

The open discovery initiative & promoting transparency in discovery

Geoffrey Morse^{a,*} and Ken Varnum^b

^a*Interim Head, Research Services, Northwestern University Libraries, 1970 Campus Drive, Evanston, IL, USA*

^b*Senior Program Manager, University of Michigan Library, Ann Arbor, MI, USA*

Abstract. NISO recently updated the Open Discovery Initiative Recommended Practice (<https://www.niso.org/publications/rp-19-2020-odi>), which outlines best practices for working with library discovery services. It defines ways for libraries to assess the level of content provider participation; streamlines the process by which libraries, content providers and discovery service providers work together; defines models for “fair” linking; and suggests usage statistics that should be collected for libraries and for content providers. The recommendations in this document, created by members of the Open Discovery Initiative Standing Committee, enable libraries, discovery service providers, and content providers to work together to the full extent of their abilities - providing the most effective and rich experience to end users.

Keywords: Discovery, ODI, open discovery initiative

1. Overview of ODI

1.1. History

The Open Discovery Initiative (ODI) project grew out of a meeting of interested parties held at the American Library (ALA) Annual 2011, convened by Marshall Breeding (an independent library consultant) and Jenny Walker (then of Ex Libris). At the time, web-scale discovery services - massive databases that indexed both multiple vendors’ abstracting and indexing records as well as the full text of articles from scholarly and other journals - were first coming to market. Serials Solutions’ Summon product (later a division of ProQuest) arrived first, joining OCLC’s FirstSearch and, soon thereafter, by the EBSCO Discovery Service (EDS). As these products were coming to market, librarians and content providers wanted to better understand what data was available in a web-sale discovery index relative to the independent products that comprised it, what metadata was indexed in them, to what extent a library needed to subscribe to both the vendor product and the discovery product to receive benefit of their investments, how linking decisions were made, and much more.

Out of those first conversations grew a NISO Working Group, the Open Discovery Initiative (ODI), to explore these questions and develop a NISO Recommended Practice around web-scale discovery. ODI’s

* Corresponding author: Geoffrey Morse. E-mail: gmorse@northwestern.edu.

members, representing libraries, content providers, and discovery providers, met intensively over three years from 2011–2014 to explore mutual concerns and issues among the three parties and develop a Recommended Practice (RP). This original RP was designed to establish basic information sharing so that all three partners in the discovery process could understand each other, what metadata and full text were expected to be made available through the ecosystem, what was indexed, and what was displayed in the discovery interface. This first RP included such items as affirmations of core/required metadata (basic information elements about an item that included titles, authors, abstract, publication dates, etc.) and additional metadata elements that were recommended to be shared by content providers and indexed by discovery providers, but were not required to be in ODI conformance. The Recommended Practice also noted a number of metadata transfer formats, depending on the type of content, to ease information sharing between content providers and discovery providers.

With publication of the original Recommended Practice in 2014, the original Working Group transitioned to a Standing Committee to oversee the RP and determine when the time was appropriate to work on a new version of the document. By 2017, the discovery ecosystem had evolved to the point where there were new content types in the discovery environment that were not previously well described. New questions arose about how choices were made by discovery providers as to what source would be the default link for full text links and the need for library-focused best practices in configuring and maintaining their discovery systems. With these opportunities in mind, the ODI Standing Committee embarked on what turned out to be a three-year process to update the 2014 Recommended Practice, releasing a revised version in August 2020, after internal discussions, research, and a public comment period.

1.2. Goals and benefits

The Open Discovery Initiative's tagline is "promoting transparency in discovery". That phrase is at the center of the Recommended Practice. ODI's goal is to allow each of the three partners in discovery - libraries, content providers, and discovery providers - to understand what is expected of them, what is available to them, and how data flows from its source to the end user.

From the library perspective, web-scale discovery systems are increasingly important to the work of libraries in service of their users. As these discovery systems become more complex, librarians are less able to understand or explain to their users what content is included or how. The Open Discovery Initiative's recommended practice represents a significant opportunity for libraries to understand what is indexed by their chosen discovery service, the sources from which that content is derived, and how best to search against it. Additionally, the library conformance statement introduced with the 2020 revision provides libraries with a checklist of regular tasks and activities to help ensure that library investments in discovery are managed effectively.

From the content provider perspective, being able to address discovery and library concerns and ensure that their end customers (libraries) are getting the full benefit of the content that they license is important. Especially for smaller content providers, knowing the "normal" paths to providing their content to discovery indexes can save effort and investment. The content provider conformance statement, through its standardization, provides content providers with a straightforward way to communicate with libraries about the metadata and content libraries should expect to see in their discovery environment.

And finally, for discovery providers, being able to affirm to both libraries and content providers how metadata and links to full text are indexed, displayed, and managed in the discovery interface can build trust in the discovery-to-delivery ecosystem.

Ultimately, of course, it is the user of library resources who benefits from the transparency created through the Open Discovery Initiative's conformance statements, though the user is likely unaware of this hidden work behind the scenes.

2. Everyone has a role

2.1. Content provider role

For the library user to receive these benefits, content providers, discovery providers, and libraries need to work together to achieve maximum transparency and usability in discovery systems. Each party has its own role to play in the successful implementation of a discovery layer. Content providers, as mentioned above, can address library and discovery concerns by providing transparency and access to their content that is indexed in a discovery system. The most crucial service to libraries and discovery systems a content provider can undertake is to ensure that there is core metadata for each of the records supplied to the discovery system. For material to be discoverable in the enormous universe of discovery system content, at a minimum, the following core metadata elements need to be included when applicable: *title, author, author identifier, publisher name, volume, issue, page, date(s), item identifier, component of title, component of title identifier, item URL, Open Access Designation, Full text flag, Content type, Content format, Language, Indexing data, and Abstract*. For discovery systems to work effectively, the metadata must include enough description to make the sources discoverable amongst millions and sometimes billions of records.

The availability of core metadata has a direct impact on the discovery experience for users. Several studies have demonstrated the influence of good metadata on search results. In a 2021 study examining how discoverable MARC records are within a discovery system, Woolcott, Payant, Skeen, and Daybell found that certain metadata fields, particularly the contents notes and the summary notes, will often contain the search phrase or search terms that the user has entered in their search [1]. In a 2005 study Gross and Taylor found that controlled vocabulary subject headings were important in the return of search results and that without this metadata users would be 35% less successful in their searching [2]. A more recent study done by Irina Trapido in 2016 found that metadata quality had an impact on the success of the discovery system user and that it is important to "... ensure consistency, granularity, and uniformity of metadata, especially as libraries are increasingly faced with the problem of integrating heterogeneous pools of metadata into a single discovery tool" [3]. The ODI Recommended Practice provides a framework that encourages the necessary consistency, granularity, and uniformity that will ensure that the end user is better able to effectively use the discovery system.

Content providers will often include data points that support direct linking to an article or other resource. Even if these data points exist, the recommended practice stresses that direct linking should not replace OpenURLs, as omitting this link could harm the library's ability to choose the platform to which they direct users or to provide a fallback access method should the direct link not be successful. While there is concern that OpenURLs can be unreliable, it is important to note that with the correct metadata OpenUrl is accurate. In most cases these failures can be traced back to metadata errors [4,5]. To counter these potential issues NISO has produced a recommended practice, *Knowledge Bases and Related Tools (KBART)* [6]. As citations to content will often appear from multiple content providers it is important to librarians to be able to control both the order of links to the platforms and which platforms through which the library user will link to content. This does not preclude direct linking, but direct linking should not take the place of OpenURLs.

A third area of importance to content providers is the indication of whether an item is viewable without any authentication. If an item can be viewed with no authentication it should have “free-to-read” included in the metadata. It is worth noting that Open Access materials are included in the free-to-read specification and that the recommended practice includes a mechanism to include metadata at the record level stating the availability of that content viewed by the end consumer. When this information is present in an item record, discovery platforms will pass it on to the user interface.

2.2. *Discovery provider role*

Above it is stated that discovery providers need to provide clarity on what is included in their indexes, both in terms of the specific journal or other resource titles that are added to the index, and the larger market product bundles (for example, databases that index a range of publications). Libraries and their end users need to understand what is available through the discovery system’s central index so that they can use the discovery layer in the best and most effective manner possible. In 2013, Buck and Steffy noted that many librarians consider an important part of discovery tool instruction to be providing an explanation of how the tool works and what it encompasses [7]. In their 2011 presentation, Howard and Wiebrands noted the propensity of librarians to be concerned with the coverage of specific databases within the discovery layer. A lack of coverage information has a negative impact on a librarian’s ability to communicate to students what was being searched [8].

Furthermore, a key pillar of the *Framework for Information Literacy for Higher Education*, Strategic Exploration, requires that a student be able to evaluate potential information sources. For a user to figure out scope, design and refine search strategies, and identify what parties may produce information about a particular topic, they need to understand not just that the discovery system is an organized system, but that it is not complete and does not contain “everything” [9].

The elements that discovery providers need to be included in a content report include, at the collection level, *Provider, Market product (database name), number of titles in central Index, number of titles in knowledge base, number of unique records in central index, percent of records full text searchable in central index, percent of records from provider that are subject searchable in central index, percent of articles that are free to read, date of last market product update, and date of last report update*. The ability to drill down to a detailed title list is also important. This means fields such as *title, format, and ISSN number* should also be available. The Discovery Provider also needs to consider the following issues:

2.2.1. *Fair linking*

Another key area the recommended practices covers is “fair linking”. This term encompasses the ideal that regardless of any arrangements a discovery provider may have with content providers, that these arrangements do not influence end user results in any way. ODI recognizes that there will be business relationships between these parties, but they should not include requirements for prioritizing or de-emphasizing content, result relevance, or link order. As the library technology industry sees increasing consolidation, it is important that content available through discovery services is equitably managed.

2.2.2. *Metrics and reporting*

With the introduction of discovery layers and central indexes, content providers are losing some control of their metadata. When metadata is in the central discovery index, the content provider no longer has a way to tell how often and who is accessing their content. Furthermore, when users access content through

a discovery provider the usage patterns of individual databases may change. Because of these factors, discovery services should supply metrics that conform to the COUNTER Guidelines [10]. Evelhoch noted that in the three years following a discovery layer implementation at a university with a FTE (Full Time Equivalent) of 12,000, statistics on specific journal and database platforms declined while at the same time the discovery layer showed an increase in usage [11]. COUNTER-conforming metrics can provide content providers with additional usage information and demonstrate that only licensed users are getting access to their content. This will supplement the statistics that content providers are already obtaining from their own platforms. These statistics can not only tell if their content is being found in the discovery layer, but also provide information on the effectiveness of their metadata.

2.2.3. *Open Access*

Content providers need to provide a free-to-read attribute in the metadata for material that is open access. It is equally important that discovery services present this indication of Open Access to the end user in the item record.

2.3. *Librarian's role*

In addition to the responsibilities of the Content and Discovery Providers, there are several areas to which libraries need to pay close attention: configuration, training and conformance.

2.3.1. *Configuration*

It is important that the library follow the discovery provider's configuration guidelines. To do this, the responsible library workers should understand how the system is supposed to work. When the Open Discovery Initiative Standing Committee surveyed libraries, a majority reported that extensive configuration work was required to implement the discovery system. However, the survey results also made clear that in many cases the implementation of the discovery system fell primarily to one staff member. While in some cases the size of the library may prevent multiple staff from being involved in configuration work, if possible, there should be staff assigned to oversee specific aspects of configuration. Having specific staff members assigned to aspects of configuration will ensure that the system is set up correctly. For instance, it is crucial that subscribed content is enabled in discovery and it would make sense if the responsibility for this is assigned to an individual or group.

Most libraries surveyed reported that significant configuration work was required when they implemented their discovery system. Because of this, it is important to document all configuration decisions. A few survey respondents were unable to answer specific configuration questions because the individual who had done the initial work was no longer working at the library and there was no record of why specific decisions were made [12]. Documenting configuration decisions and processes will help current and future staff troubleshoot problems and understand the system as they implement new features.

2.3.2. *Training*

Concomitant to documenting configuration decisions and understanding how the discovery system works is ensuring that staff and patrons are trained to use the system in the most effective manner possible. When establishing a training program for a discovery system one needs to consider that there are multiple user communities. Library public services staff need to know how to best take advantage of search features

and be made aware of any potential problems. Technical services staff need to know best practices for troubleshooting problems as well as having a basic understanding of how the system works, particularly as it affects their areas. Additionally, while discovery systems are designed to be intuitive, there are many features that are not evident to the end user. For libraries to be sure end users are getting the most out of their discovery system, training should be offered. This is particularly true after the implementation of a new system, but it is also needed on an ongoing basis.

2.3.3. *Conformance*

Libraries should also complete and publish a conformance statement. In addition, libraries should follow up on conformance issues with their vendor partners and become advocates for increasing ODI conformance for Content Providers and Discovery Service Providers.

3. Conclusion

By reviewing the recommended practice and completing a conformance statement, a party, whether a content provider, discovery provider, or a library will have contributed to the aspiration enunciated in the NISO ODI tagline: “promoting transparency in discovery”. In addition to following up on the recommendations for their section, all can assist others by making their conformance statements public and pointing out areas where others would benefit by meeting the recommendations. As discovery systems grow in scope it is incumbent on all parties to maintain as much transparency as possible. The NISO Open Discovery Initiative Recommended Practice, when implemented, provides content providers assurance that the end user is getting access to what discovery systems are providing; provides discovery providers assurance that their products are being implemented in the most effective manner; and provides libraries with the assurance that content and discovery providers are working with them to provide the maximum benefit to the end user.

About the authors

Geoffrey Morse is Interim Head of Research Services at the Northwestern University Libraries and is liaison to the departments of History and Religious Studies. Morse has held various roles at Northwestern including supervising the research assistance desk, chairing the Resource Discovery Operations Group and, in his role as a public services librarian, configuring the library’s discovery system Primo and the library services platform ALMA. Morse holds an MLIS from the University of Wisconsin–Madison and a BA in History from Hampshire College. Email: gmorse@northwestern.edu; Phone: (847)-467-1866.

Ken Varnum is Senior Program Manager and Discovery Strategist at the University of Michigan Library. Working within Library Information Technology’s Design and Discovery unit, he is responsible for technology and project management for the library’s discovery interfaces (Library Search, the library’s integrated search interface), link resolver (“MGet It”) and other delivery interfaces, and the library’s evolving and emerging analytics infrastructure. From 2007–2015, he was manager of the Web Systems department at the University of Michigan Library, where he oversaw the library web site and development of new features and functionality. Ken received his Bachelor of Arts from Grinnell College and two masters degrees from the University of Michigan (one from the School of Information and the second from the Center for Russian and East European Studies. Email: varnum@umich.edu.

References

- [1] L. Woolcott et al., Missing the MARC: Utilization of MARC fields in the search process, *Cataloging & Classification Quarterly* **59**(1) (2021), 28–52.
- [2] T. Gross and A.G. Taylor, What have we got to lose? The effect of controlled vocabulary on keyword searching results, *College & Research Libraries* **66**(3) (2005), 212–230.
- [3] I. Trapido, Library discovery products: Discovering user expectations through failure analysis, *Information Technology & Libraries* **35**(3) (2016), 9–26.
- [4] J. Wisniewski, The missing link (resolver), *Online Searcher* **39**(1) (2015), 74–76.
- [5] O. Pesch, The KBART's potential beyond OpenURL linking, *Serials Librarian* **67**(3) (2014), 231–239.
- [6] NISO, KBART: knowledge bases and related tools. A recommended practice of the National Information Standards Organization (NISO) and UKSG, in: *NISO Recommended Practices*, NISO, 2014.
- [7] S. Buck and C. Steffy, Promising practices in instruction of discovery tools, *Communications in Information Literacy* **7**(1) (2013), 66–80.
- [8] D. Howard and C. Wiebrands, Culture Shock: Librarians' Response to Web Scale Search, Australian Library and Information Association (ALIA) Information Online Conference, 2011.
- [9] Association of College and Research Libraries, Framework for Information Literacy for Higher Education, Association of College and Research Libraries, a division of the American Library Association, 2015.
- [10] COUNTER. *About COUNTER*, 2021 [cited 2021 June 4]; Available from: <https://www.projectcounter.org/about/>, accessed July 18, 2021.
- [11] Z. Evelhoch, Where users' find the answer: Discovery layers versus databases, *Journal of Electronic Resources Librarianship* **30**(4) (2018), 205–215.
- [12] NISO, *The NISO Open Discovery Initiative: Promoting Transparency in Discovery*. Baltimore, MD, 2020, p. 65.