Questions to Consider

Abstract

For numerous reasons, libraries contemplating the purchase of a web scale discovery service should very carefully complete their homework. The following set of questions can serve as a primer for those engaging in their own evaluations of the increasingly competitive library web scale discovery space. Questions are divided into several topical areas: general and background questions; local library resources; publisher and aggregator indexed content; open access content; relevancy ranking; authentication and rights management; and the user interface.

The following questions can serve as a springboard for prospective library customers seeking more information from vendors as part of their own evaluation of library Web scale discovery services.

Section 1: General and Background Questions

1. **Customer Install Base**

    - How many current customers do you have that have implemented the product at their institutions (i.e., the tool is currently available to users or researchers at the institution)?

    - How many additional customers have committed to the product?

    - How many of these customers fall within our library type (e.g., higher ed academic, public, K–12)?

2. **References**

    - Can you provide website addresses for live implementations that you feel serve as a representative model matching our library type?

    - Can you provide references—name and contact information—for the lead individuals you worked with at several representative customer sites that match our library type?

3. **Pricing Model, Optional Products**

    - Describe your pricing model for a library type such as ours, including initial up-front costs and ongoing costs related to the subscription and technical support.

    - What optional add-on services or modules (federated search, recommender services, enrichment services) do you market that we should be aware of and that are related to and able to be integrated with your discovery solution?

4. **Technical Support and Troubleshooting**

    - Briefly describe options (including hours of availability) customers have for reporting mission-critical problems and for reporting observed non–mission-critical apparent glitches.

    - Briefly describe any consulting services you may provide above and beyond routine support services (e.g., consulting services related to harvesting of a unique library resource for which an ingest/transform/normalize routine does not already exist).

- Is there a process for suggesting enhancements for potential future incorporation into the product?

5. **Size of the Centralized Index.** How many periodical titles does your preharvested centralized index encompass? What is the current count of unique indexed items?

6. **Statistics**

   - Please describe what you feel are some of the more significant use, management, or content-related statistics available out of the box with your system.

   - Are the statistics COUNTER compliant?

7. **Ongoing Maintenance Activities, Local Library Staff**

   - For instances where the interface and discovery service are hosted on your end, please describe any ongoing local library activities associated with maintaining the service for the local library's clientele (e.g., ongoing maintenance associated with periodic local resource harvest updates, etc.).

   - For instances where the hardware and interface are hosted at the local library, please describe any ongoing local library activities associated with maintaining the service for the local library's clientele (e.g., backup responsibilities, hardware troubleshooting, etc.). Note: Most systems are hosted by the vendor; this question applies to only a few select vendors.

Section 2: Local Library Resources

8. **Metadata Requirements**

   - What mandatory record fields for local resources must exist for the content to be indexed and discoverable within your platform (e.g., title, date)?

   - To what degree can collections from different sources have their own unique field information that is displayed or figures into the relevancy-ranking algorithm for retrieval purposes? For example, one particular digital collection at the library has unique fields pertinent to that collection. How would such unique fields be handled by the discovery service?

9. **Crosswalking and Existing Ingestors**

   - Are both local and remote content normalized to a single schema? If so, please offer comments on how local and remote (publisher or aggregator) content is normalized to this single underlying schema.

- Please verify that your platform has existing connectors and standard ingest/transform/normalize tools or application profiles for the following schema or standards used by local systems at our library (e.g., MARC 21 bibliographic records; Unqualified/Qualified Dublin Core, EAD, etc.).

- If existing ingestors do not exist, please describe any tools your discovery platform may offer to assist local staff in crosswalking between the local library database schema and the underlying schema within your platform.

10. **Schedule**

    - For records hosted in systems at the local library, how often do you harvest information to reflect record updates, modifications, deletions?

    - Can the local library invoke a manual harvest of locally hosted resource records on a per-resource basis (e.g., if the library launches a new digital collection and wants the records to be available within the discovery service shortly after they are available in our locally hosted digital repository, is there a mechanism to force a harvest prior to the next regularly scheduled harvest routine)?

    - After routine harvesting, how long does it typically take for such updates, additions, and deletions to be reflected in the discovery service?

11. **Policies and Procedures.** Please describe any general policies and procedures not already addressed of which the local library should be aware as they relate to the harvesting of local resources.

12. **Examples of Local Collections Incorporated into the Discovery Solution**

    - Our library uses the ABC digital collection management software. Do you have any existing customers who also utilize this platform, whose digital collections have been harvested and are now exposed in their instance of the discovery service?

    - Our library uses the ABC institutional repository software. Do you have any existing customers who also utilize this platform, whose IR materials have been harvested and are now exposed in their instance of the discovery service?

13. **Consortial Union Catalogs.** Can your discovery service harvest and provide access to items within a consortial or otherwise shared catalog (e.g., an Innovative Interfaces INN-Reach catalog). Please describe.

Section 3: Publisher and Aggregator Agreements and Indexed Content

14. **Publisher and Aggregator Agreements:** General

    • With approximately how many publishers and aggregators have you forged content agreements?

    • Are there major publisher or aggregator agreements that you feel are especially significant for your service? If so, which ones and why (e.g., other discovery platform vendors may not have such agreements with those particular providers, the amount of content was so great that it greatly augmented the size and scope of your service, etc.)?

    • Are these agreements indefinite, or do they have expiration dates?

    • Have you entered into any exclusive agreements with any publishers or aggregators (i.e., the publisher or aggregator is disallowed from forging agreements with competing discovery platform vendors or disallowed from providing the same deep level of metadata or full text for indexing purposes).

15. **Metadata or Full Text Provided by Publishers and Aggregators.** Could you please provide some comments and detail on the level of metadata provided to you, for indexing purposes, by the majority of major publishers and aggregators with whom you've forged agreements? Please describe to what degree the following data elements are provided by these agreements and play a role in your discovery service:

    A. Basic bibliographic information (article title, journal title, author, publication information)

    B. Subject descriptors

    C. Keywords (author-supplied?)

    D. Abstracts (author-supplied?)

    E. Full text

16. **Topical Content Strength**

    A. Do you feel there is a particular content area your discovery service covers especially well or leans heavily toward (e.g., humanities, social sciences, sciences)?

    B. In what subject or content areas, if any, do you feel the discovery service may be somewhat weak? Are there current efforts to mitigate these weaknesses (e.g., publisher agreements on the horizon)?

    C. Do you feel there is a particular format type that your discovery service covers very well or leans heavily toward (scholarly journal content, magazine content, newspapers, conference proceedings, etc.)?

17. **Content Overlap Analysis.** Please describe what sort of content overlap analysis you can provide comparing our local library holdings with titles covered in your discovery service.

    • **Content Considered Key by Local Library (by publisher).** Following is a list of some major publishers whose content, licensed by the library, is considered key. Has your company forged agreements with these publishers to harvest their materials? If so, please describe the scope of the agreement and whether the materials are preharvested into your central index or incorporated through an optional federated search mechanism. How many titles are covered for each publisher? What level of metadata are they providing to you for indexing purposes (e.g., basic citation-level metadata [title, author, publication date], abstracts; full text)?

        A. ex. Elsevier

        B. ex. Sage

        C. ex. Taylor and Francis

        D. ex. Wiley-Blackwell

    • **Content Considered Key by Local Library (by title).** Following is a list of the top 100 major journal or newspaper titles whose content, licensed by the library, is considered key. Could you please indicate whether such titles are incorporated into your preaggregated central index or as a target in an optional federated search mechanism? If so, please describe the level of indexing (e.g., basic citation-level metadata [title, author, publication date], abstracts; full text).

        E. ex. Nature

        F. ex. American Historical Review

        G. ex. JAMA

        H. ex. Wall Street Journal

18. **Google Books/Google Scholar/Hathi Trust.** Do any agreements exist at this time to harvest the data associated with Google Books, Google Scholar, or Hathi Trust into your discovery service? If so, are such items incorporated into your preharvested central index or as a target in an optional federated search product offered by your company? Please describe the level of indexing (e.g., basic citation-level metadata [title, author, publication date], abstracts, full text).

19. **WorldCat Catalog.** Does your service include the OCLC WorldCat catalog records? If so, are such items incorporated into your preharvested central index or as a target in an optional federated search product offered by your company? What level of information is included? The complete record? Holdings information?

20. **E-Book Vendors.** Does your service include items from major e-book vendors?

21. **Record Information.** Given the fact that the same content (e.g., metadata for a unique article) can be provided by multiple sources (e.g., the original publisher of the journal itself, an open-access repository, a database or aggregator, another database or aggregator, etc.), please provide some general comments on how records are built within your discovery service. For example:

   A. You have an agreement with a particular publisher or aggregator that agrees to provide you with rich metadata for its content, perhaps even provide you with indexing it has already done for its content, or provide you with the full text for you to be able to "deep index" its content.

   B. You have an agreement with a particular publisher that happens to be the only publisher or provider of that content. It may provide you rich information, or it may provide you rather weak information. In any case, you choose to incorporate this into your service, as it is the only provider or publisher of the content. Or alternatively, it may not be the only publisher or provider of the information, but it is the only publisher or provider with which you've currently entered into an agreement for that content.

   C. For some items appearing within your service, content is provided from multiple sources with which you've made agreements. There may be some or many cases of overlap for unique items, such as a particular article title. In such cases, do you create a merged, composite, or super-record where your service utilizes metadata from each of the multiple sources, creating a strong single record built from these multiple providers of content?

22. **Deduplication.** Related to the question immediately above, please describe your discovery service's approach to deduplicating items (or not) in the central preaggregated index. If your discovery service incorporates content for the same unique item from more than one content provider, does your index retrieve and display multiple instances of the same title? Or do you create a merged, composite, or super-record that is the only record displayed? Please describe.

Section 4: Open-Access Content

23. **Does your discovery service automatically include (out of the box, at no additional charge) materials from open-access repositories?** If so, could you please list some of the major repositories included (e.g., arXiv.org e-Prints; Hindawi Publishing; Directory of Open Access Journals; etc.)?

24. **Open-Access Content Sources: Future Plans.** In addition to the current open-access repositories that may be included in your discovery service, are there other repositories whose content you are planning to incorporate in the future?

25. **Exposure to Other Libraries' Bibliographic, Digital Collection, or IR Content.** Would ILS bibliographic records from other customers using your discovery service be exposed for discoverability in our instance of the discovery service? Would digital collection records? Institutional repository records?

Section 5: Relevancy Ranking

26. **Relevancy Determination.** Please describe some of the factors that figure into your discovery service's relevancy algorithm (fields and elements that play a role and their relative weighting in the relevancy determination).

27. **Currency.** Please comment on how heavily currency of an item is weighted in relevancy determination. Does currency factor more heavily for certain content types (e.g., newspapers)?

28. **Local Library Influence.** Does the local library have any influence over the relevancy determination? Can it choose to "bump up" particular items or local collections for a search? Please describe.

29. **Local Collection Visibility.** Please offer some detail on how local content (e.g., ILS bibliographic records, digital collections) remains visible and discoverable within the larger pool of content indexed by your discovery service. For example, local content may measure a million items, while your centralized index may cover half a billion items.

30. **Exposure of Items with Minimal Metadata.** Some items likely have weaker or less rich metadata than other items. Could you please offer some comments on how your discovery service ensures discoverability for items with lesser or minimal metadata?

31. **Full-Text Searching.** Does your discovery service offer the capability to search the full text of materials (i.e., is a user searching a full-text keyword index)? If so, approximately what percentage of

items within your service are indexed at the full-text level?

32. **Please describe what is presented by your discovery service when no results are retrieved for a search.** Does your system enable "best-match" retrieval—that is, something will always be returned or recommended? What elements play into this determination; how is the user prevented from having a completely dead-end search? What Did You Mean? functionality is included with your discovery service?

Section 6: Authentication and Rights Management

33. **Open or Closed Nature of Your Discovery Service.** Does your discovery service offer an unauthenticated view or access? Please describe and offer some comments on what materials will not be discoverable or visible for an unauthenticated user versus an authenticated user. It's assumed that licensed full-text materials would be available only to authenticated users (such as through a login or IP authentication). Please offer comments on the discoverability of the following:

    A. Records solely sourced from abstract and indexing databases (i.e., That database is the only provider of that specific content within your discovery service.)

    B. Citation information for content sourced from multiple content providers (Please describe the general level of citation information available to an unauthenticated user versus the level available to an authenticated user.)

    C. Enrichment information (e.g., book cover images, tables of contents, abstracts)

34. **Exposure of Nonlicensed Resource Metadata.**

    • Directly related to the question above, if one weren't to take into account *any* e-journal, publisher package, or database subscriptions and licenses the local library pays for, is there a base level of citation information that's exposed and available to all customers of your discovery service? This may include open-access materials or bibliographic information for some publisher or aggregator content (which often requires a local library license to access the full text). Please describe.

    • Would a user need to be authenticated to search (and retrieve results from) this base index?

    • How large is this base index that all customers

may search, regardless of local library publisher and aggregator subscriptions.

35. **Rights Management.**

    • Please describe how rights management is initialized and maintained in your discovery service for purposes of determining whether a local library user should have access to the full text (or, if not full text, full resolution for non–full-text resources—e.g., full resolution to the abstract or more complete citation as may be found within an A&I resource).

    • Our library uses the ABC link resolver. Our library uses the ABC A–Z journal listing service. Our library uses the ABC electronic resource management system. Is your discovery service compatible with one or all of these systems for rights management purposes? Is one approach preferable to the other, or does your approach explicitly depend on one of these particular services?

Section 7: User Interface

36. **Openness to Local Library Customization.** Please describe how open your discovery service is to local library customization. For example, please comment on the local library's ability to

    A. Rename the service

    B. Customize the header and footer hyperlinks and color scheme

    C. Choose which facet clusters appear

    D. Define new facet clusters

    E. Embed the search box in other venues

    F. Create "canned" precustomized searches for an instance of the search box

    G. Define and promote a collection, database, or item such that it appears at the top or on the first page of a relevant, toipically related search

    H. Develop custom widgets offering extra functionality, or download widgets from an existing user community

    I. Incorporate links or direct inline viewing for external enriched content (e.g., Google Books previews; Amazon.com item information)

    J. Other things you feel are worth noting

37. **Social Features.** Please describe some current social Web features present in your discovery service (e.g., user tagging, ratings, reviews, etc.).

What, if any, plans do you have to offer or expand such functionality in future releases?

38. **User Accounts.**

- Does your discovery service offer user accounts?

- What services does the user account provide?

  A. Save a list of results to return to at a later time?

  B. Save queries for later searching?

  C. See a list of recently viewed items?

  D. Perform typical ILS functions such as viewing checked-out items, renewals, and holds?

  E. Create customized RSS feeds for a search?

39. **Mobile Interface.** Please describe the mobile interface(s) available for your discovery service.

Is it a browser-based interface optimized for small screen devices? Is it a dedicated iPhone-, Android-, or BlackBerry-based application?

40. **Usability Testing.**

- Please describe how your discovery service incorporates established best practices in terms of a customer-focused, usable design including modern Web interface elements expected by today's users.

- What usability testing have you performed or do you conduct on an ongoing basis?

- Are you aware of any existing customers that have conducted usability testing of their own related to your discovery service?