

# Editorial Board Thoughts

## A&I Databases: the Next Frontier to Discover

Mark Dehmlow

---

I think it is fair to say that the discovery technology space is a relatively mature market segment, not complete, but mature. Much of the easy-to-negotiate content has been negotiated, and many of the systems on the market are above or approaching a billion records. This would seem a lot, but there is a whole slice of tremendously valuable content still not fully available across all platforms, namely the specialized subject abstracting and indexing database content. This content has a lot of significant value for the discovery community—many of those databases go further back than content pulled from journal publishers or full-text databases. Equally as important is that they represent an important portion of humanities and social sciences content that is less represented in discovery systems as compared to STEM content. For vendors of A&I content, the concerns are clear and realistic, differently from journal publishers whose metadata is meant to direct users to their main content (full text), the metadata for A&I publishers is the main content. According to a recent NFAIS report, a major concern for them is that if they include their content in discovery systems, they “risk loss of brand awareness” and the implications are that institutions will be more likely to cancel those subscriptions.<sup>1</sup> The focus therefore seems to have been how to optimize the visibility of their content in discovery systems before being willing to share it.

In addition to the NFAIS report, some of the conversations I have seen on the topic seem to focus on wanting discovery system providers to meet a more complex set of requirements that will maximize leveraging the rich metadata contained in those resources, the idea being that utilizing that metadata in specific ways will increase the visibility of the content. In principle I think it is a commendable goal to maximize the value of the comprehensive metadata A&I records contain, and the complexities of including A&I data into discovery systems need to be carefully considered - namely blending multiple subject and authority vocabularies, and ensuring that metadata records are appropriately balanced with full text in the relevancy algorithm. But I also worry that setting too many requirements that are too complicated will lead to delayed access and biased search results. It is important that this content is blended in a meaningful way, but determining relevancy is a complex endeavor, and it is critically important for relevancy to be unbiased from the content provider perspective and instead focus on the user, their query, and the context of their search.

Another concern that I have heard articulated is that results in discovery services are unlikely to be as good as native A&I systems because of the already mentioned blending issues. This is likely

---

*Mark Dehmlow ([mark.dehmlow@nd.edu](mailto:mark.dehmlow@nd.edu)), a member of the ITAL Editorial Board, is Program Director, Library Information Technology, University of Notre Dame, South Bend, IN.*

---

to be true, but I think it is critical to focus on the purpose of discovery systems. As Donald Hawkins recently wrote in a summary of a workshop called “Information Discovery and the Future of Abstracting and Indexing Services,” “A&I services provide precision discipline-specific searching for expert researchers, and discovery services provide quick access to full text.”<sup>2</sup> Hawkins indicates that discovery systems are not meant to be sophisticated search tools, but rather a quick means to search a broad range of scholarly resources and I think sometimes a quick starting point for researchers. Because of the nature of merging billions of scholarly records into a single system, discovery systems will never be able to provide the same experience as a native A&I system, nor should they. Over time, they may become better tuned to provide a better overall experience for the three different types of searchers we have in higher education: novice users like undergraduates looking for a quick resource, advanced users like graduate students and faculty looking for more comprehensive topical coverage, and expert users like librarians who want sophisticated search features to hone in on the perfect few resources. Many of the discovery systems are working on building these features, but the industry will take time to solve this problem, and I tend to look at things from the lense of our end users—non-inclusion of this content directly impacts their overall discovery experience.

One might ask, if the discovery system experience isn’t as precise and complete as the native A&I experience, why bother? In addition to broadening the subject scope by including many of the more narrow and deep subject metadata, there is also the importance of serendipitous finding. That content, in the context of a quick user search, may drive the user to just the right thing that they need. In addition, my belief is that with that content, we can build search systems that are deeper than Google Scholar, and by extension provide our end users with a superior search experience. And so I advocate for innovating now instead of waiting to work out all of the details. I am not suggesting moving forward callously, but swiftly. The work that NISO has done on the Open Data Initiative has resulted in some good recommendations about how to proceed. For example, they have suggested two usage metrics that could be valuable for measuring A&I content use in discovery systems: search counts (by collection and customer for A&I databases) and results clicks (number of times an end user clicks on a content provider’s content in a set of results).<sup>3</sup>

While I think these types of metrics are aligned with the types of measures that libraries evaluate A&I database usage by, I think at the same time they don’t really say much about the overall value of the resources themselves. Sometimes in the library profession, our obsession for counting stuff loses connection with collecting metrics that actually say something about impact. Of the two counts, I could see perhaps counting the result clicks as having more value. In this instance, knowing that a user found something of interest from a specific resource at the very least indicates that it led the user some place. I think the measure of search counts by collection is less useful. At best it indicates that the resource was searched, but it tells us nothing about who was searching for an item, what they found, or what they subsequently did with the item once they found it. I do think we in libraries need to consider the bigger picture. Regardless of the number of searches

(which doesn't really tell us anything anyway), we need to recognize the value alone of including the A&I content, and instead of trying to determine the value of the resource by the number of times it was searched, focus more on the breadth of exposure that content is getting by inclusion in the discovery system.

I think a more useful technical requirement for discovery providers would be to provide pathways to specific A&I resources within the context of a user's search—not dissimilar to how Google places sponsored content at the top of their search results, a kind of promotional widget. In this case, using metadata returned from the query, the systems could calculate which one or two specific resources would guide the user to more in depth research. By virtue of inclusion of the resource in the discovery system, those resources could become part of the promotional widget. This would guide users back to the native A&I resource which both libraries and A&I providers want, and it would do that in a more intuitive and meaningful way for the end user.

All of the parties involved in the discovery discussion can bring something to the table if we want to solve these issues in a timely way. I hope that A&I publishers and discovery system providers make haste and get agreements underway for content sharing and I would recommend that instead of focusing on requiring finished implementations based in complex requirement before loading content, both of them should instead focus on some achievable short and long term goals. Integrating A&I content perfectly will take some time to complete and the longer we wait, the longer our users have a sub-optimal discovery experience. Discovery providers need to make long term commitments to developing mechanisms that satisfy usage metrics for A&I content, although I would recommend defining measures that have true value. A&I providers should be measured in their demands: while their stakes in system integration is real, there runs a risk of content providers vying for their content to be preferred when relevancy neutrality is paramount for a discovery system to be effective. I think it is worth lauding the efforts of a few trailblazing A&I publishers such as Thomson Reuters and ProQuest who have made agreements with some of the discovery providers and are sharing their A&I content already, providing some precedent for sharing A&I content. Lastly, libraries and knowledge workers need to develop better means for calculating overall resource value, moving beyond strict counts to thinking of ways to determine the overall scholarly/pedagogical impact of those resources and they need to make the fact alone that an A&I publisher shares its data with a discovery provider indicate significant value for the resource.

---

## REFERENCES

1. NFAIS, *Recommended Practices: Discovery Systems*. NFAIS, 2013.  
[https://nfais.memberclicks.net/assets/docs/BestPractices/recommended\\_practices\\_final\\_aug\\_2013.pdf](https://nfais.memberclicks.net/assets/docs/BestPractices/recommended_practices_final_aug_2013.pdf).
2. Hawkins, Donald T., "Information Discovery and the Future of Abstracting and Indexing Services: An NFAIS Workshop." *Against the Grain*. , 2013. <http://www.against-the-grain.com/2013/08/information-discovery-and-the-future-of-abstracting-and-indexing-services-an-nfais-workshop/>.
3. Open Discovery Initiative Working Group, *Open Discovery Initiative: Promoting Transparency in Discovery*. Baltimore: NISO, 2014.  
[http://www.niso.org/apps/group\\_public/download.php/13388/rp-19-2014\\_ODI.pdf](http://www.niso.org/apps/group_public/download.php/13388/rp-19-2014_ODI.pdf).