
New Discovery Services and Library Bibliographic Control

MYUNG-JA HAN

ABSTRACT

To improve resource discovery and retrieval, libraries have implemented new discovery services, such as next generation catalogues, federated search, and Web-scale discovery, in addition to their traditional integrated library systems. These new discovery services greatly improve the user experience by utilizing existing cataloguing records housed within the library system or in combination with metadata from other sources, both in and outside of libraries. However, to maximize the functionality of these discovery services, libraries must reexamine current cataloguing practices and the way libraries control the bibliographic description to better serve the user's needs. This report discusses how new discovery services use the cataloguing records and the challenges that libraries encounter in bibliographic control to work with new discovery services, including the quality of cataloguing records, granular levels of bibliographic description, and integration of user-generated metadata into the cataloguing records. Each of these aspects requires further discussion.

INTRODUCTION

The way that libraries organize and manage their resources has changed along with the formats of resources to which libraries provide access, advancement of information technology, the development of tools, such as integrated library systems (ILSs), and metadata standards that are used for recording bibliographic information, storing cataloguing records, and employing them in search and discovery of resources that libraries hold. Since the late 1960s, libraries have been using MACHine Readable Cata-

loguing (MARC) as the metadata standard for bibliographic description and Anglo-American Cataloguing Rules (AACR) as the content standard for creating cataloguing records, mostly for describing resources in print format.

In recent years, however, because of digital library developments and increasing acquisition of resources in digital format, libraries now create records in different metadata standards and use different content standards other than MARC and AACR, especially because they must capture information, such as technical and administrative metadata, that is not required for description of print format resources. These changes also require libraries to reexamine their approaches to controlling bibliographic descriptions and evaluating the quality of cataloguing records, even as they continue to manage traditional cataloguing records in the MARC format.

Another change facing libraries is the arrival of new and advanced discovery and retrieval systems, notably next-generation catalogues (NGCs), federated search, and Web-scale discovery. Many libraries are experimenting with such systems to improve access to the resources they hold. The experience of users and the performance of these new systems depend heavily on the quality of the cataloguing records in a given system and the approach each library takes to the control of bibliographic records. Considerations include whether a library provides different levels of granularity for bibliographic descriptions and whether it uses metadata standards that are useful in describing resources across a variety of different formats.

This report outlines the University of Illinois Library's experiences with multilevel discovery services, including VuFind, an open-source solution for NGC implementations, and Easy Search, which facilitates federated search technology, as well as the traditional WebVoyage online public access catalogue (OPAC) as part of its Voyager ILS. The focus here is on the manner in which these discovery systems use cataloguing records and the challenges to current bibliographic control that supports the discovery systems in meeting user needs and improving access to and retrieval of resources in diverse formats from diverse sources.

NEXT-GENERATION CATALOGUE SERVICE

The term *next-generation catalogue* first appeared in 2006 when the North Carolina State University Library implemented a new discovery software, Endeca's Information Access Platform, on top of its cataloguing system. Also known as a discovery layer, the NGC enables use of existing cataloguing records, which Antelman, Lynema, and Pace refer to as "rich metadata trapped in MARC format records for enriching the collection browsing" that enhances discovery and retrieval of resources (2006, p. 128). The NGC offers services that the traditional cataloguing system (i.e., integrat-

ed library system [ILS]) could not offer to users, such as new browsing capabilities using faceted navigation, subject access, and, most important, an intuitive interface. Many libraries have implemented NGCs on top of their traditional OPAC; in other words, both systems use the same cataloguing records, but they provide different discovery services to users.

According to several user studies published since 2006 on NGCs, users found the faceted navigation feature very helpful since they could narrow down the search results set in their OPAC using categories, such as date, topic, and format, with information from traditional MARC format records (Antelman, Lynema, & Pace, 2006; Denton & Coysh, 2011; Emanuel, 2011; Ho, Kelly, & Garrison, 2009; Olson, 2007). Usability tests also found that the effective performance of new NGCs depends on the quality of cataloguing records. Denton and Coysh's research revealed that inconsistencies, inaccuracies, and incompleteness of metadata caused less than ideal results of faceted navigation display (2011). Emanuel's research (2011) based on the University of Illinois Library's VuFind implementation also found that inconsistencies in the cataloguing record, such as using noncontrolled vocabularies, impeded effective use of the faceted navigation feature display.

Although NGCs are best known for faceted navigation, another feature that makes these systems attractive to many libraries is their ability to use data from multiple sources other than the OPAC, such as digital repositories and article databases. With additional tools and technical support, NGCs can harvest data from other sources and index them for searching and browsing, together with the library's own cataloguing records to augment system performance. This can provide article and chapter-level access to material, which, according to Yang and Hoffman (2011), is one of the features that users most want.

However, this new feature is only possible in systems that dynamically search article databases and indexing and abstracting databases using the patron's search term (i.e., federated search) or by ingesting additional metadata into the system and indexing with library metadata (i.e., Web-scale discovery). This seems to imply that in order to make these new services possible, the upgraded NGC should be capable of using the metadata from heterogeneous collections (i.e., both locally owned and subscription content) with resource descriptions in diverse formats that were created using different metadata standards, and controlled vocabularies (Sadeh, 2008). Yang and Hoffmann's (2011) study of the OPACs of 260 academic libraries showed that federated search was one of the three weakest areas in library implementations of NGC systems; the other two weak areas were relevancy assessments based on circulation statistics and patron recommendations based on previous transactions. The study also found that none of the surveyed libraries had a federated searching feature and

added that the federated search was the most important—but also the most difficult—of all NGC features to accomplish because of “technical reasons and political complications” (p. 275).

DISCOVERY SERVICES IN THE UNIVERSITY OF ILLINOIS LIBRARY

The University of Illinois Library provides its users three different discovery services: NGC (VuFind), the Classic Catalogue search using WebVoyage, and Easy Search that exploits a federated search technology. Since each discovery service has its own pros and cons and serves a different user group, the library decided to keep all three services available to users (fig. 1).

Next-Generation Cataloguing—VuFind

As a member of the Consortium of Academic and Research Libraries in Illinois (CARLI), the library participates in the I-Share program, along with seventy-six member libraries (CARLI, 2011a) among CARLI’s 153 member institutions. I-Share uses the Voyager software developed by the ExLibris

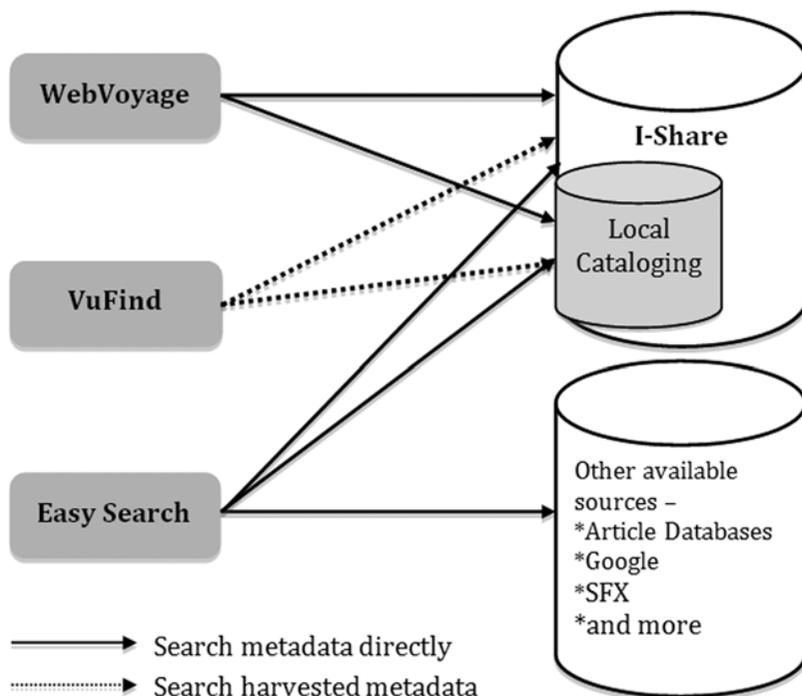


Figure 1. University of Illinois Library provides three different discovery services.

Group as the online cataloguing system and provides member libraries with “an online catalogue of their own collection as well as a merged, union catalogue of the holdings of all I-Share Libraries” (CARLI, 2011). In 2007, CARLI implemented VuFind (<http://vufind.org/>), an open-source interface used as a discovery layer of the Voyager ILS. As a member of the I-Share Libraries, the University of Illinois Library changed its default OPAC discovery interface to VuFind from WebVoyage (fig. 2). To work with VuFind, cataloguing records are indexed and stored in a separate server for discovery services, which means that the information used by VuFind is one-day old, and the changes made on the cataloguing records are not reflected in the search and discovery by VuFind until the following day. The features introduced in CARLI’s VuFind are almost the same as at other libraries that have implemented next generation cataloguing (e.g., a single simple search box and facets to refine search results; CARLI, 2011b).

Among Yang and Hoffmann’s (2011) checklist of twelve features of NGC discovery tools, the University of Illinois Library’s VuFind provides six features: a state-of-the-art interface, faceted navigation, a simple keyword search box with a link to the advanced search on every page, enriched content, user-generated content, and persistent links. However, the library’s other discovery service, Easy Search, provides two additional fea-

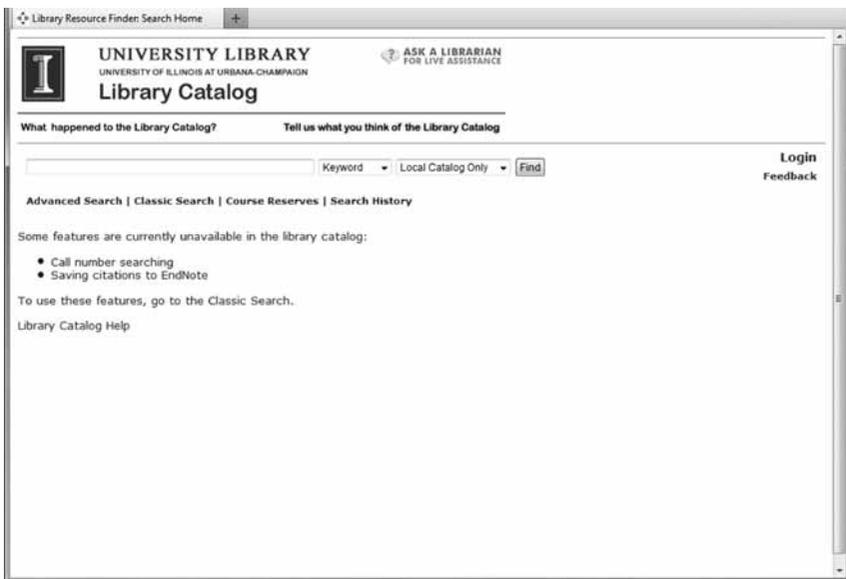


Figure 2. University of Illinois VuFind offers a simple keyword search box.

tures common to NGCs, such as single-point entry for all library resources and a *did you mean?* feature. Compared to the list of the NGC features identified by Yang and Hoffmann (2011), the University of Illinois Library search service has eight of the twelve possible features, which makes it one of only seven among 260 surveyed academic institutions that have more than eight features implemented in their discovery systems.

The number of faceted navigation options provided by the University of Illinois Library's VuFind implementation is more extensive than is provided by other libraries. For example, when users choose a default search option (i.e., keyword searching), the search results are displayed with ten faceted navigation options from which users can narrow down the search results: format, location, author, topic, subject area, language, genre, era, region, and title. In order to provide these faceted navigation options, the system indexes selective information from MARC-format bibliographic records as shown in table 1 (CARLI, 2011b). The information used for this rather extensive list of faceted navigation options comes from twelve descriptive MARC data fields and one fixed field (or twenty-five data fields/subfields) as well as library holdings records. In addition, the system indexes other information from the MARC format record in order to display additional views such as "VuFind Record Page" and "More Details" that include publisher, edition, series, ISSN, and ISBN.

In addition, the University of Illinois Library's VuFind has extra features that enhance the user experience, such as cover images, book reviews, table of contents, abstracts, comments, and user tagging features. (WebVoyage also displays cover images and table of contents.) Tagging and comments features are also provided as Web 2.0 services, allowing users to add their own personal tags and comments to items, or to recommend an item to their friends. Whether such user-generated metadata or other enhanced content should be integrated into the cataloguing records and how these might be used in the future have not yet been discussed.

Table 1. Information used for faceted navigation options

Faceted Navigation Options	MARC Data Field/Subfield Used
Title	245 subfield ab
Format	245 subfield h, RecType+BibLvl
Location	Holdings record
Author	100 subfield a
Topic	650 subfield ax, 690 subfield ax
Subject Area	050a, 090a
Language	008 bytes 35-37
Genres	655 subfield a
Era	650y, 651y, 690y, 691y
Regions	650z, 651az, 690z, 691az

From VuFind FAQ (2011).

WebVoyage

The University of Illinois also maintains the traditional WebVoyage OPAC. The Library decided to keep WebVoyage and named it the “Classic Catalogue.” This interface is available as part of the “Library Catalogue” page (<http://www.library.illinois.edu/catalogue/>) that provides two separate searches: a “Quick Search” and an “Advanced Search.” Both types of search are used in known-item searches by scholars and graduate students alike, and many users are already familiar with the “Classic Catalogue” interface. In addition, since the bibliographic records and holdings records in WebVoyage are refreshed dynamically whenever the database is refreshed, users can search with newly updated data at any given time, while VuFind data are not updated in real time, and the available data are from the previous day. Because of the time delay in refreshing bibliographic data in VuFind, several scripts used by the library’s research projects operate using WebVoyage and not VuFind when harvesting bibliographic records. Detailed usage statistics of WebVoyage after the introduction of VuFind are not currently available.

Easy Search

Easy Search is a locally developed discovery service that exploits federated search technology. As mentioned earlier, more and more users want to search for books, articles, and all available resources with one search. Easy Search meets the users’ needs by simultaneously searching the local catalogue, I-Share Libraries, article databases, indexing and abstracting services, Google, and many other available sources directly instead of harvesting and indexing metadata in a local server, as does Web-scale discovery. The search results are then displayed in a list that indicates the total number of resources located in each source.

The search categories offered in Easy Search are simple (i.e., keyword, title, and author) because the further refinement options are available in the database provider’s page when users select resources located in a given database. This also means that the information used in Easy Search is far fewer than the twenty-five data fields/subfields that VuFind uses for its faceted navigation options. Easy Search added a script that provides a *did you mean?* service and author-name normalization to improve the precision and recall rate of the search results set. The *did you mean?* feature corrects typos and suggests controlled terms that are unfamiliar to users but are used in cataloguing records and also terms that are commonly used in similar searches. The author-name correction feature was added because there are inconsistencies in how each database and source uses personal names (e.g., order of names and form of names such as full name or abbreviated name), which impedes the interoperability of metadata (Lagoze, 2011). The script added in Easy Search checks the search terms, and if the terms are recognized as a personal name or users choose

an author search, it automatically converts the name with the right order, last name and first name, to make the term more interoperable and understandable by the target source systems on which the searches are performed.

DISCOVERY SERVICES AND BIBLIOGRAPHIC CONTROL

The implementation and transition of discovery services from WebVoyage to VuFind and Easy Search brought up several issues on bibliographic control that should be discussed and researched further.

First, the quality of the traditional cataloguing record is of critical importance. Until recently, the issues of the quality of cataloguing records and how the ILS uses the information stored in MARC format records have not been an overly frequent topic of discussion. Although the Library of Congress published “Minimal Level Record Examples” (2003), it is hard to enforce the creation of even minimal-level cataloguing records. This state of affairs results in inconsistent and incomplete records in cataloguing systems, both in the local system and in the union cataloguing system provided by Online Computer Library Center (OCLC). Also, records that do not meet minimal requirements have not been clearly visible in traditional systems because of the way the ILS uses cataloguing records. The traditional ILS deals with records one at a time for search and discovery. After an initial search, results are displayed individually in WebVoyage. However, in a new discovery system, the faceted navigation service displays terms as used in individual records for each category, which clearly reveals the quality of records, especially inconsistencies in the use of controlled terms and incomplete records that lack useful information for faceted navigation options (Denton & Coysh, 2011).

Second, there is a discrepancy between the content of cataloguing records and what new systems need in order to provide optimal discovery services. Comparison of the top thirty-six occurring data fields/subfields in MARC format records reported in Moen’s study (2007) with the twenty-five data fields/subfields used by the library’s faceted navigation service in VuFind found that only nine fields appeared in both lists. Also, among twenty-five data fields/subfields, only three fields consistently appeared in the minimal-level record examples from the Library of Congress. Although library catalogue records contain not only descriptive information but also administrative information, such as a local bibliographic number and control numbers, these two cases show the possibility that some resources are not discoverable if the cataloguing records do not include the data fields/subfields used for discovery by faceted navigation services.

Third, increasing the granular levels of cataloguing records should also be explored more fully. Until now, libraries have been describing books or journals as a whole unit in MARC format, rather than providing article-level or chapter-level descriptions. Increasing the granularity of metadata

for books or journals would enhance the discoverability of the resources at a more in-depth level; thus, libraries should seek to identify tools to create and manage metadata at various levels of granularity. Although some of the basic information that would enhance this approach is available in article databases and indexing and abstracting services, libraries could also find ways to share their expertise in organizing information and using controlled vocabularies to make database records more consistent in quality. In addition, since creating article-level, chapter-level, and item-level records requires resources and time, libraries and publishers could work to find ways to automatically generate more granular levels of metadata for cataloguing records to provide better service to patrons.

Fourth, integration of user-generated metadata and added content into cataloguing records is another matter that should be discussed. Since the library provides a service for user tagging and comment features, as well as enriched content such as cover images, table of contents, and abstracts, libraries need to decide whether the user-generated metadata and enriched contents should be integrated into or preserved along with local and shared cataloguing records. Prior to making this decision, an evaluation of the usefulness of patron-contributed information and determining what kind of information needs to be captured (e.g., name, date, and contents) should be undertaken. Such an approach requires technological developments in capturing and storing metadata as well as the selection and implementation of a metadata standard that works best for this purpose.

Last, the question remains of whether libraries should keep using MARC as a container for library bibliographic description. MARC is not the ideal format for describing chapter-level and article-level information, nor is it ideal for user-generated metadata. The Library of Congress (2011) announcement *A Bibliographic Framework for the Digital Age* mentioned that the community should move into a “robust, open, and extensible carrier for our rich bibliographic data” other than MARC. However, there is currently no clear plan after MARC, but the new format should be simpler and easier for humans as well as for machines.

CONCLUSION

Since the new discovery service systems, including NGCs, federated search, and Web-scale discovery, use traditional cataloguing records in MARC format for resource discovery and retrieval, the quality of cataloguing records assumes a critical role in user services. Usability tests on NGC systems have already revealed that the variable quality of cataloguing records could impede optimal functioning of faceted navigation services, especially because of inconsistencies, incompleteness, and inaccuracies in existing cataloguing records (Denton & Coysh, 2011; Emanuel, 2011).

While discussions of metadata quality in a digital environment have ac-

tively progressed in recent years in research and training programs, notably with the project Metadata for You and Me (<http://images.library.uiuc.edu/projects/mym/>), the quality of traditional cataloguing records has not often been discussed in this context. One of the reasons could be that it is commonly understood that metadata for digital resources should be shareable in order to improve the discoverability of such resources in an aggregated environment, but the importance of the same kind of interoperability has not been taken into consideration for library bibliographic records. However, traditional cataloguing records no longer reside only in each library's local system. The information contained in the records is often extracted and indexed along with other information in federated and aggregated environments, for use both within and outside of an institution.

With each of the new discovery services, libraries need to examine the current practice of how cataloguing records are created and what they contain and find better ways to create records that are good enough for the *user*. Ultimately, this important step will improve the management, search, and discovery of resources.

REFERENCES

- Antelman, K., Lynema, E., & Pace, A. K. (2006). Toward a twenty-first century library catalogue. *Information Technology & Libraries*, 25(3), 128–139.
- Consortium of Academic Research Libraries in Illinois. (2011a). *I-Share* Retrieved November 30, 2011, from <http://www.carli.illinois.edu/index.php?page=i-share>
- Consortium of Academic Research Libraries in Illinois. (2011b). *Frequently Asked Questions About Vufind*. Retrieved November 28, 2011, from <http://www.carli.illinois.edu/mem-prod/i-share/opac/vufind-faq.html>
- Denton, W., & Coysh, S. J. (2011). Usability testing of VuFind at an academic library. *Library Hi Tech*, 29(2), 301–319. doi:10.1108/07378831111138189
- Emanuel, J. (2011). Usability of the VuFind next-generation online catalogue. *Information Technology & Libraries*, 30(1), 44–52.
- Ho, B., Kelley, K., & Garrison, S. (2009). Implementing VuFind as an alternative to Voyager's Web Voyage interface: One library's experience. *Library Hi Tech*, 27(1), 82–92.
- Lagoze, C. (2011). Disciplinary culture and interoperability, an incompatible mix? [PowerPoint slides]. Retrieved November 28, 2011, from <https://netfiles.uiuc.edu/trainer1/errtslides/lagoze%20slides.ppt>
- Library of Congress. (2011, October 31). *A Bibliographic framework for the digital age*. Retrieved November 29, 2011, from <http://www.loc.gov/marc/transition/news/framework-103111.html>
- Library of Congress, Network Development and MARC Standards Office. (2003). *Appendix C: Minimal level record examples*. MARC 21 format for bibliographic data. Retrieved November 28, 2011, from <http://www.loc.gov/marc/bibliographic/bdapndxc.html>
- Moen, W. E. (2006, August 7). Examining MARC records as artifacts that reflect metadata utilization decisions. *First Monday*, 11(8). Retrieved November 28, 2011, from http://www.unt.edu/wmoen/publications/MARCPaper_Final2003.pdf
- Olson, T. A. (2007). Utility of a faceted catalogue for scholarly research. *Library Hi Tech*, 25(4), 550–561.
- Sadeh, T. (2008). User experience in the library: A case study. *New Library World*, 109(1/2), 7–24.
- Yang, S. Q., & Hofmann, M. A. (2011). Next generation or current generation? A study of the OPACs of 260 academic libraries in the USA and Canada. *Library Hi Tech*, 29(2), 266–300. doi:10.1108/07378831111138170

Myung-Ja “MJ” Han has been a metadata librarian and assistant professor of library administration at the University of Illinois since 2008. Prior to this position, she was a metadata librarian and visiting assistant professor of library administration for the Illinois Harvest Project. As a metadata librarian, her job responsibility consists of developing best practices and creating metadata for the library’s digital collections. Her research interests include interoperability of metadata, linked data in a library environment, and relationships between collection-level metadata and item-level metadata.