical media (CD/DVD or USB flash drive).

Responses should contain the following information:

- Company legal/registered name.
- Company mailing and physical address.
- Company website URL (if available).
- Name of company representative who will be the primary point of contact for inquiries, and contact information (telephone number, e-mail address).
- RFI Response
- Any brochures, material to support RFI Response

Questions and Responses for this RFI may be directed in writing to the Harvard contact:

Laura Morse
Director, Library Systems
Library Technology Services
Harvard University
90 Mount Auburn Street
Cambridge, MA 02138
617-495-3724
A. Harvard University Background

Founded in 1636, Harvard University is preeminent among the world’s research universities, with an FTE of 7,000 undergraduate and 13,000 graduate students. Over 16,000 people work at Harvard, including approximately 2,100 faculty members. An additional 10,000 individuals have faculty appointments in Harvard-affiliated teaching hospitals.

Harvard includes 12 colleges and schools. The Faculty of Arts and Sciences—comprising Harvard College, the School of Engineering and Applied Science, the Graduate School of Arts and Sciences, and the Division of Continuing Education—is the oldest and largest of the Harvard faculties. The University’s world-renowned graduate and professional schools include Harvard Business School, Harvard Divinity School, Harvard Graduate School of Design, Harvard Graduate School of Education, Harvard Kennedy School, Harvard Law School, Harvard Medical School (including the School of Dental Medicine), Harvard School of Public Health, and Radcliffe Institute for Advanced Study.

The Harvard Library is a 73-unit library system whose holdings include 17 million volumes, more than 21 million digital files, 10 million microforms, 8 million photographs, and an estimated 400 million manuscript items. The library currently uses the Ex Libris Aleph Integrated Library System as the basis for
its catalog including holdings and circulation information, and has fully implemented the MARC21 format for holdings.

B. Objectives

The Harvard Library has defined a strategic objective to: Enable effective access to the world of knowledge and data through intuitive discovery, networks of expertise, and global collaborations. To achieve this objective, a Discovery Platform Investigation group has been charged with identifying and evaluating options currently available and in development for a new web-scale discovery system for the Harvard Library. Web-scale here is defined as a system spanning local and global resources, regardless of format or location. The system should have an intuitive interface, support interdisciplinary research, integrate special collections and research guides, and aggregate local, licensed, and open metadata about the scholarly record, including non-textual media and grey literature. The system should also be capable of integrating information allowing users to determine the extent of the resources available to them, regardless of format.

C. Web-scale discovery

1. Please describe how the system supports searching across licensed databases, indexes, and local content.
   a. How does the system integrate article-level results with results from a customer’s integrated library system, archival system with finding aids, image system, institutional repository, geospatial resources, social science datasets, etc.?
   b. How does the system handle relevance-ranking for these different data sources?
   c. What options are available for local configuration of the relevance-ranking feature?
   d. How does the system handle faceting for these different data sources? How is local data included in facets?
   e. How do you extend search beyond content that is included in the central index? How are these results presented in relation to central index and local holdings search results?

2. The vendor-supplied core index is a key feature for web-scale discovery tools.
   a. Please identify all of the types of content are included in your index.
   b. How frequently is the content of the aggregated index updated? What process is in place for correcting data issues and errors in aggregated index records?
   c. How frequently is new content added?
   d. How does the system support on-demand lookup for patrons to determine if a journal or package is indexed in your tool? What does the user see?
   e. Please provide your 6 month and 12 month roadmap for adding major data to aggregated index.

3. Please describe how the system ingests/harvests other data sources such as institutional open access repositories.

4. What features are available for collaborative discovery of resources in partner institution catalogs?

5. What data enrichment options are currently available (e.g., published reviews, abstracts/summaries, ratings, cover images, table of contents, sample pages, altmetrics)?

6. What is supported for Demand-Driven Acquisitions (DDA)/Patron Driven Acquisitions (PDA) services?
D. Aggregated index coverage analysis

1. Please deliver a content coverage comparison for electronic holdings. We will provide an export of our electronic holdings. Please return:
   a. A deduplicated list of our ISSN that are covered.
   b. A deduplicated list of our ISSN that are not covered.
   c. Please include complete information on each for depth of inclusion, such as full-text, abstract, subject keywords, or metadata.
   d. Please include range of coverage (to/from dates).
2. Please describe all types of content included in the aggregated index in addition to journals. How are major changes to content disclosed to customers?
3. What ebook platforms are indexed? Is full-text indexed?

E. Data types / indexing

1. To what extent are the advanced search indexes configurable? What indexes are not configurable?
2. How does the system handle authority files? Are they displayed and actionable? For example, can you click on a subject term in a record to re-execute a search, or follow an author or geographic cross-reference? How are cross-references from the authority files used in the system? Which authority files are employed by the system?
3. What metadata standards can be ingested into the system (MARC, EAD, VRACORE, local xml (flat or nested), etc)? Please provide example sites for each type of data that has been ingested.
4. How does the system handle the MARC holdings format? Can data from holdings be indexed / faceted?
5. How does the system handle Aleph item data? Can data from items be indexed / faceted?
6. How does the system handle local MARC tags in bibliographic and holdings records? Can the data be indexed? Can display be configured?
7. How does the system handle the normalization of punctuation and abbreviations?
8. How does the system handle vernacular data, e.g., CJK, Arabic, Greek, Cyrillic, Hebrew, Devanagari, etc.? How does it handle indexing and display of these scripts? Are these scripts/languages supported fully in facets and spelling correction?
9. To what extent does the system normalize precomposed Unicode characters (extended Latin, CJK, Arabic, Greek, Cyrillic, Hebrew) for search and retrieval, including phonetic modifiers?
10. How does the system handle traditional Chinese and simplified Chinese for searching?
11. How does the system handle data from objects linked within records? For example, can the system index full-text of locally digitized objects that are linked to from a bibliographic record?
12. How will RDA and LC's Bibliographic Framework initiative be handled?
13. Please describe how the system supports linked data. Does it create/publish linked data or incorporate linked data to enhance records or incorporate linked data into system functionality?
14. Is there a defined export format required for local data?
15. How is the system positioned to respond to changes and innovations in the way library data is distributed?

F. Licensed Content Management

1. How are subscriptions for licensed content maintained and activated within the system?
2. How does the system interface with the institution's OpenURL resolver to determine the linked copy?
3. What benefits are there to end users in using your OpenURL resolver and knowledge base, versus using our current SFX system?
4. What process is in place to synchronize from our local SFX knowledge base to the discovery knowledge base (if required)? How frequently can this synchronization happen?

G. Interface Features
1. How are mobile / tablet interfaces supported? Please provide examples of mobile sites.
2. How does the system query underlying ILS system to display accurate availability? Are there limitations on the number of holdings/items the system can query for a given resource?
3. Can links be opened in new tabs/windows? Can each result be opened in a new tab? Can hyperlinked searches be opened in new tabs?
4. How is RSS or other SDI supported? What can the feed be based on (new records, searches, subjects)? How are new records determined? Is there any automated tracking for new records added to the system for use in new book lists, etc?
5. Please describe any "recommender" services.
6. Please describe facet configuration options. What data is used to build facets? Can we map our local data, at a field level, into existing or new facets? Can facets be built "on the fly" for things like availability? Can location data be mapped into multiple facets? Can facets be re-ordered and re-labeled and removed? Can data points be re-ordered (for example, year ranges)
7. Are there limits to the number of facets? Or the numbers of data points that can be used to build the facet (either record number or fields)?
8. Please describe what happens if a record is eligible for multiple facets entries because it qualifies for more than one (i.e., format: a single record may be both Book and Online)?
9. Do the facets offered to users change based on their selection?
10. What record clustering functionality is offered? Is this FRBR based? Can it be customized?
11. How is relevance ranking determined? Please provide as much detail as possible.
12. How is relevance configured for local records? Merged records? Aggregated index records? Is relevance locally configurable at a field/subfield level for all records?
13. How are left-anchored / browse searches supported? For which data points is this supported (author, subject, call number, title, local classifications?)
14. How is relevance handled?
15. How is stemming supported during simple searches? Is fuzzy searching supported (Mac vs. Mc)?
16. For search results, is there a "Select All" feature? What functions can be performed once selected? Does it apply to a page of results or an entire result set? Can the select be modified (i.e., select all results and then remove a few)?
17. If a user navigates to a full record from a result set and then returns to the result set, what happens? Are they returned to the point in the list they were at when they left?
18. Please describe options for visualization of search results (e.g., tag clouds)? Can this be customized?
19. Please describe what functions are available to manage a result set (sorting by author/date/title, ascending or descending, etc., search within, modify original search, etc.). What is the default setting? Can we change default?
20. What is presented to a user when a search fails? How can failed-search text be customized? Can links or widgets be included in the failed search display?
21. Can fields in records be hyperlinked to yield new searches? Is this configurable?
22. How are persistent links to records supported? To searches? To searches with selected facets?
23. Are search terms highlighted? Does this apply to all data sources, including any additional record enhancement data, such as tables of content or review services?
24. Are there next/previous buttons for navigating result sets? Is the context presented to the user (e.g., Record 3 of 10)? How can users navigate within a result set (e.g., set number of results per page, toggle between first/last page of the result set, jump to a specific page or record)?
25. Please describe advanced search options. Are Boolean searches supported? Can multiple formats/locations be selected?

H. Localization
1. To what extent can the system be branded? Headers, footers, images, styling, format icons, labels, etc.
2. To what extent can the core system be "rebranded" for internal groups of constituents? Is this just the skin, or can this affect facets/relevancy?
3. How do we add links to external services within the interface, such as links to account systems, reference systems, and external fulfillment systems? Which sections of the interface support such links?
4. To what extent can the displays for various record types be configured?
5. How does the system support electronic bookplates? Please describe the options for display, indexing, and hyper-linking options for donor result pages. This functionality can be seen in the right hand column of full record view in our current discovery system (please see this record for the Katherine B. Loker Fund for Widener Library). The display utilizes a budget string embedded in a local field in the bibliographic data to map to the appropriate display name for the fund, pulls the thumbnail image for the book plate from our digital repository, and creates the link to the library’s donor recognition page. The fund string in the marc record is indexed in the discovery system; this allows for execution of a search to display items purchased on that fund.
6. What support is there for adding library-provided search suggestions? For example, if a user searches for "hours" the system responds with "Are you looking for library hours?" with a hyperlink to a local web page.

I. User Features
1. What is available in terms of user-specific tools such as: save searches, create favorites, save lists, perform actions on lists, create folders, subfolders.
2. How would the user account be established (e.g. tied to local authentication systems? Or requiring a username and password internal to the system)?
3. What login mechanisms are supported (e.g., IP, EzProxy, LDAP, Shibboleth)?
4. What types of user exports are available (e.g, print records, email records, text)? Is this supported for both individual records and lists of records?
5. Please describe any available hooks for integrating with other systems to access account features.
6. What functions are available for users to add content to records (tags, reviews, annotations, links)? How is this data used by the system for indexing/facets? Is this data aggregated across all customers and available to support search?

7. What social media functions are available?

8. What export functions to citation managers are included (RefWorks, EndNote, Zotero, Bibtex, Mendeley, etc.)? Are these available for all data formats?

9. Which parts of your system require authenticated users for access? Can non-affiliated users access our local catalog data?

10. How does the system offer spelling help/correction?

11. How does the system provide search term auto-completion? Can this be customized?

12. How well does the interface respond to browser functions, like back, print, etc?

13. Please describe your system’s end-user interfaces for the following functions when integrated with a standards-based Integrated Library System (Aleph), ILL System (ILLiad), Consortial Borrowing System (Borrow Direct), Special Requests System (Aeon). Indicate which, if any, of these user interfaces cannot be rendered natively in your system, or which require an OPAC or other end-user system.
   a. Viewing of circulation account information, initiating renewals, viewing of fines information.
   b. Placement of circulation and/or digitization requests (recall, campus delivery, intercampus delivery, electronic delivery) viewing of holdings and item details.

14. What type of user help / documentation is offered? Is this configurable?

15. Do facets include options to exclude as well as include a value? Can a user select multiple values per facet?

J. Accessibility
   1. To what extent does the system adhere to WCAG accessibility guidelines?
   2. To what extent has the system been tested with user-assistive software?

K. Statistics
   1. What usage statistics are logged in the system? What granularity for user actions? Can this data be harvested for ingest into other reporting environments?
   2. Are raw usage logs and documentation of what is logged provided to facilitate local analysis?
   3. Can statistics be broken down by affiliation, user type (faculty, grad, undergrad) etc.? How are these data reported?

L. Support
   1. How are upgrades scheduled and deployed?
   2. What is the frequency of major and minor releases?
   3. Please describe your SLA model for this system.
   4. Please describe the failover environment for the discovery interface (if hosted) and the core index. We maintain 24x7 availability for our high-use systems.
   5. Please describe the Dev/QA environment (if discovery interface is hosted).
   6. What tools are available to help users connect with users at other institutions?
7. How do you collect user feedback about the system? Who provides input on which user requested features are incorporated? Please provide a recent example of enhancing the system based on user feedback.

M. Architecture
1. What platform/OS does the system run on?
2. What is the server architecture?
3. What are the server requirements?
4. What API’s does the system offer?
5. Do these API’s exclude any content?
6. Please describe how the local institution can add widgets or other functionality within the provided discovery framework. Does the system come pre-configured with any widgets (Google Book Search, Hathi Trust, etc.)?
7. How does the system expose local data for external harvesting? For example, can our local holdings be exposed to Google, etc.?
8. What options are available for scaling to customers of our size? Please describe large customers and provide information about peak response time.
9. How does the system handle incremental updates (e.g., adds, replaces, deletes)? What is the frequency for ingesting incremental updates for library catalog content (hourly, daily, weekly)? What is the frequency for ingesting updates for other library content (hourly, daily, weekly)? What is the typical load rate for records per hour? For indexing? What are load and indexing rates sensitive to (e.g., record size, total database size, total number of records, processing speed, I/O, available storage space)?
10. What methodology is used to check availability in external systems such as ILS?
11. What database is used to store data?
12. What technology used to create indexes?
13. What browsers/versions/operating systems are supported by the user interface?

N. Costs
1. What are the initial fees for implementation?
2. What are annual license/subscription fees? Does this include technical support? If so, please describe.
3. How are fees assessed (data, FTE, etc.)?
4. How many local staff, and what roles, will be needed to support this system?

O. Implementation Process
1. Please describe the available models/service plans for local implementation (e.g., hosted, or SaaS).
2. Describe qualifications of likely project personnel from your organization? Describe the tasks and roles you would expect the customer institution to fill?
3. What tasks are performed during implementation? Who performs these tasks?
4. Please provide project timelines for a project of our scale and provide and estimate for your availability to start on this project.
P. Company information
   1. Provide a brief description of what distinguishes your company from others.
   2. Provide a current customer list.
   3. Describe how your discovery platform offering fits into your company portfolio and your plans for developing this product over the next 3-5 years.