

UKSG webinar – Optimizing the discovery experience through dialogue – a community approach, 10 January 2017

Speakers:

Lettie Conrad, Publishing and Product Development Consultant
Kathleen Donovan, Research Librarian, Harvard University
Bruce Heterick, Vice President, Outreach & Participation Services, ITHAKA/JSTOR
Rachel Kessler, Product Manager, ProQuest
Alexa Pearce, Head of Social Sciences at the University of Michigan Library, University of Michigan

E-mails:

lettieconrad@yahoo.com
kathleen_donovan@harvard.edu
bruce.heterick@ithaka.org
Rachel.Kessler@proquest.com
alexap@umich.edu

Should you have a question arising from the webinar and not answered below, please e-mail the appropriate speaker(s) who will be happy to respond.

Questions arising from the Q&A session and the registration forms:

Q1: Libraries want to ensure that content and collection information from publishers is correct and updated. How can we help that business case?

A: Answered in session. Please review the recording.

A: From Bruce: I would just like to add to the answer provided in the session and say that it is important that libraries reach out to the publisher when they notice that the content and/or collection information is not correct or updated in the discovery service knowledgebase. As a content provider, sometimes we are not even aware when this is an issue. We do our best to keep the discovery providers informed of new content, new packages, new collections, etc., but there can be delays in seeing those changes reflected in the knowledgebases and administrative modules. So, be sure to let the content provider know if you are noticing discrepancies.

Q2: How can discovery systems be tuned to provide for specific local user populations and their needs?

A: Answered in session. Please review the recording.

Q3: For Kathleen -- she said that users were concerned about conformity to usual search expectations. What does this mean?

A: From Kathleen: I was referring to choices discovery system vendors make about basic screen layout, navigation and links. Users expect search systems to follow the conventions that are familiar to them from their use of major commercial sites. For example, links should be able to be opened in new tabs, or selecting an item should result in some visual confirmation of that action. Other examples would be filter/facet placement or placement and naming of help links or icons.



UKSG webinar – Optimizing the discovery experience through dialogue – a community approach, 10 January 2017

Q4: ODI risks obsolescence by focussing on legacy metadata standards - the role of machine augmentation, in terms of natural language processing and text/data mining, has been mentioned in passing but what is the role of ODI in consolidating innovation and supporting adoption of those technologies at scale?

A: From Lettie: ODI recommends those metadata standards that are required by discovery service providers for indexing scholarly resources. I think the responsibilities here lie more so with each individual constituency: discovery service providers should be leading the way in what metadata is optimum to perform within their technological environments; libraries should be applying market pressure to demand the innovations that best fit their visions for localized content discovery; and publishers should be working with discovery providers to deliver on library expectations. The role of ODI is to support and respond to any technological evolutions within the community as they occur.

A: From Alexa: I agree with Lettie's response. The CRL model license (<http://liblicense.crl.edu/licensing-information/model-license/>) provides text for TDM licensing that libraries may be trying to negotiate. If ODI is successful with its first goal of encouraging more effective data sharing for discovery of scholarly and popular resources, it will be in a better position to include considerations around these additional technologies.

A: From Bruce: My personal belief is that the discovery services we know today will have to adapt to the natural language processing and machine augmentation described above, just as other starting points for research will be required to do so by the market. I'm not sure if ODI has a role in this or not, as I believe the "market" driving this adaptation will be at the user level, rather than the library. I think it would be a mistake to look at ODI and expect it to be more than what it was intended to be when it was created. It may simply have a "natural life" that extinguishes once its usefulness is outlived.

Q5: I think one thing Google is better at than discovery tool systems is the speed with which new content is indexed and made available. Our discovery tool often takes several weeks to make the latest content available - whereas Google is much, much quicker. Do the speakers have any thoughts on this (especially the library speakers)? Is it an issue for their students? Or isn't access to the latest research seen as critical?

A: From Lettie: Amen! Really wished we had time to discuss, as this is a very important issue for everyone involved! I spent a good deal of time unpacking the [Stumbling Blocks](#) presented by Roger Schonfeld of Ithaka S+R and found that temporal gaps were the biggest culprit in delaying the indexing and linking required for authenticated access to the newest research. In my experience, this is most often caused by publishers not supplying sufficient, accurate, or timely metadata that comply with industry standards or discovery service provider requirements. Actually, this is a great example why ODI champions the legacy metadata standards required by discovery providers (see Q4).



UKSG webinar – Optimizing the discovery experience through dialogue – a community approach, 10 January 2017

A: From Alexa: This is a good point, though my sense is that it's uneven – Google has some content faster than some discovery services but I don't think that is true for all content and all discovery systems. For me, this comes down to using the right tool for the right job. When I'm helping people do research, I never frame selection of a search interface as a mutually exclusive thing, i.e., "use ArticlesPlus and NOT Google." Google is one part of the search ecosystem. Sometimes a user is best served through a Google search, sometimes through a discovery service search, and sometimes through a subject specific A&I tool. Many times, through combinations of the above and some toggling between various tools. There is not one tool that does it all.

A: From Kathleen: First, I agree completely with Alexa about seeing the discovery system as one tool in your research toolbox. In most of my interactions with students, I end up showing a mix of these tools, as often they are complementary. Currency in discovery systems is difficult to measure and bears further investigation. I think I worry more about currency than the students I assist, but it is a little less critical for researchers in the social sciences. My science librarian colleagues would have a very different perspective on this issue.

Q6: Might the reason for JSTOR low referrals be due to lack of subject indexing? Will JSTOR ever provide this?

A: Answered in session. Please review the recording.

Q7: Is ODI considering comparisons of different types of discovery systems - not just those that rely on an index of pre-harvested metadata?

A: From Rachel: ODI does not compare discovery systems but rather promotes transparency via the best practices and tools like conformance checklists that encourage discovery systems to disclose their level of participation, thus making it easier for libraries to compare discovery systems. That said, most of the recommendations do focus on the discovery system central index as the committee since its inception has felt that the index was the greatest cause of ambiguity. As a group, we are currently exploring emerging technologies in discovery, and we'd encourage these providers to submit conformance statements as well whenever relevant.

Q8: If as a university library you identify a publisher's content (which you have an active subscription to) is not or not fully being discoverable in the discovery service - who do you reach out to; discovery service vendor or publisher?

A: Answered in session. Please review the recording.

Q9: I've recently started working at University of Strathclyde, where we use Primo. We use the system Metalib. I'm very new to it, but my first impression is that it seems to be quite dated and not fully supported in terms of making content cross-searchable in the discovery layer; and I was wondering are there alternative systems to Metalib?

A: From Kathleen: Because of its out-dated look and limited cross-search functionality, we use a very stripped-down version of Metalib to host our A-Z database list, and rely on vendor platforms or our



UKSG webinar – Optimizing the discovery experience through dialogue – a community approach, 10 January 2017

discovery system to get cross-search functionality. Some sites are using LibGuides to manage their A-Z list (see Wesleyan University: <http://libguides.wesleyan.edu/az.php?>)

A: From Rachel: MetaLib relies on federated search technology, which returns results much more slowly than discovery services. The advantage of MetaLib is 1) It provides an A to Z list of subscribed databases; 2) It allows users to search a subset of desired databases rather than all available content; and 3) It allows users to search databases that might be missing from a discovery central index.

In terms of compatibility with Primo, MetaLib is the federated search product most compatible with Primo and most functionality should be supported within the Primo interface. If you have specific questions, I would be happy to help answer them or put you in touch with someone from Ex Libris.

Q10: Question for Bruce - please could you tell us some more about stripping out of persistent identifiers in discovery systems?

A: Answered in session. Please review the recording.

Q11: Do any of the speakers have any comments about the risks for libraries (and research) of concentrating patrons' research activity into one subscription discovery tool?

A: From Lettie: Sure – I think there's a risk to all of us if users concentrate all their information seeking on a single tool. Until there's a monopoly ☺ I think we all need to encourage and educate users on the value of different tools to serve their research needs.

A: From Alexa: Agreed and my answer above to Q5 is also applicable here. I don't know any librarians who think we can direct users to a single tool.

A: From Kathleen: I do have concerns about this since we still need more transparency about discovery system content, updating, results ranking and linking. I worry even more about the risk of users doing all their research activity in one non-subscription tool (Google) because it is even less transparent.

Q12: What testing do university libraries do to check their subscription to a publisher's content is fully discoverable in the discovery service? E.g. Linking on campus and off campus as well as article counts/content sets

A: From Alexa: Most university libraries do not have the resources to do extensive testing. We have to operate on good faith – if a publisher tells us that they provide their content, we will be likely to believe them. Of course we pay attention to our observations, do testing, and reach out to publishers and vendors if it seems that their content is not actually included.

A: From Kathleen: To extend this one step beyond discovery, I'd add a concern about whether full text we subscribe to is accurately identified as available. Students are as concerned about immediate



UKSG webinar – Optimizing the discovery experience through dialogue – a community approach, 10 January 2017

access to full text as they are about the discovery of content. Since most libraries don't have the resources to do ongoing monitoring, problems are often identified by feedback from staff and other users.

Q13: Bruce et al - is the 4% quoted on the pie chart common for all content providers. Is this the figure for first search....

A: From Bruce: Let's remember that this is data specific to JSTOR, which has its own unique characteristics as compared to publishers. In aggregate, across 10,000+ participating institutions, we typically see 4-8% of our usage originate from discovery services like Primo, Summon, EDS, and WorldShare. There are lots of caveats to those numbers, however – one of the most prominent caveats being the fact that we (as a community) haven't demanded that the discovery providers provide a persistent identifier in their referring URL (see Question 10). Without that identifier, we (the content provider) can't truly measure the impact of these services on our usage, and you (the library) can't truly measure your return on investment. To me, solving this problem is paramount.

With that being said, I will also tell you that I've seen plenty of usage analysis from individual institutions where the discovery service was driving 20-25% of the usage to JSTOR. So, like they say, statistics are like bathing suits: what they reveal is interesting, but what they conceal is essential.

Q14: It is common that publishers submit their metadata to discovery services but it is then not made discoverable by the discovery vendors. How best can publishers resolve such issues and meet the institution's needs.

A: From Lettie: Yes, this happens – but, in my experience, never for sinister reasons, more often technological or operational errors. I think publishers and discovery providers must work together to ensure timely, accurate indexing – where publishers dedicate resource to semi-automated metrics and auditing exercises, working then with their discovery service partners to remediate gaps and safeguard against future errors. There's an [IEEE case study on discovery service gap analysis](#) that I think is a great example of ideal teamwork across the sectors.

A: From Bruce: I agree with Lettie, but I also believe this is a place where ODI can play a more prominent role. As a content provider, it is maddening that the discovery providers do not feel compelled to proactively communicate with us to let us know (a) they've received our content, and (b) that content has been processed. We send, we hope, we assume, but we never actually get confirmation (unless we ask). I honestly don't know why we put up with the opaqueness of it all.

Q15: What support is there for metadata in non-MARC formats, e.g. metadata in XML. Specifically looking for the same support as MARC data in terms of supporting updates and deletions without a re-index of the entire content which is very time consuming.

A: Answered in session. Please review the recording.



UKSG webinar – Optimizing the discovery experience through dialogue – a community approach, 10 January 2017

Audience comments shared in session:

Ideas to help libraries and publishers audit their discovery index contents:

01. Consistent names and unique identifiers for eResource collections, so we can know with certainty whether the content we license is the content that is turned on in discovery (and link resolvers, etc.)
02. Make a standard format for sharing collection information that the library, discovery vendors, and publisher can use to communicate to their eresources and database holdings to each other.
03. Create a tool the library can use to run an audit. The tool should allow the library to export details about of all the enabled content in *that library's* index (name of collection, its unique ID, any content coverage notes. Ideally, the tool would allow the library to get further detailed reports, such as title and holding coverage for a specific collection).
04. Publishers/vendors: provide report of a list of the content that they think the library should be able to enable in the index, using the name or unique ID that matches the name.

