Open Discovery Initiative: Promoting Transparency in Discovery

A Recommended Practice of the
National Information Standards Organization

Approved: June 25, 2014

Prepared by the
Open Discovery Initiative Working Group

About NISO Recommended Practices

A NISO Recommended Practice is a recommended "best practice" or "guideline" for methods, materials, or practices in order to give guidance to the user. Such documents usually represent a leading edge, exceptional model, or proven industry practice. All elements of Recommended Practices are discretionary and may be used as stated or modified by the user to meet specific needs.

This recommended practice may be revised or withdrawn at any time. For current information on the status of this publication contact the NISO office or visit the NISO website (www.niso.org).

Contents

Foreword

About This Recommended Practice

The ODI was initiated in June 2011 by Oren Beit-Arie of Ex Libris, Jenny Walker, independent consultant working on behalf of Ex Libris, and Marshall Breeding, a library industry consultant (previously at Vanderbilt University) who, at the 2011 ALA Annual Conference in New Orleans, invited senior industry managers to meet and explore areas of mutual interest related to library discovery services. Following this initial meeting, a proposal was submitted to NISO; later in the year, the NISO Discovery to Delivery Topic Committee accepted the proposal as a new NISO work item. This Recommended Practice is the outcome of that project.

The NISO Open Discovery Initiative (ODI) group started its work in early 2012 to define best practices for the new generation of library discovery services. These services use an aggregated central index to enable searching across a wide range of library related resources—both licensed and free—from multiple providers. They also offer more sophisticated capabilities and faster performance than those provided by systems relying on federated search technologies. Several major discovery products, based on centrally indexed search, have been released to the market since early 2009 and have been widely deployed in libraries globally.[1]

This investment by libraries in centrally indexed search as a primary means for users to discover and access licensed content brings with it new requirements for industry practices in a number of areas. To facilitate improved communications and clarity, the ODI established the following goals:

- Create ways for libraries to assess the level of content providers' participation in discovery services

- Help streamline the process by which content providers work with discovery service vendors

- Define models for fair or unbiased linking from discovery services to publishers' content

- Determine what usage statistics should be collected

Libraries expect their entire collection, including licensed and purchased electronic content, to be made available through the discovery service of their choice. When acquiring licensed content, libraries expect a clear explanation of the degree of availability of that content in that discovery service.

Based on the input from a survey done early in the project (see 2.7), the ODI group agreed to develop recommended practices in the following areas:

1. Technical recommendations for data format and data transfer, including method of delivery and ongoing updates.

2. Recommendations for the communication (automated or through reporting) of libraries' rights to distribute or display specific content (e.g., restricted to subscribers versus open to all users). These recommendations are to include technical specifications on how data will be exchanged and procedural specifications regarding update frequency and other logistical details.

3. Clear descriptors regarding the level of indexing performed for each item or collection of content and the level of availability of the content.

4. Definition of fair linking from discovery service to the published content.

---

[1] Marshall Breeding. "Automation Marketplace 2013: The Rush to Innovate." *The Digital Shift*, April 2, 2013. Available at: www.thedigitalshift.com/2013/04/ils/automation-marketplace-2013-the-rush-to-innovate/

5. Determination of what usage statistics should be collected, for whom, and how these data should be disseminated.

Further, the Working Group agreed to develop mechanisms to evaluate conformance with the Recommended Practice.

To work towards the development of recommended practices to address the above issues, five subgroups were formed as follows:

1. Technical formats

2. Communication of library's rights

3. Level of indexing

4. Fair linking

5. Usage stats

Early in the process subgroups 2 and 3 merged.

A further goal of the ODI was to develop mechanisms to evaluate conformance with the Recommended Practice. To help libraries better understand the position of content providers and discovery service providers, these organizations can take measures to demonstrate the extent to which they conform with the recommended practices issued by the Open Discovery Initiative. We anticipate that these conformance statements will be voluntarily issued by content providers and by discovery service providers respectively. In the absence of voluntary statements, libraries can use the presence or absence of these factors to infer conformance.

NISO D2D Topic Committee Members

The following individuals served on the NISO Discovery to Delivery Topic Committee that had oversight for this work and approved this Recommended Practice:

Kristin Antelman
North Carolina State University Libraries

Pascal Calarco, Co-Chair
University of Waterloo Library

Lucy Harrison, Co-Chair
Florida Virtual Campus

Juli Marsh
The Library Corporation (TLC)

Peter Murray
Lyrasis

Tim Shearer
University of North Carolina Chapel Hill Libraries

Chris Shillum
Reed Elsevier

Christine Stohn
Ex Libris, Inc.

ODI Working Group Members

The following individuals served on the NISO Open Discovery Initiative Working Group that developed and approved this Recommended Practice:

Marshall Breeding, Co-Chair
Independent Consultant
(formerly with Vanderbilt University)

Jenny Walker, Co-Chair
Independent Consultant (formerly affiliated with Ex Libris)

Anya Arnold (through July 2012)
Orbis Cascade Alliance

Scott Bernier (from December 2013)
EBSCO Information Services

Jamene Brooks-Kieffer
University of Kansas Libraries (formerly with Kansas State University Libraries)

Sara Brownmiller
University of Oregon Libraries

Lettie Conrad
SAGE Publications

Michael Gorrell (through November 2013)
EBSCO Information Services

Jeff Lang
Thomson Reuters

John Law
ProQuest

David Lindahl
University of Missouri-Kansas City Libraries
(formerly eXtensible Catalog Project)

Laura Morse
Harvard University

Michele Newberry
Unaffiliated (formerly University of Florida)

Peter Noerr (through June 2012)
MuseGlobal

Roger Schonfeld
Ithaka S+R

Ken Varnum
University of Michigan Library

Aaron Wood
Alexander Street Press

Acknowledgments

The ODI Working Group would like to offer a special thanks to the following individuals for their assistance:

- SAGE Publications, especially Elisabeth Leonard, for support of the ODI Survey.

- Lucy Harrison (Florida Virtual Campus), Bonnie Lawlor (NFAIS), and Jeff Penka (OCLC), who were observers on the Working Group and provided valuable input and feedback.

- The many individuals and organizations that responded to the information-gathering survey issued by the ODI.

- Nettie Lagace and Cynthia Hodgson from NISO for their support and advice.

Trademarks and Service Marks

Wherever used in this recommended practice, all terms that are trademarks or service marks are and remain the property of their respective owners.

<div style="background-color:#2e4a7a; color:white; padding:10px; text-align:center;">

Section 1: Introduction

</div>

1.1 Purpose and Scope

1.1.1 Purpose

The Open Discovery Initiative (ODI) aims to facilitate progress through exploration of relevant issues and the development of recommended practices for the current generation of library discovery services based on centrally indexed search. The domain of index-based discovery services involves a complex ecosystem of interrelating issues and interests among content providers, libraries, and discovery service creators.

This model of discovery relies on an index populated with metadata, full text, or other representations of the content items—such as journal articles, book chapters, e-books, research reports, reference sources, images, maps, datasets, AV materials, and other selected material—which a library provides to its users. The content comes from a range of information providers and products, such as commercial and non-profit publishers, universities and other research institutions, and many other types of organizations. The content of interest to ODI includes any materials that libraries would consider within their collection, regardless of the business model for acquisition or the type of license, such as commercially restricted or open access.

Several major discovery products have been released to the market since early 2009 that are based on the model of centrally indexed search—largely influenced by the Google search model and users' expectations for a single, unified discovery experience. An increasing number of libraries, especially those that serve academic or research institutions, have invested in index-based discovery services. These products serve as one of the interfaces through which the library's patrons gain access to the rapidly growing breadth of information that may be available to them. These discovery services play an increasingly strategic role in the way that libraries provide users with access to their collection, represent a growing segment of the library technology industry, and may become a factor in how libraries select content products. These factors draw attention to the discovery services arena for any improvements that might be gained through this Recommended Practice.

To work effectively, discovery services need to be as comprehensive as possible in their content coverage. Libraries expect their uniquely licensed and purchased electronic content to be indexed within their discovery service of choice. Further, they require comprehensive and clear representation of each category of content in the discovery service. Content items not represented in a discovery service present a challenge to libraries in how they might otherwise ensure that these materials are discovered and accessed. Libraries have an interest in knowing whether any content providers are excluded or underrepresented in any given discovery service.

The Open Discovery Initiative aims to facilitate increased transparency in the content coverage of index-based discovery services and to recommend consistent methods of content exchange or other mechanisms. Full transparency will enable libraries to objectively evaluate discovery services and to deal with daily operational issues surrounding these products.

Discovery services depend on the cooperation of content providers with discovery service creators to provide access to metadata or full text of information resources in order to create effective indexes. The inclusion of data in the indexes of the current slate of discovery services is based on private agreements and ad hoc exchange methodologies between information providers and discovery service creators. Index-based discovery can potentially benefit content providers through enhanced exposure of their materials. It also presents some concerns, such as enabling library patrons to bypass the specialized interfaces created by content providers, potentially reducing or eliminating branding and losing control in how content is presented to the end user. And, as libraries' uptake of these services increases, the usage (and perceived

value) of publisher products can be greatly influenced by how successfully discovery services drive readers to a content providers' assets.

ODI investigated the need for standard protocols to make the transfer of data from content providers to discovery service creators. Consistent practices in the exchange and formats of data aim to lower the level of complexity as content flows through this ecosystem, mitigating technical issues that might hinder broader participation by content providers or potential discovery service creators.

Libraries need a clear understanding of the degree of exposure for the content that they have acquired as represented in a discovery service. This understanding is essential as libraries evaluate and select a discovery service and on an ongoing basis once it is implemented. Libraries require specific information on exactly which articles, databases, and other sources are represented; whether they are indexed in full text, by citations only, or both; and whether the metadata derives from aggregated databases or abstracting and indexing resources.

In the operation of an index-based discovery service, many different factors contribute to how it presents and orders results and how it connects users to content resources. For any given item of content, multiple metadata elements contributed from different content providers may be indexed by the discovery service. For a journal article, for example, its full text might be contributed by the primary publisher, citation data from the provider of an aggregated database, and abstracts or controlled vocabulary terms may be provided by yet another provider. Content providers are motivated to contribute to discovery services in order to gain more access from the patrons associated with the libraries that implement the discovery services. It is therefore important to each type of content provider that its contributions are appropriately recognized. If a record contributed by an abstracting and indexing (A&I) service, for example, leads to the selection of a full-text resource from another provider, how does the A&I service gain benefit from the discovery transaction? A subgroup of ODI on Fair Linking was established to explore and make recommendations on these issues.

The Open Discovery Initiative recognized and aimed to address perceptions regarding bias and concerns about the possibility of bias in discovery services. Special concerns surround the possibility of bias when discovery services are owned by the same corporate parent as content products or services. Concerns also arise through exclusive arrangements or other business relationships made by a discovery service with a content provider that might introduce bias. Some of these recommended practices were developed with the intent of helping discovery services providers mitigate concerns that exist in the community about conflicts of interest and other relationships that introduce bias. By explaining the nature of their business connections with related content providers and third parties alike, and affirming the neutrality of their discovery offerings, these services will be positioned to reassure both libraries and content providers about the nature of their practices.

1.1.2 Scope

In broad terms, ODI focuses primarily on the issues related to the composition of the central index associated with these discovery services and not with the design of user interfaces. The initiative does not seek to intrude into areas of proprietary innovation that distinguish each of the discovery service products.

The arena of index-based discovery spans many different issues, some of which lend themselves to a more open and standard treatment, while others remain in the realm of product development. The Open Discovery Initiative recognizes that even among the issues that might potentially benefit from its attention, some rank at a higher priority and that others may need to be addressed through possible follow-up activities.

<u>In Scope</u>

A primary area of interest for ODI involves the arena of content coverage of the discovery services represented in their central index. The content used to create the indexes of a discovery service comes

from a range of information providers and content products, such as commercial and non-profit publishers, universities and other research institutions, and many other types of organizations. The content of interest to ODI includes any materials that libraries consider within their collection, regardless of the business model for acquisition or the type of license, such as commercially restricted or open access.

This initiative aims to address the following questions in the realm of content coverage:

- The quantity of content that a provider makes available to discovery service providers relative to its total offerings

- The form of that content, such as whether it consists of citation-level metadata or if it also includes full-text representations

- Whether the discovery services operate in a way that results presented to the user do not favor or disfavor items from any given content source or material type

- The specific metadata fields provided within metadata records

- The specific metadata fields indexed in the discovery services

- Whether any controlled vocabularies or ontologies are included

- How abstracting and indexing products relate to discovery services

- How branding of content providers is presented in a discovery service

This initiative aims to address the following technical issues:

- The transfer mechanisms or protocols by which data are delivered from content providers to discovery service creators

- The format in which the records are delivered by content providers to discovery service creators

One area of focus deals with the definitions of the metadata delivered by content providers to discovery services, as well as the data made available to licensed customers. A perceived lack of transparency across these data flows prompted the need to develop definitions of the data points and to clarify what metadata or data elements are made available to which parties and under which conditions. For example, a content provider might allow certain metadata elements to be included in the search index for retrieval purposes, but not allow those elements to be displayed in the final user interface. Conversely, libraries might not have understanding of what elements from which full text or abstracting and indexing services are available and in which circumstances. Some elements might be displayed to authenticated users, and some not, but definitions of these distinctions are sometimes vague, if they are described at all.

Another topic concerns factors related to whether or not a discovery service functions with a bias toward certain databases or content products based on business relationships rather than user needs or library preferences. It was deemed important to propose practices that facilitate the presentation of unbiased links to a user following the execution of a search through a discovery service and in support of this objective to ensure transparency about discovery service practices.

Out of Scope

This initiative does not address issues related to performance or features of the discovery services, as these are inherently business and design decisions guided by competitive market forces.

Aspects of index-based discovery not deemed within the scope of ODI include:

- **User interface issues** – The user interface components of a discovery service may depend on the content indexes, but are out of scope for this ODI initiative.

- **Relevancy ranking** – The specific methods that a discovery service employs to order items in a search result set may fall within the realm of proprietary technologies used competitively to differentiate commercial offerings.

  Further, the demands of relevancy ranking—necessary for satisfying user expectations in keeping pace with open web search applications—require continuous enhancement of supporting technologies and algorithms. As such it would be both impractical and an impediment to require that service providers continuously publish the highly dynamic and substantially detailed workings of their search relevancy algorithms. Therefore, the ODI Working Group concludes that, while disclosure of the broad aspects of a relevance ranking algorithm is encouraged, there should be no expectation that a discovery service provider would explain in any level of detail the ranking algorithms it applies. (As noted above, however, whether rankings result in the bias of results presented to the end user is within scope.)

- **APIs exposed by discovery services** – Several discovery service providers offer access to their discovery index through an Application Program Interface (API). This is a set of protocols that allows a computer program to query the index and receive search results. Some libraries build their own interfaces to the search index; others use the API to embed data retrieved from the search index into other applications, like pull lists of citations into course sites or other web pages.

  The ODI Working Group determined that APIs were out of scope for this initial foundational stage of the initiative. This decision was made somewhat reluctantly based on the growing need for libraries to have access to data via API. Because the provision of an API to customers is largely a business decision by the vendor, it was felt the existence of an API could not be suggested as a best practice. Additionally, the desire to recommend standardization of API results would have added significant complexity to the workgroup's recommendations. Thus, the current committee concluded that best practices concerning discovery service APIs should be deferred until a later round of work. See Section 4.2 for recommended next steps.

- **Fair linking** – While the concept of "fair linking" was identified as a highly shared value among all stakeholders, it was determined that fully qualifying how links are presented and ordered fairly is beyond practical scope and intrudes into areas that stakeholders consider to be of a proprietary nature.

- **Protocols for data exchange** – In the arena of the technical mechanisms involved in the transfer of data between content providers and discovery services, the ODI concludes that the existing protocols and methodologies previously defined provide adequate options and that it is not necessary to create a new protocol specifically for use within the discovery services ecosystem. There has been much work and standards development in the area of file formats, schemas, naming practices, transport mechanisms, etc.

1.2 Terms and Definitions

In order to ensure consistent use of terminology, one of the early tasks of ODI was to create a glossary of definitions relevant to the field of investigation. The terms as used in this recommended practice, have the meanings indicated below.

Term	Definition
Search Models and Related Concepts	
central index	The result of storing and indexing content in a central location. Disparate content sources are aggregated with consistent formatting, indexing, and ranking algorithms. Also referred to as *centrally indexed search engine*.
federated search	Method for searching multiple disparate content sources with one query. Federation of searching involves restructuring queries to appropriate forms and sending them to multiple remote search engines. Results are coordinated and displayed to the user.
metasearch	Method for searching multiple disparate content sources with one query. Metasearch is sometimes synonymous with federated search; however, metasearch can also describe searching across disparate content sources stored at a single location.
full-text search	Method for searching the text of the works rather than metadata about the work.
Data Definitions	
full text	The complete text of the work. The text may be in display format or be a simple character representation of the complete text of the work.
index only	Limitation of rights of the search service or aggregator to only allow indexing of portions of the content with limited or no rights to display content to users. Most often applied to full text requiring the user to be directed to another source for display capabilities.
snippet	A small portion of the text of the work, often including text immediately around terms matching the query. Used for display of search results to give the end user context of the retrieved result.
basic metadata	Metadata excluding textual descriptions extracted directly from the work including abstracts, tables of contents, and sample pages.
enhanced metadata	Metadata that is augmented by textual descriptions extracted directly from the work including some or all of abstracts, tables of contents, and sample pages
Methods of Data Exchange	
harvest	Method of extracting indexing and/or full text from remote web-accessible sites for the purpose of providing search and/or display from a different location.
syndication	Method of pushing content to remote indexing, abstracting, or display services.
RSS (rich site summary)	Standard method for advertising the availability of frequently published content that includes metadata, publication date, and authorship information.
AtomPub (Atom publishing protocol)	Standard method for publishing syndication feeds similar to RSS. AtomPub provides a more robust method than RSS for publishing beyond blog and webpage entries.

Term	Definition
screen scraping	Simplest method of harvesting content that places no technical burden on the publishing site. The human-readable text is extracted and indexed, formatted, and searched by various engines.
linked data	Method for publishing structured data intended for further processing by machines rather than directly by human readers.
API access (application programming interface access)	Generic description of a method for exposing services and data through an advertised programming interface. Specification of the programming interface through either a proprietary or published standard is required.

Actors

Term	Definition
end user	The final consumer in an information retrieval session.
licensee	The institution or individual who has acquired rights to access others' content or services.
licensor	The rights holder granting search and/or access rights to others. Also referred to as *rights owner*.
content provider	The organization providing dissemination of the content (literature or information). May be a publisher, aggregator, OA or institutional repository, or A&I service provider. The same content may be available from multiple content providers.
publisher	A person or organization whose predominant activity is to commission, create, collect, validate, and edit information in printed and/or in electronic form. A publisher may also act as a content provider for its material.
aggregator	The organization that collects information from varied sources and provides consistent search, presentation, and/or access.
Open Access Repository/Institutional Repository	The entity that collects and disseminates content created during the research process at academic institutions.
A&I service provider	The organization that provides abstracting and indexing services, including controlled vocabularies, for published content.

License Terms

Term	Definition
authenticated access	Access to search, metadata display, or content display via a known method for verifying end user identity or institution affiliation.
subscription	Provision of access to search, metadata, or content through specified license terms for a period of time, usually annually.
mutual subscriber	A collaboration between two service organizations agreeing to provide equivalent access terms to a single individual or institution.
market product	A specific, defined collection of resources made available for license or use by a content provider to a library. This could be a citation database, a set of e-journals or e-books, or other constellation of titles or objects commonly licensed as a single entity.

1.3 Stakeholders

The Open Discovery Initiative worked to include participation and input from each of the categories of organizations involved in discovery services. These organizations were considered in three stakeholder groups:

- **Content Providers (CPs)** – These organizations offer content products or services, primarily intended for access by library patrons or the general public. The content provided by these organizations is used to generate the central indexes associated with the discovery services. Content providers include commercial and non-profit organizations. Many different access and license models may apply, including those restricted to individuals affiliated with subscribing organizations and those based on open access licenses.

- **Discovery Service Providers (DSPs)** – These organizations create index-based discovery services intended to enable end users to search the broad universe of content made available through their library.

- **Libraries** – These organizations—which may be affiliated with universities, research institutes, or commercial firms—acquire content from a variety of content providers and may also implement an index-based discovery service.

Section 2: The Evolution of Discovery and Delivery

Libraries have relied on different tools and technologies in recent decades to provide their users the ability to search for materials within their collection and gain access to items of interest. The successful discovery and delivery of resources has been a key library mission from the time when collections were primarily composed of print materials until the current time when library collections include massive amounts of electronic content. A succession of products has addressed these tasks in response to evolving demands for discovery and retrieval of library materials.

2.1   Catalogs and Indexes

During the time when library collections were composed primarily of print materials, the online catalog module of the integrated library system was the primary tool for patrons to search for books and journals. A variety of indexes was published in print to find journal articles. The online catalog provides search and request features for the content managed by the integrated library system (ILS), which is broadly the content the library owns and holds locally, but does not address materials managed through other platforms. It also does not address the most granular view of resources. Journal titles, including the years and issues received by the library may be available, but not the articles contained within each issue. Libraries acquire indexing and abstracting services, individual e-journal subscriptions, and aggregated databases of articles, each with their own search interface, to provide access at the article level. Working through each of the possible electronic resources that might have material for a given research topic was often unwieldy for library patrons, which led to the creation of utilities able to search across these resources simultaneously.

2.2   Abstracting and Indexing Resources

Abstracting and Indexing (A&I) services form an important component of the development of resource discovery tools and the current index-based discovery ecosystem. In their original form, abstracting and indexing services produced printed indexes that allowed researchers to find articles according to topics of interest. These index and abstracting products predate online technologies. H.W. Wilson's *Reader's Guide to Periodical Literature*, for example, began publication in the early 1900s.

The production of abstracting and indexing services depends on manual efforts, usually performed by individuals with specific knowledge in a given domain, to compose summaries or abstracts of each article and to assign appropriate subject terms. These cover general categories of content, but many address narrow scholarly or professional disciplines, with specialized thesauri, indexing techniques, or other value-added methods. The production of these resources comes at significant business costs, but is a service that researchers value.

A&I products today are delivered primarily though web-based platforms, some offered directly by the organization that produces the resource and others by organizations that license and aggregate multiple resources into products that span broader areas of subject coverage. Aggregators may also license from the primary publishers the full text of the articles covered by the A&I products, allowing them to provide a comprehensive service that includes both search and full-text delivery within their interface. These products may also use proprietary or OpenURL linking mechanisms to link from citations to the full text of articles not embedded within their products. An A&I product functions much like a discovery service, though operating within a narrow scope of content.

Libraries function as the primary distributors of these A&I products, subscribing to the resources that best match the requirements of their researchers. Libraries also subscribe to the e-journals referenced by the A&I products, either directly from the primary publishers or as part of aggregated products. To enable

coverage of all of the disciplines represented in their collections of articles, libraries subscribe to multiple A&I products. Library users, in turn, need to understand which of these products that they need to use to complete their literature search. To a very large extent, index-based discovery services emerged to simplify the research process relative to the complex process of identifying and using many different individual content products, each with its own interface and search procedure.

A&I services have a complex set of issues relative to the ecosystem of index-based discovery. The producers of these services naturally have an interest in preserving their value, especially in being assured that libraries will continue to maintain their subscriptions should they contribute to discovery services. Decisions regarding whether to participate in discovery services are not straightforward. Discovery services not tuned to make use of the specialized vocabularies, abstracts, and other mechanisms inherent in the native A&I product may underexpose the resource. Aggregators that license A&I content and full-text resources from other providers may not have the rights to further distribute that content. Discovery services must limit access to proprietary content, such as abstracts and specialized vocabularies to authenticated users affiliated with mutually subscribing institutions. Given these factors, among many others, A&I resources must be treated with special consideration within the body of content providers that potentially contribute to discovery services.

2.3   Metasearch Utilities

One of the techniques that emerged in the early 2000s to provide simultaneous search across a library's electronic resources consisted of a utility that would prompt the researcher for a query and then transmit that query to multiple targets. The utility would accept the results returned by each of the resource targets and present them to the researcher. How they might be presented would vary according to the tool's capabilities and configuration and might include grouping by source or interfiling and ranking results. This approach, called metasearch or sometimes federated search, comes with some inherent limitations, such as the number of live simultaneous sessions that could be maintained with resource targets. One technique implemented to address this limitation involved the creation of disciplinary sets that would be selected by the user in order to direct the query to the best set of targets for the topic at hand.

On a technical level, metasearch initially worked through communication sessions with each target that essentially emulated the search session presented by a user. This approach required parsing the HTML delivered by each target into a structured form that could then be handled by the metasearch utility for sorting, presentation, and other tasks. This method of HTML parsing proved to be fragile, with any change made by a resource provider requiring updates to the parsing algorithms. Specialized connectors were developed, with some organizations specializing in creating and maintaining connectors on behalf of metasearch providers. Some content targets offered Z39.50 or SRU/SRW servers, enabling metasearch products to make use of these well-established library standard protocols. Other techniques such as XML gateways were also developed for some resource targets designed to respond to metasearch requests with structured data and with server processes insulated from production instances of their content products. Metasearch, though a pragmatic approach, had inherent performance limitations due to the multiple communications streams and processing performed in real time in response to user queries. The slow performance, limited number of content targets that could be included in a query, and the limited number of results returned were factors that impacted the success and satisfaction with metasearch.

The NISO Metasearch Initiative was launched in May 2003 to define standards and recommended practices in support of this search model.

2.4   Enhanced Library Catalogs

The online catalogs provided with integrated library systems (ILS) were often not easily understood by library users, both in terms of operating the interface and in the scope of content addressed. In contrast,

search engines such as Google were used effortlessly and helped to establish much higher expectations for the simple and comprehensive nature of any search facility offered by libraries. Enhanced library catalogs, or discovery interfaces, that included characteristics such as improved user interfaces, simplified and advanced query options, results ordered according to relevancy rankings, alternative query recommendations, faceted navigation, and enriched content such as cover art, summaries, or tables of contents emerged in the mid-2000s. Most of these discovery interfaces were designed to work with any of the major ILS products and were based on local indexes built from the contents of the ILS, as well as other local content repositories. The scope of these products could be expanded to include article-level content through the use of optional metasearch components.

2.5 Index-Based Discovery Services

Index-based discovery services, initially introduced in 2009, aim to provide a search interface that addresses a very broad representation of a library's collection through a single search box accessing an index populated with the article-level metadata and full-text content offered by the library—whether in print or via other resources managed in the local automation environment. The creators of these index-based discovery services make arrangements with library-oriented content providers to receive citation and/or full-text metadata of their resources. The indexes are constructed to include as much of the potential content as possible to which libraries might subscribe or otherwise add to their online collections. The discovery service also includes a mechanism to profile the content according to the subscriptions associated with any given library implementing that discovery service. Metadata from the library's local integrated library system or other local repositories may also be included in the index and periodically synchronized. Library patrons who use the discovery service are presented with lists of results that match their query. When the user selects an entry and requests to view the associated content item, the discovery service links to the content provider's platform/server, potentially via the library's OpenURL link resolver. Items selected from the library's print materials, managed through the ILS, are either presented as links to the corresponding entry in the local online catalog or to the appropriate service options integrated into the discovery service interface.

This model of index-based discovery generally bypasses the native interface of the content providers' services, but delivers library users to those same content items for viewing, downloading, printing, or other available access options. The purpose of these discovery services is not to re-publish materials represented in the index, but to provide an additional channel for connecting library users with that content.

This model of index-based search offers advantages over previous approaches implemented in libraries. By indexing content in advance, discovery services have the ability to deliver more sophisticated services with instant performance, compared to the metasearch techniques used previously.

2.6 Related Initiatives

A number of initiatives produced recommendations related to the use of discovery services. The most relevant of these to the ODI work are described below. Links to the projects and reports described are available in the Bibliography.

COUNTER (www.projectcounter.org/)

> COUNTER (**C**ounting **O**nline **U**sage of **N**etworked **E**lectronic **R**esources) is an international initiative serving librarians, publishers, and intermediaries by setting standards that facilitate the recording and reporting of online usage statistics in a consistent, credible, and compatible way.

> COUNTER published its latest *Code of Practice*—Release 4—in April 2012 and COUNTER-compliant content providers must comply with this new release by the end of 2013. A significant

number of content providers offer their library customers COUNTER-compliant usage reports. (For the full list of prescribed and recommended reports see Table 1 (page 5-6) in the *Code of Practice for e-Resources: Release 4*.) Index-based discovery services may or may not directly embed full-text content, and therefore a number of the prescribed reports that deal with full text are not relevant.

COUNTER Release 3 specifically addressed the needs of metasearch services but these services are now rapidly being replaced by index-based discovery services. Limited interaction with a member of the COUNTER board and the COUNTER Director indicated that COUNTER Release 4 did not specifically address usage related to index-based discovery services. However, it did add some new metrics to the standard COUNTER Database Report 1 for the "Number of Searches" metric that might be relevant:

- **Result Clicks** –Records the number of times a user clicks on any link in a particular result. Note that if a user clicks on a number of links for a particular record then this would only count as one result click. The intent of this metric is to record that the user has been sufficiently interested in this result to investigate further.

- **Record Views** –A subset of Result Clicks that counts only those times that the user asked to see the Detailed Record.

COUNTER Code of Practice for Usage Factors (www.projectcounter.org/usage_factor.html)

"The overall aim of the Usage Factor project is to explore how online journal usage statistics might form the basis of a new measure of journal impact and quality, the Usage Factor (UF)." Existing assessment metrics, such as the Journal Impact Factor, are generally focused at the journal level. The Usage Factor utilizes the data in the COUNTER *Code of Practice for E-Resources* reports. Usage of discovery services that results in retrieval of full-text articles will feed into the data for calculating the usage factor.

Digital Library Federation ILS Discovery Interface Task Group (ILS-DI)
(old.diglib.org/architectures/ilsdi/)

"In 2007-2008, the DLF convened a Task Group to recommend standard interfaces for integrating the data and services of the Integrated Library System (ILS) with new applications supporting user discovery." The group's technical recommendations define an API for "effective interoperation between integrated library systems and external discovery applications."

International Coalition of Library Consortia (ICOLC) (icolc.net/)

From their *Revised Statement on the Global Economic Crisis and Its Impact on Consortial Licenses*: "Principle 3 (added June 2010). We encourage publishers to allow their content to be made available through numerous vendors appropriate for their subject matter. We also encourage online providers and aggregators to allow their metadata to be included in emerging discovery layer services on a non-exclusive basis. Multiple access platforms will permit libraries and consortia to select content and discovery tools that are suitable and affordable for their constituents. We encourage vendors to provide options that match the range of needs that libraries have for any particular content as to degree of importance, currency, interfaces, access, archiving, preservation, and metadata. It is in the common interest of publishers, database vendors, consortia, libraries, and information consumers to work collectively to provide affordable access to licensed content, while preserving the businesses integral to our collective success."

JISC Discovery Programme (discovery.ac.uk/)

"The JISC-funded Discovery programme was launched in May 2011 to create 'a metadata ecology' to support better access to vital collections data in libraries, archives and museums and facilitate new services for UK education and research." The project work included the publication of *Discovery*

Open Metadata Principles and invited stakeholders to sign up in support of the principles and to enact them.

KBART–Knowledge Base And Related Tools (www.niso.org/workrooms/kbart)

The NISO and UKSG joint KBART project was initiated to "develop and publish guidelines for best practice to effect smoother interaction between members of the knowledge base supply chain." The initial recommended practice (NISO RP-9-2010) provided best practices for formatting and distributing title lists by content providers to knowledge base developers to improve the accessibility of resources obtained through the use of OpenURL link resolvers. A revised recommended practice is now available which expands the original recommendations and focuses on the more granular, complex issues that cause problems in metadata supply. Since the metadata supplied for knowledge bases is often the same as that supplied to indexed-discovery providers, implementation of the KBART recommendations should also improve the data used in discovery services.

Music Discovery Requirements (committees.musiclibraryassoc.org/ETSC/MDR)

"The Music Discovery Requirements document addresses the unique needs posed by music materials which must be considered for successful discovery. The document was created under the auspices of the Music Library Association's Emerging Technologies and Services Committee and officially approved by the Music Library Association Board of Directors. […] This document discusses the issues and when possible gives concrete recommendations for discovery interfaces. […] Three appendixes compile technical details of the specific indexing recommendations in spreadsheets."

NFAIS Recommended Practices: Discovery Services
(info.nfais.org/info/Recommended_Practices_Final_Aug_2013.pdf)

The National Federation of Advanced Information Services (NFAIS™) "members agreed that it would serve the community to develop a code of practice that would outline the rights and obligations of all participants in a discovery service relationship in order to move forward with an understanding of mutual expectations." The recommended practice is intended to assist their Content Owner members who contribute their content to discovery services by providing guidelines that will help maintain an equitable balance of the interests of all participants in the information distribution process. The NFAIS Recommended Practice was developed to be "the ideal towards which the Information Community should strive as technology advances, even if today's technology presents barriers to full implementation."

2.7 Information Gathering and Results

An important component of the ODI work involved gathering information from the key stakeholders: libraries, content providers, and library discovery service providers. Although the NISO ODI membership represented all three of these key stakeholder groups, the Working Group conducted a wider survey of all three constituencies; survey results helped each ODI sub-group prioritize its work and informed these recommendations.

Each of the ODI sub-groups recommended suitable questions for a survey, following a series of interviews used to assess the most pressing needs of the stakeholders. The full survey was announced on September 11, 2012, and closed on October 4, 2012. The survey questions and results can be found in Appendix A.

The merged Communication of Library's Rights and Level of Indexing sub-groups compiled a set of data elements useful for all parties in understanding what data are ingested into discovery indexes, what data are used to generate search results, what data are displayed on the screen, and what uses discovery service consumers might make of these elements. These elements set the stage for determining the subgroup's survey questions for its stakeholders (libraries, content providers, and discovery service providers).

During the process of reviewing the survey results, it was clear that some of the questions posed by the merged Communication of Library's Rights and Level of Indexing Working Group caused confusion for many respondents. In particular, responses from content providers were considered insufficient for informing best practices and led to a follow-on survey for content providers, to both validate the initial survey responses and further explore potential barriers to content provider participation in discovery services. The major barriers related largely to business modeling concerns (how to continue to prove their value as a service within the larger discovery environment), acknowledgement for their content/services within the discovery interface via appropriate branding, lack of metrics related to use of provided content, "unfair" linking, and concerns about lack of advance options for their content in a discovery system. Many comments throughout the research revealed a need for more transparency and educational dialog between discovery providers, content providers, and libraries.

Section 3: Recommendations

3.1 General Recommendations

1. Create an oversight group (Standing Committee or Working Group) to promote educational opportunities about adoption of these recommended practices, provide support for content providers and discovery providers during adoption, provide a forum for ongoing discussion related to all aspects of discovery platforms for all stakeholders (content providers, discovery providers, libraries), and determine timing for next steps for ongoing work.

2. Content providers and discovery service providers can take specific measures to assert their conformance with the recommended practices of the ODI. These measures will be voluntarily made by the organizations. Self-check conformance lists are provided in Appendix B and Appendix C. In this phase, conformance will not be assessed by NISO or any other formal organization, but through general community review.

3.2 Recommended Best Practices for Content Providers

3.2.1 Metadata Elements Provided to Discovery Services

The trend toward index-based discovery requires cooperation between content providers, discovery service providers, and libraries to ensure that the broadest spectrum of materials can be fully exposed through discovery platforms.

The elements listed in 3.2.1.2 and 3.2.1.3 are based on the KBART (NISO RP-9-2014) metadata encoding schema—which is intended primarily for serial and monograph titles to improve A-Z lists, link resolution, and OpenURL—but has been extended here to capture elements for subject, abstract/description, content type, and content format. This extension is intended to serve as a minimum basis from which all types of content may be described.

3.2.1.1 General Requirements

1. Content providers should make available to discovery service providers core metadata, and underlying full-text/original content for complete offerings, for the purposes of indexing to meet licensed customers' and authenticated end users' needs.

2. To this aim, all content providers should make available to discovery service providers, at a minimum, the core set of metadata elements (see 3.2.1.2) for each item they submit for indexing.

3. Content providers should provide the content item (full text, transcript, etc.) and additional descriptive content (abstract/description and controlled and/or uncontrolled keywords) (see 3.2.1.3), for as much of their content as possible.

3.2.1.2 Core Metadata

The fields shown in Table 1 collectively constitute the minimum metadata record that must be provided by Content Providers (CPs) to Discovery Service Providers (DSPs) for each item provided to the DSP for indexing. Fields must be provided where they are applicable to the content type. Examples are provided for each metadata element in Table 2.

NOTE: It is recognized that many content providers merge Content Type and Content Format in their systems. Providing separate fields for this data is preferred, but the current practice of a single field may continue if separating the data is too burdensome.

Table 1: Core metadata elements to be provided by content providers

Field Name	Definitions
Title	The main title of the item.
Authors	The author(s) of the item. Individual authors should be listed in lastname, firstname order.
Publisher Name	The name of the publisher of the item.
Volume	Volume number of the resource, where applicable.
Issue	Issue number of the resource, where applicable.
Page(s)	Page numbers of the resource, where applicable.
Date/Date Range	The date of publication. For a serial run, coverage dates included for the serial.
Item Identifier	One or more standard identifiers for the print or online version of the item (e.g., ISSN, OCLC number, ISBN, DOI, etc.). The identifier should be preceded by a label indicating the type of identifier.
Component Of Title	Describes the publication or serial of which the individual item is a part (e.g., for journal articles, the serial title; for tracks on a CD, the album title; etc.).
Component Of Title Identifier	Provides a standard identifier for the component title defined above (e.g., ISSN, OCLC number, ISBN, DOI, etc.). The identifier should be preceded by a label indicating the type of identifier.
Item URL	Either an OpenURL or a direct link for the specific item's full text.
Open Access Designation	To comply with the NISO Open Access Metadata and Indicators (OAMI) group's recommendations, if an item is open access, this status should be indicated with "free_to_read" and otherwise left blank. See www.niso.org/workrooms/oami/.
Full Text Flag	A yes/no statement describing whether the content provider makes this item available in full text (or for non-print media, a full-length or high-resolution version) to the DSP for the purpose of indexing. It is expected that this will be disclosed by DSPs to libraries in future when describing indexing coverage for a title or collection.
Content Type*	Intended to be used to identify whether the content being described is textual, a visual recording, a sound recording, etc. The textual descriptors from the controlled list established in the MARC 21 Type of Record position (06) of the Leader field is recommended to be used for this field's content.
Content Format*	Intended to be used to indicate whether the nature of the content being described is monographic, serial, a component part, collection, etc. The textual descriptors from the controlled list established in the MARC 21 Bibliographic Level position (07) of the Leader field is recommended to be used for this field's content.

* It is recognized that many content providers merge Content Type and Content Format in their systems. Providing separate fields for this data is preferred, but the current practice of a single field may continue if separating the data is too burdensome.

Table 2: Examples of core metadata elements provided by content providers

Field Name	Monograph & Serials	Articles/ Documents	Manuscripts/ Unpublished Content	Media Content Image	Media Content Recording	Media Content Motion Image
Title	Nature	Derivation of pluripotent stem cells horn cultured human primordial germ cells	The evolution of cooperative signal comprehension in the domestic dog (Canis familiaris)	Portrait of Francoise en costume tailleur	Panama	The Scent of Green Papaya
Authors		Shamblott, MJ; Axelman, J; Wang, SP; Bugg, EM; Littlefield, J; Donovan, PJ; Blumenthal, PD; Huggins, GR; Gearhart, JD	Wobber, Victoria Elizabeth	Picasso, Pablo	Armstrong, Louis	Tran, Anh Dung ; Lu, Man San; Truong, Thi Loc ; Yen-Khe, Tran Nu ; Delhomme, Benoit ; Abril, Jean-Philippe ; Dedieu, Nicole ; Roques, Jean-Pierre ; Rossignon, Christophe ; Negre, Alain ; Ton-That, Tiet
Publisher Name	Macmillan Journals Ltd.	New York Academy of Sciences			Future Noise Music Ltd	Kino International
Volume		95				
Issue		23				
Page(s)		13726				
Date/Date Range (first issue for serials)	19950105-	19981110	20060929	1946	2008	1993
Item Identifier	ISSN 0028-0836 eISSN 1476-4687	DOI 10.1073/pnas.95.23.13726	OCLC 157003171		ISRC GBSUW0716131	LCCN 97505399

Field Name	Monograph & Serials	Articles/ Documents	Manuscripts/ Unpublished Content	Media Content Image	Media Content Recording	Media Content Motion Image
Component Of Title	N/A	Annals of the New York Academy of Sciences.	N/A	N/A	Classic Song Book, Vol. 4	N/A
Component Of Title Identifier	N/A	0077-8923; 749-6632	N/A	N/A	UPC 884385621520	N/A
Item URL	http://www.nature.com /nature/index.html	http://linkresolver/?pid =username:password& id=doi: 10.1073/pnas.95.23.13 726&noredirect=true		http://www.bridgeman education.com/ImageV iew.aspx?balid=16684 1	http://jazz.alexanderstree t.com/View/1030219	http://feat.alexanderstree t.com. View/1825104
Open Access Designation	N	N	free_to_read	N		N
Full Text Flag	N	N	N	N		N
Content Type	Book, Language Material	Article, Journal Article	Manuscript	Image, DVD, Movie	Musical sound recording	Projected medium
Content Format					Monographic component part	Monograph

3.2.1.3 Enriched Content

The elements in Table 3 may be provided by CPs to DSPs in total or in part for each item provided to a DSP for indexing. Inclusion of enriched content in indexes and as used for relevancy ranking greatly improves the discovery experience for users; it brings particular benefit to librarians and advanced researchers who are accustomed to controlled vocabularies. Examples are provided for each metadata element in Table 4.

Table 3: Enriched content to be provided by content providers

Field Name	Definitions
Indexing data	One or more keywords (from controlled or uncontrolled vocabularies) to describe the content of the item.
Full Text/Transcript	For text items, the entirety of the document. For audio or video content, a full transcript of the spoken content of the material. May not be relevant for all indexed content.
Abstract/ Description	Either a text summary on the content or (for non-text materials) a description of the item.

Table 4: Examples of enriched content elements provided by content providers

Field Name	Monograph & Serials	Articles/ Documents	Manuscripts/ Unpublished Content	Media Content Image	Media Content Recording	Media Content Motion Image
Indexing Data	Science	Stem Cells; Regenerative Medicine; Mesenchymal Stem Cells; Differentiation; Therapy	Human-animal communication. Domestication	Female; Picasso's Mistress, Gilot, Drawing, Lover	Music & Performing Arts--Jazz	Music & Performing Arts—Film Romantic Relationships
Full Text/Transcript	[full text]	[full text]	[full text]		[transcript]	[transcript]
Abstract/ Description		Aging is associated with a progressive failing of tissues…				An Academy Award nominee for Best Foreign Language Film, Tran Anh Hung's "luxuriant, visually seductive debut" (New York Times) recreates antebellum Vietnam through both the wide eyes of childhood and the deep blush of first love. In 1951 Saigon, 10 year old Mui (Lu Man San) enters household service for an affluent but troubled Vietnamese family. Despite her servile role, Mui discovers beauty and epiphany in the lush physical details that envelope her, while earning the fragile affection of the household's grieving matriarch. As she comes of age, the now grown Mui (Tran Nu Yen-Khe) finds her relationship with a handsome pianist she has admired since childhood growing in depth and complexity. Though steeped in writer-director Tran Anh Hung's southeast Asian heritage, The Scent of Green Papaya was realized entirely within a Parisian soundstage. The film's heady, scrupulously detailed and wholly authentic depiction of a

Field Name	Monograph & Serials	Articles/ Documents	Manuscripts/ Unpublished Content	Media Content Image	Media Content Recording	Media Content Motion Image
						society in decline, a family in quiet turmoil, and lovers on the threshold of romance earned the Camera D'Or at the Cannes Film Festival. A timeless evocation of life's universal enchantment and a powerful portrait of a vanished world, The Scent of Green Papaya is "a film to cherish." (Roger Ebert).

3.2.2 Metadata Elements Provided by Content Providers to Libraries

The ODI recommends that content providers disclose their level of participation in discovery services to library subscribers. For each market product (journal collection, A&I database, etc.), content providers should disclose the coverage depth and content depth provided, as described in Table 5.

Table 5: Summary of metadata elements to be provided by content providers to libraries

For each Market Product Included	Disclose the Following			
	Coverage	Content Provided to DSP for Indexing	Content Provider Type	Content Provided To
For example: • Journal Collection • E-book Collection • Image Collection • Audio Collection • Video Collection • Database • Data sets	For journals: • One or more date range(s) For other products: • All OR • Part (describe scope of part)	List all that apply to content provided: • Metadata • Indexing data • Abstract/ Description • Full text	The type of agent the content provider acts as for the Market Product. Options are: • Publisher • Aggregator • A& I service • Open Access Repository/ Institutional Repository For full text items, this is generally the publisher, aggregator, or open access repository/ institutional repository. For A&I-provided items, this is generally the A&I service.	Discovery Services the content is made available to. Options are: • Available to all DSP agents – when content is made available for harvest by any interested party • Available to DSPs by agreement [provide list of DSPs under agreement]

3.2.3 Disclosure

We recognize that some discovery service providers and/or content providers may wish to incorporate non-disclosure agreements (NDAs) and similar provisions into the contracts and other agreements that they enter into with one another. But NDAs should not be used to avoid the disclosures and transparency recommended in Section 3.3.2, and elsewhere in this document.

3.2.4 Technical Formats

The ecosystem of index-based library discovery includes the transmission of data from content providers to discovery service providers for the purpose of building a central index. Although the number of current providers of index-based discovery services remains fairly limited, the number of potential content providers is immense. Given the vast amounts of data being exchanged, it is in the interest of all stakeholders to employ the most efficient transfer mechanisms available.

Section 3.3.3 describes the major metadata encoding schemas, file format conventions, and methods of transfer currently in use within the discovery services ecosystem.

The ODI recommends that the transfer of data from content providers to discovery service providers should make use of existing standards where applicable. Some of the standards and protocols most directly applicable include the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and KBART. ResourceSync also has strong potential as a mechanism for data transfer for discovery services once it is published by NISO and software tools become available for its implementation.

3.3 Best Practices for Discovery Service Providers

3.3.1 Discovery Service Content Listings

Discovery service providers should make available to prospective and current customers sufficient information about the content of their repositories to ensure an adequate evaluation of that content against the customers' needs.

3.3.1.1 Metadata Elements

While there is a need for full description of metadata at levels from the title to the item, ODI recommends an initial phase of implementation for title-level metadata. The following metadata (where relevant for each content type) should be made available:

1. Publication title

2. Dates of coverage (journals) or date of publication

3. Standard identifier(s), e.g., ISSN, ISBN

4. Content format and content type

5. Depth of coverage, as indicated by content indexed

Table 6: Summary of metadata elements to be provided by discovery service providers to libraries

Field Name	Definitions
Publication Title	The main title of the item.
Date/Date Range	The date of publication. For a serial run, coverage dates included for the serial.
Identifier	One or more standard identifiers for the print or online version of the item (e.g., ISSN, OCLC number, ISBN, DOI, etc.). The identifier should be preceded by a label indicating the type of identifier.
Content Format*	Intended to be used to indicate whether the nature of the content being described is monographic, serial, a component part, collection, etc. The textual descriptors from the controlled list established in the MARC 21 Bibliographic Level position (07) of the Leader field is recommended to be used for this field's content.
Content Type*	Intended to be used to identify whether the content being described is textual, a visual recording, a sound recording, etc. The textual descriptors from the controlled list established in the MARC 21 Type of Record position (06) of the Leader field is recommended to be used for this field's content.

Field Name	Definitions
Content Indexed*	List all that apply to content indexed: • Metadata • Indexing data • Abstract/Description • Full Text

* It is recognized that many content providers merge Content Type and Content Format in their systems. Providing separate fields for this data is preferred, but the current practice of a single field may continue if separating the data is too burdensome.

The intention is for the Content Format field to be used to indicate whether the nature of the content being described is monographic, serial, a component part, collection, etc. The textual descriptors from the controlled list established in the MARC 21 Bibliographic Level position (07) of the Leader field is recommended to be used for this field's content. The Content Type field is intended to be used to identify whether the content being described is textual, a visual recording, a sound recording, etc. The textual descriptors from the controlled list established in the MARC 21 Type of Record position (06) of the Leader field is recommended to be used for this field's content.

Table 7: Examples of data elements to be provided by discovery service providers

Field Name	Examples				Media Content Recording	Media Content Motion Image
	Monographic Content	**Journal Content**	**Manuscripts / Unpublished Content**	**Media Content Image**		
Publication Title	Roses and rose culture	Nature	The evolution of cooperative signal comprehension in the domestic dog (Canis familiaris)	Portrait of Francoise en costume tailleur	Panama	The Scent of Green Papaya
Date / Date Range	1892	19950105-20131212	20060929	1946	2008	1993
Identifier	OCLC 21388020	ISSN 1476-4687			ISRC GBSUW0716131	LCCN 97505399
Content Format	Monograph	Serial	Manuscript	Image	Monographic component part	Monograph
Content Type	Language Material	Language material	Language material	Visual Material	Musical sound recording	Projected medium
Content Indexed	Metadata & Full Text	Metadata & Abstract	Metadata & Full Text	Metadata	Metadata	Metadata & Full Text (Transcript)

3.3.1.2 Distribution of Metadata Elements

1. The metadata in 3.3.1.1 should be provided in a downloadable form. It is recommended that the KBART practices be followed where relevant. However, the quantity and breadth of data handled in the discovery service supply chain necessitate some variance from the practices recommended by KBART.

2. **Format** – ODI, like KBART, recommends the use of tab-delimited text files for the transmission of metadata.

3. **File Naming** – For files provided by discovery services, files should be named as follows, encoded in UTF-8:

   `ProviderName_YYYY-MM-DD.csv`

   with the date information indicating the date of file generation, e.g.:

   `ExLibrisPrimo_2013-01-28.csv`

4. **Frequency of Updates to Metadata** – Current customers should receive regular updates of the metadata specified in 3.3.1.1 from their discovery service provider on a quarterly basis. Prospective customers should be provided equivalent data on demand.

5. **Accessibility of the Reports** – Access to reports by library staff should be secured using Internet Protocol address controls or password protection. Discovery Provider is responsible for securing access to the reports.

3.3.2 Linking

ODI recommends that discovery services implement the following protocols to ensure fair linking:

1. Discovery services should not discriminate, based on business relationships, among content providers or products (especially their own) in the methods that are used to generate results, relevance rankings, or link order.

2. In cases where the same content is available through multiple content platforms (such as a primary publisher and a secondary database, or multiple secondary databases), discovery service providers should provide mechanisms to enable libraries implementing the service to establish preferences regarding which platforms to present to users as link targets, and in what order or priority.

3. Discovery service providers should issue a statement annually to all customers (or generically on their website in an area available to all customers) explaining their business connections with content providers, including those with direct or indirect ownership relationships and those with which they have negotiated paid or other agreements for metadata deposit, direct linking, or other special arrangements.

4. Discovery service providers should offer an affirmative statement of the neutrality of their algorithms for generating result sets, relevance rankings, and link order with respect to ideological/political viewpoint, content provider source (especially but not limited to any content that it or its parent organization may provide), and any other relevant factors.

5. Discovery service providers should make other aspects of link presentation associated with a given result (including the number of links presented, the order in which links are presented within a given result, and how libraries' "get the full text" link is labeled / branded) configurable options by libraries.

6. Discovery service providers should offer a seamless link (not necessarily a full-text link) from the discovery screen(s) to the A&I interface(s) utilized in the discovery service index, when such link can be provided by the A&I provider.

7. Discovery service providers should supply content providers and libraries with information when material changes are made to the discovery service that could impact the result set or relevance rankings or link order of results.

3.3.3 File Formats and Methods of Transfer

ODI recommends that discovery services implement the following practices for the transfer of data from one party to another:

1. Discovery service providers should clearly describe their capabilities, limitations, and preferences regarding how content providers should transfer data to them for the most effective indexing by the discovery service.

2. Discovery service providers should communicate to content providers when format, schema, or transport mechanisms will have an impact on features or performance of the discovery service.

3. Metadata Encoding Schemas – Where possible, it is recommended that robust encoding schemas be used to describe a content provider's metadata. Appropriate, supported schemas should be agreed to by the content provider and discovery service provider, and preferably disclosed to customers. Table 6 lists the primary metadata encoding schemas in wide use in the library and publishing communities. ODI will maintain an updated list of such schemas on its workroom website (www.niso.org/workrooms/odi/).

Table 6: Primary encoding metadata schemas

Schema	Schema Link
MARC	http://www.loc.gov/marc/
MODS	http://www.loc.gov/standards/mods/mods-schemas.html
METS	http://www.loc.gov/standards/mets/
VRA	http://www.loc.gov/standards/vracore/
DC	http://dublincore.org/documents/dcmi-terms/
KBART	http://www.niso.org/workrooms/kbart/
EAD	http://www.loc.gov/ead/
ONIX	http://www.editeur.org/8/ONIX/

4. The method and schedule of delivery of metadata to the discovery service should also be agreed to by both parties. The file formats in Table 7 and methods of transfer in Table 8 are recommended.

Table 7: File formats

File Format	Example
text	csv, mrc, txt, etc.
xml	MARCXML, MODS XML, DC XML, etc.

Table 8: Methods of transfer

Transfer Method	Description
oai-pmh	Protocol for Metadata Harvesting from the Open Archive Initiative
resourcesync	Defined in the forthcoming NISO standard *ResourceSync Framework Specification* (Z39.99)
api	Standardized or proprietary application programming interface
ftp	Standard IETF File Transfer Protocol

3.3.4 Usage Statistics

3.3.4.1 Recommendations to Support Content Providers

3.3.4.1.1 Recommended Metrics Provided to Content Providers

Discovery service providers should offer a set of simple usage metrics on a monthly basis that are relevant to most content providers. More complex metrics should be considered for a future phase of this initiative. The recommended metrics are:

1. **Total number of searches** – A search is a query initiated by a user from a search box and is typically equated to submitting the search form of the online service.

   When possible, the discovery service should provide search counts by collection and customer for A&I databases and similar types of collections since these resources are often restricted to authenticated users; this will help to assess the use of these resources in the discovery process.

   For most content providers whose content is included in all searches on a discovery service, this metric is likely not useful as it would not be reflective of the provider's content or users' pursuit of it; rather it would merely be a report of the total number of searches run against the discovery service's index.

   Note that a search should not be counted when the application of filters, sorts, or pages to the original query happens. These alterations should be recorded as actions, not additional searches. A search initiated by a user should be tracked by purpose, such as a search results page, an RSS feed, etc. Additionally, any automated searches created by the discovery service, such as recommendations or multiple queries for one user, should not be tracked as a user-initiated query. (The definition here is intended to be consistent with the COUNTER *Code of Practice*.)

2. **Result clicks** – Report the number of times an end user clicks on a content provider's content in a set of results offered by the discovery service. Note that where there are multiple content sources that contributed article metadata to the discovery service, each content provider is credited a result-click increment. Further, note that a result should be counted only once, regardless of how many clicks a user made on the same record; i.e., if a user clicks on a single record multiple times to see the details, then the count should only be 1.

   When possible, the discovery service should provide Result Click counts by collection and customer for A&I databases and similar types of collections since these resources are often restricted to pre-authenticated users; this will help the content provider to assess the use of these resources in the discovery process

3. **Total number of click-throughs** – This metric reports how frequently a content provider's full content items are selected for access by end users. Discovery services can only measure the requests for the content and not the success or failure of content retrieval. Click-through counts may also include indication of whether a user's request was serviced by an OpenURL link or via a direct link from the discovery service.

   NOTE: Where there are multiple content sources that contributed article metadata to the discovery service, each content provider is credited a click-through increment.

The recommended metrics should be reported on a monthly basis in a simple CSV format so that content providers can readily analyze and collate the data as desired.

It is recommended that the metrics above be incorporated by COUNTER in a future version of the *Code of Practice* that specifies details of format and distribution.

3.3.4.1.2 Link Source identification

In order for content providers to identify, and therefore quantify, the type of user traffic originating natively from a discovery service to their own content platforms, it is recommended that discovery service provider links include a referrer URL identifying the discovery service. (The expectation is that intermediary link resolvers would pass along this referrer identification to the content provider with the OpenURL request.)

3.3.4.2 Recommended Metrics Provided to Libraries

Discovery service providers should offer a set of structured metrics to illustrate to libraries how the discovery service is utilized. More complex metrics should be considered for a future phase of this initiative. The current recommended metrics are as follows:

1. **Total number of searches per month** – Report the total number of times a user performed a search of the discovery service, providing for tracking of month-over-month usage trends. Additionally, when possible, the discovery service should provide search counts by collection for A&I databases and similar types of collections since these resources are often restricted to authenticated users; this will help to assess the use of these resources in the discovery process.

2. **Total number of unique visitors per month** – Report the total number of unique visitors by month to provide library clients with month-to-month trending of usage among their constituency.

3. **Total number of click-throughs per month** – Report the total number of times a user has clicked on a link to request full-text content. This metric provides an indication of how frequently the discovery service yielded a result perceived by the user to be valuable to their research. When possible, the discovery service should provide click-through counts by collection for items in A&I databases and similar types of collections since these resources aid discovery of items that may not be the ultimate target destination for the user; this will help to assess the contribution of these resources in the discovery process.

4. **Top 500 search queries for the last period** – Report on the 500 most frequently submitted search queries for the specified period. Search query data enables a library to consider the types of searches—and topics—being researched when using the discovery service.

5. **Top 100 referring URLs to the discovery service for the last period** – Report on the 100 most frequently used URLs to bring the user to the discovery service. This metric provides the library with an indication of how users are getting to the discovery service.

To allow libraries to analyze the usage of their discovery service, they will need to have the ability to access the aforementioned metrics and integrate them into a usage analysis or library management system. To satisfy this need, it is recommended that the reporting of the data points be provided in CSV formatted reports. Note that implementation of a discovery service might result in declining search counts from native content platforms. This is not unexpected since the library is promoting access to its content via search of the discovery service's search index. However, full-text retrievals from native product platforms typically increase upon introduction of a discovery service. As full text is the artifact patrons seek, libraries should find full-text document retrievals a better metric for assessing the value of a content product rather than search counts on native product platforms.

It is recommended that the metrics above be incorporated by COUNTER in a future version of the *Code of Practice* that specifies details of format and distribution.

Section 4: Recommended Next Steps

The ODI workgroup identified a number of items that need to be addressed once the baseline recommendations, outlined above, are in place.

4.1 Collaborative Discussion

The ODI workgroup recommends the establishment of a collaborative mechanism for further discussion among all ODI stakeholders, broadening the discussion beyond the ODI working group members. This should provide educational opportunities to promote adoption of the ODI recommended practices, provide support for content providers and discovery providers during adoption, provide open dialog mechanisms to discuss ongoing issues related to all aspects of discovery platforms, and determine timing for next steps for ongoing work.

4.2 Application Programming Interfaces (APIs)

As libraries increasingly use Application Programming Interfaces (APIs) to access data managed in discovery services, there is a growing need to have a clear understanding of the usage rights that come with discovery service records and, indeed, with individual data elements contained in a record. Some libraries (e.g., the University of Michigan, Villanova, and North Carolina State University) are providing access to discovery services exclusively or primarily through APIs. Other libraries are considering ways to use citation information pulled dynamically from discovery services as feeds into other applications. There is no current standard around access rights to data other than a distinction between "authenticated user" and "unauthenticated user" attached to the entire record. When is it appropriate for a library to use selected fields retrieved from a discovery service outside the context of a discovery interface? And where appropriate, what fields may be displayed to authenticated and unauthenticated users? These are some of the issues that should be discussed and potentially standardized.

4.3 Managing "Restricted" Content in Discovery Services

Some content providers stipulate that their content—or some of their content (typically abstracting and indexing databases)—be available only to discovery service users from institutions that subscribe to said content. Today discovery service providers that offer this feature have created mechanisms for periodic reporting on activation/usage of this restricted content. Some content providers have suggested automated mechanisms for discovery services to query their systems to check subscription information; for this a standard is recommended to prevent the development of multiple proprietary mechanisms. The development of such a standard was considered out-of-scope for ODI; but should be considered for a future phase if content providers continue to place search restrictions on their content when indexed in discovery services.

4.4 Reporting on Discovery Service Content at a Collection Level

ODI addresses the need by libraries to receive current reports on the contents of the discovery services. However, only the most important need is captured here: to provide reports at the level of the "work," i.e., journal title, book title.

Librarians have also expressed the need to have information at the collection level. Some of the specific requests are as follows.

1. For journal collections, report on the entire contents of the applicable packages included, specifically:

   - Aggregator database, i.e., if the entire contents has been ingested into the discovery repository or, if not, what the limitations are

   - Specialized databases, i.e., A&I databases that are not multi-disciplinary aggregated databases

   - E-book collections, i.e., not individual titles but description of the source and general content

   - Other documents and reports, i.e., not individual titles but description of the source and general content

   - E-reference resources, e.g., encyclopedias, directories, etc.

2. For such collections it is recommended that the following elements should be addressed:

   - Collection name

   - Source of metadata, e.g., publisher feed, aggregator database, specialized database

   - Extent of metadata, i.e., minimal or enhanced (see Section 3.2.1 above)

   - Access status for the metadata, e.g., open, subscription required, other

   - Access status for the full text, e.g., open, subscription required, other

3. It is further recommended that attention be paid to KBART collection descriptions and other industry initiatives that may arise.

4.5 On-Demand Lookup by Discovery Service Users

It is recommended that a future ODI phase look at the issue of enabling end users to query their discovery service to check whether or not a particular collection, journal, or book is included in the indexed content.

4.6 Research and Analysis to Monitor Discovery Services for Fair/Unbiased Linking

It is recommended that a future ODI phase, to the extent feasible, investigate methods to further analyze linking from discovery services to content providers, to measure the occurrence of unbiased links and content usage.

4.7 Further Interaction with COUNTER

ODI recommends further interaction with COUNTER to better understand its direction and the applicability of its *Code of Practice* covering the content provider and library needs of index-based discovery services.

4.8 Identification of Features/Functionality of Discovery Services to Address Needs of A&I Service Providers

ODI recommends a future in-depth look at the specific needs of A&I service providers, including the issues identified in Section 2.2, amongst others, to identify subsequent recommendations that would address these concerns and ultimately encourage discovery participation from providers of these services.

Appendix A:
Survey Summary

During September and October 2012, the ODI surveyed librarians, content providers, and discovery service providers to learn more about the current state of satisfaction with the new research tools and to measure the value of various requirements in cross-sector practice. The survey addressed current levels of scholarly metadata delivery/indexing, technical successes/opportunities in these data exchanges, and potential benefits of greater development/distribution of discovery tool usage data.

More than 2000 individuals were contacted by ODI members and NISO staff to recruit participation in this survey (drawing on NISO's *Newsline* e-mail list and registrants of 17 industry mailing lists). Cross-sector contributors were targeted for participation, drawing primarily on those based in the United States and the United Kingdom.

A total of 871 completed survey responses were logged: 782 identified as librarians; 74 identified as publishers; and 15 identified as discovery service providers. The ODI believes these numbers reflect an active engagement by libraries, which demonstrates the value of these new tools to library patrons and to the future of library services as a whole. However, the total number of responses relative to the global library population was not evaluated.

Survey findings and analysis are provided in greater detail in the previously published report, available at: www.niso.org/apps/group_public/download.php/9977/NISO ODI Survey Report Final.pdf

Appendix B:
Content Provider Conformance Checklist

When requested by libraries, Content Providers can use the table below to indicate their ODI compliance. A 'Y' (for Yes) in column 1 indicates compliance with the indicated paragraph of this Recommended Practice. A 'P' response indicates Partial compliance for which explanatory comments should be entered in the last column. An 'N' (No) response indicates that the content provider does not comply with the recommendation. Explanatory comments may be added for any response.

Y/N/P	Recommendation	Paragraph	Comment
	Content Provider makes available to Discovery Service Providers core metadata and underlying full-text/original content for complete offerings.	3.2.1.1 (1)	
	Content Provider makes available to Discovery Service Providers, the core set of metadata elements (see 3.2.1.2) for each item submitted for indexing.	3.2.1.1 (2)	
	Content Provider provides the content item and additional descriptive content for as much of their content as possible.	3.2.1.1 (3)	
	Content Provider provides libraries, on request, with a statement of participation in the discovery services, including disclosure of coverage depth and content depth.	3.2.2	
	Content Provider's agreements with Discovery Service Providers do not include non-disclosure agreements.	3.2.3	
	The transfer of Content Provider's data to Discovery Service Providers makes use of existing standards where applicable and uses one of the metadata encoding schemes listed in 3.3.3.	3.2.4	

Appendix C:
Discovery Provider Conformance Checklist

When requested by libraries, Discovery Service Providers can use the table below to indicate their ODI compliance. A 'Y' (for Yes) in column 1 indicates compliance with the indicated paragraph of this Recommended Practice. A 'P' response indicates Partial compliance for which explanatory comments should be entered in the last column. An 'N' (No) response indicates that the content provider does not comply with the recommendation. Explanatory comments may be added for any response.

Y/N/P	Recommendation	Reference	Comment
	Discovery Service provides content listing for library customers	3.3.1	
	Linking	3.3.2	
	Discovery Service does not discriminate among Content Providers contributing to the service.	3.3.2 (1)	
	Mechanisms are offered to enable libraries to establish preferences regarding which platforms to present to users as link targets, and in what order or priority.	3.3.2 (2)	
	Discovery Service confirms with non-bias with regard to content indexed and results presented to the user. A statement in this regard is published annually.	3.3.2 (3)	
	Discovery Service uses an algorithm that is non-preferential with regard to the Content Provider for generating result sets, relevance rankings, and link order.	3.3.2 (4)	
	Link presentation associated with a given result is configurable by libraries.	3.3.2 (5)	
	A seamless link (not necessarily a full-text link) is provided from the discovery screen(s) to the A&I interface(s) utilized in the Discovery Service index.	3.3.2 (6)	
	Discovery Service provides usage metrics to content providers.	3.3.4.1	
	Report total number of searches	3.3.4.1.1 (1)	
	Report on result clicks.	3.3.4.1.1 (2)	
	Report on total number of click-throughs.	3.3.4.1.1 (3)	
	Report referrer URL identifying the Discovery Service	3.3.4.1.2	
	Discovery Service provides usage metrics to libraries.	3.3.4.2	
	Report on total number of searches per month.	3.3.4.2 (1)	

Y/N/P	Recommendation	Reference	Comment
	Report on total number of unique visitors per month.	3.3.4.2 (2)	
	Report on total number of click-throughs per month.	3.3.4.2 (3)	
	Report top 500 search queries for the last period.	3.3.4.2 (4)	
	Report top 100 referring URLs to the Discovery Service for the last period.	3.3.4.2 (5)	

Bibliography

Asher, Andrew D., Lynda M. Duke, and Suzanne Wilson. "Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources." *College & Research Libraries,* September 2013, 74 (5): 464-488. Available at: http://crl.acrl.org/content/74/5/464.full.pdf+html

Baldwin, Dee, Michael Kucsak, and Alice Eng. "Don't Touch that String! There Went the Databases." *Information Outlook*, September/October 2012, 16 (5): 24-32. Originally presented at the 2012 Special Libraries Association Annual Conference, Chicago, IL, July 16, 2012. Available at: http://digitalcommons.unf.edu/library_facpub/7/

COUNTER Code of Practice for e-Resources. Release 4. COUNTER, April 2012. Available at: http://www.projectcounter.org/code_practice.html

COUNTER Code of Practice for Usage Factors. Draft Release 1. COUNTER, March 2012. http://www.projectcounter.org/usage_factor.html

DCMI Usage Board. *DCMI Metadata Terms*. DCMI Recommendation. Dublin Core Metadata Initiative, June 14, 2012. Available at: http://dublincore.org/documents/dcmi-terms/ [Also published as ANSI/NISO Z39.85-2012, available at http://www.niso.org/standards/z39-85-2012.]

Digital Library Federation ILS Discovery Interface Task Group (ILS-DI) [webpage]. http://old.diglib.org/architectures/ilsdi/

Discovery Open Metadata Principles. London: JISC Discovery Programme, May 26, 2011. Available at: http://discovery.ac.uk/businesscase/principles/

Ekins, Andy, and Lukas Koster. Unified Resource Discovery Comparison [website]. http://sites.google.com/site/urd2comparison/

Encoded Archival Description (EAD) [website]. http://www.loc.gov/ead/

ICOLC. *Revised Statement on the Global Economic Crisis and Its Impact on Consortial Licenses*. International Coalition of Library Consortia, originally issued January 19, 2009; reissued June 14, 2010. Available at: http://icolc.net/statement/revised-statement-global-economic-crisis-and-its-impact-consortial-licenses

Information Retrieval: Application Service Definition & Protocol Specification. ANSI/NISO Z39.50-2003 (R2009). Baltimore, MD: National Information Standards Organization, approved 2003; reaffirmed 2009. Available at: http://www.niso.org/standards/z39-50-2003/

Kelley, Michael, "Stakeholders Strive to Define Standards for Web-Scale Discovery Systems." *The Digital Shift*, October11, 2012. Available at: http://www.thedigitalshift.com/2012/10/discovery/coming-into-focus-web-scale-discovery-services-face-growing-need-for-best-practices/

KBART: Knowledge Bases And Related Tools. NISO RP-9-2010. Baltimore, MD: National Information Standards Organization, February 8, 2010. Available at: http://www.niso.org/publications/rp/rp-9-2010/

Knowledge Bases and Related Tools (KBART) Recommended Practice. Phase II revision. NISO RP-9-2014. Baltimore, MD: National Information Standards Organization, March 27, 2014. Available at: http://www.niso.org/publications/rp/rp-9-2014/

Knowledge Base And Related Tools (KBART) [webpage]. http://www.niso.org/workrooms/kbart

Luther, Judy, and Maureen Kelly. "The Next Generation of Discovery." *LibraryJournal.com*, March 15, 2011. Available at: http://www.libraryjournal.com/lj/home/889250-264/the_next_generation_of_discovery.html.csp

MARC Standards [website]. http://www.loc.gov/marc/

Metadata Encoding & Transmission Standards (METS) [website]. http://www.loc.gov/standards/mets/

Metadata Object Description Schema (MODS) [website]. http://www.loc.gov/standards/mods/

Music Library Association's Emerging Technologies and Services Committee. *Music Discovery Requirements*. Middleton, WI: Music Library Association, April 23, 2012. Available at: http://committees.musiclibraryassoc.org/uploads/ETSC/MDRdocument.pdf

NISO Metasearch Initiative [webpage]. http://www.niso.org/workrooms/mi

NISO Metasearch XML Gateway Implementers Guide. NISO RP-2006-02. Bethesda, MD: National Information Standards Organization, August 7, 2006. Available at: http://www.niso.org/publications/rp/RP-2006-02.pdf

ONIX family of standards [website]. http://www.editeur.org/8/ONIX/

The Open Archives Initiative Protocol for Metadata Harvesting. Version 2.0. Open Archives Initiative, June 14, 2002. Available at: http://www.openarchives.org/OAI/openarchivesprotocol.html

Postel, J., and J. Reynolds. *File Transfer Protocol*. RFC 959 aka STD9. Internet Engineering Task Force, October 1985. Available at: http://www.rfc-editor.org/rfc/rfc959.txt

Quint, Barbara. "The Undiscovered Discovery." *Information Today*, July/August 2010, 27 (7): 7-11. Available at: http://www.infotoday.com/IT/jul10/Quint.shtml

Recommended Practices: Discovery Services. Philadelphia, PA: National Federation of Advanced Information Services, August 30, 2013. Available at: http://info.nfais.org/info/Recommended_Practices_Final_Aug_2013.pdf

ResourceSync Framework Specification. NISO Z39.99-2014. National Information Standards Organization and Open Archives Initiative, April 21, 2014. Available at: http://www.niso.org/standards/z39-99-2014/

Rowe, Ronda. "Web-Scale Discovery: A Review of Summon, EBSCO Discovery Service, and World Cat Local." *The Charleston Advisor*, July 2010, 12 (1): 5-10. Available at: http://dx.doi.org/10.5260/chara.12.1.5

searchRetrieve. Version 1.0. OASIS Standard, January 30, 2013. Available at: http://www.loc.gov/standards/sru/oasis/

Somerville, Mary M., Barbara J. Schader, and John Sack. *Improving the Discoverability of Scholarly Content in the 21st Century: Collaboration Opportunities for Librarians, Publishers, and Vendors.* Thousand Oaks, CA: SAGE Publications, n.d. (est. 2011). Available at: http://www.sagepub.com/repository/binaries/librarian/DiscoverabilityWhitePaper/

Vaughan, Jason. "Web Scale Discovery Services." *Library Technology Reports*, January 2011, 47 (1). Available at: http://dx.doi.org/10.5860/ltr.47n1

VRA Core [website]. http://www.loc.gov/standards/vracore/

Way, Doug, "The Impact of Web-Scale Discovery on the Use of a Library Collection." *Serials Review*, 2010, 36 (4): 214-20. Available at: http://dx.doi.org/10.1016/j.serrev.2010.07.002