

## **Users and Uses of Bibliographic Data**

### **Background Paper for the Working Group on the Future of Bibliographic Control**

February 25, 2007

Nancy J. Fallgren

College of Information Studies, University of Maryland

Consultant to the Working Group on the Future of Bibliographic Control

Since the advent of the Internet, our conceptions of information resources, information seeking behavior, information access, and information use have evolved. These changes are leading us also to question the validity of the bibliographic data that we record in a more fluid information environment. Any knowledge organization system, such as a bibliographic database, is useful only to the extent that it meets its users' needs and requirements; therefore, in reflecting on bibliographic data, we first need to understand its users and uses.

#### **The Users**

Traditionally, bibliographic data is considered the content of a library catalog, union catalog, or abstract and indexing service, used to connect library users with resources that will fulfill their information needs. However, the stereotypical image of bibliographic data users is evolving along with the information environment. In addition to the traditional library end-user, bibliographic data is used by internal library staff, metadata developers, and commercial enterprises.

Traditional library end-users interact with bibliographic data via library catalogs, on-site or on the Internet, via Internet search engines, and via major indexing and abstracting tools. These users are as diverse as our population, with an equally diverse range of prior knowledge, research skills, and information needs. Our academic/research, special, school, and public libraries serve end-users who may range in age from toddlers to senior citizens, in knowledge from elementary school children to undergraduates to scholars, in computer skills from barely literate to expert programmers, and so on.

Library staff use bibliographic data in their daily work for a variety of purposes. Reference librarians interact with bibliographic data for research and in the creation of subject guides. Bibliographic data is used in collection management for inventory control and collection development, via reports based on the data. Catalogers consult external sources of bibliographic data in order to copy or create bibliographic records in their own catalogs, using shared services, such as OCLC, or the catalogs of other institutions. In addition, catalogers may browse or create reports from the bibliographic data in their own catalogs for purposes of quality control.

Museums and archives are increasingly focused on digitizing their collections for broader access to researchers and for preservation. These institutions are capturing bibliographic data in

metadata schemes using XML. Mapping of XML tags to MARC tags affords interoperability between the two data formats and sharing of data.

Other users of bibliographic data include commercial enterprises, such as Google and Microsoft. Google Scholar uses bibliographic data for finding and accessing items held in the collections of member libraries, while Google Book Search and Microsoft's Live Book Search use bibliographic data to identify digitized books. Developers of bibliography management programs access bibliographic data to auto-create citations for bibliographies. LibraryThing, a social cataloging Web site, uses existing MARC records from Web-based library catalogs to create catalogs for its users' personal collections.

### **User Expectations**

The Internet has altered forever our expectations for discovering, accessing and using information. In the past, when library collections comprised print, audio, audio-visual, etc., materials physically housed in the library building, bibliographic data helped satisfy a user's information need by making the right information in the right format discoverable. Today, information resources include licensed electronic resources, such as e-books, e-journals, and e-audio, as well as digital archives and the vast array of information and reference resources available via the Internet. While the mission still is to find the right information in the right format to meet the user's information need, both users and their expectations have changed in conjunction with the availability of new formats.

The traditional bibliographic access points of author, title, and subject now constitute a small proportion of the data that can be retrieved with full text keyword searching. Electronic journals and books allow users to search within the content for discovery of relevant articles and chapters. Full text database aggregators provide users with access to information across a plethora of resources with just one search. Relevant information buried within text has become more easily accessible.

The ability to search within content is best exemplified by search engines that crawl content on the Internet. On university and college campuses, students acknowledge that they begin their information seeking on the Internet, usually with Google, and then progress to library Web sites. (De Rosa & et al., 2005, pp. 142, A-17) Research indicates that Google's attraction is attributed to several factors that all relate to ease of use, including the following factors that relate to the underlying data (Markey, 2007):

- it takes little prior topical knowledge to get started, i.e., searches do not need to be highly targeted;
- the ability for users to make relevancy decisions from brief displays; and
- links take users directly to electronic full-text, if available.

In essence, Markey is stating that users do not need to know controlled vocabularies to conduct a successful search, there is enough content made available for an initial relevancy decision, and navigation between the search results and the desired resource is simple and fulfilling.

The popularity of searching Amazon.com for information resources is attributed to enhanced and user-added data.

Amazon relentlessly enhanced the [original bibliographic] data, adding publisher-supplied data such as cover images, table of contents, index, and sample material. Even more importantly, they harnessed their users to annotate the data, such that after ten years, Amazon, not Bowker, is the primary source for bibliographic data on books, a reference source for scholars and librarians as well as consumers. (O'Reilly, 2005, p. 3)

Amazon promotes user interaction with its resource metadata, a concept upon which social tagging Web services, such as Del.icio.us, are based. Allowing users to apply their own subject tags and to search for Web sites with subject tags supplied by other users is a growing Web phenomenon. Rather than depending on complex controlled vocabulary terms, users can supply simple, common, current language to identify Web pages of interest.

Faceted browsing is another feature of Internet information discovery that has gained in popularity because it helps guide the user to more specific search results. Faceted browsing is used with many Internet shopping Web sites, such as Barnes & Noble, Home Depot, and Wal-Mart. It lets the user begin with a broad category search, and then offers subdivisions within that category to narrow the search. The process continues until there are no narrower subdivisions or the user's information need is met. This allows metadata in the Web site to do the work for the user in information discovery, rather than forcing the user to arbitrarily generate search terms with the hope that they will work.

In sum, our information users have come to expect that information should be easy to discover (with one search and/or guided searching), easy to analyze for relevance, and readily available. Specific to bibliographic information, research indicates that users want information to be cross-searchable (i.e., not bounded by format or containers) and they want to interact with the metadata by appending their own tags and comments. (Calhoun, 2006, pp. 35, 40) Commercial users want access to the underlying metadata, as well as interoperability across data formats, so that the data can easily be incorporated into their programs. It remains to discern how bibliographic control should evolve to meet these user expectations and needs and to discover what other user needs we have not considered.

### **Bibliographic Data for the Future**

The foreseeable uses of bibliographic data can be grouped into three areas: discovery and delivery, inventory management, and cross-compatibility with related data. Some examples, by no means exhaustive, follow. Each example illustrates that a thoughtful approach to the bibliographic data that is recorded and how it is recorded will directly affect success in meeting the spectrum of user needs and uses in the future.

Toward providing richer discovery and delivery, IFLA's Functional Requirements for Bibliographic Records (FRBR) (1998) already offers a framework for user-centric bibliographic data. FRBR identifies four user goals: to find, identify, select and obtain information resources. In order to meet those goals, the underlying bibliographic data needs to be reorganized into a hierarchy of works, expressions, manifestations, and items, and re-evaluated to provide the appropriate attributes for each entity. FRBR's hierarchical structure would allow the bibliographic data to work harder for users in information discovery and delivery. For example, in a FRBR database a user can find a specific work by an author, under which the various expressions of that work might be clustered by translations into different languages through their related manifestations. As a result, a user interested in comparing Russian and German translations of a work is able to identify easily relevant German and Russian expressions through their respective published manifestations. In another scenario, a user may be satisfied in identifying a paperback edition of an expression, despite originally searching for the hardback edition. Upon selecting the appropriate manifestation, users are led to information necessary to obtain the physical item itself, whether it is in a library or a retail location, with the opportunity to navigate outside the database if necessary.

Using bibliographic data as facets is another way to enhance information discovery by making the data work for the user. To illustrate, searches conducted on subject headings can be subdivided by bibliographic data provided in the MARC record, such as discipline (classification number), author name, subfields within MARC subject tags, language, etc. Users are guided by the bibliographic data to narrower results until the information need is satisfied. North Carolina State University and the State University Libraries of Florida are experimenting with faceted browsing interfaces for their library catalogs. Moreover major library management systems are responding by beginning to develop software tools to support faceted display capabilities, along with relevancy display functionality. However, not all current bibliographic data is sufficiently atomized to maximize the benefits of faceted browsing: authors' first and last names, for example, are entered as a composite string in MARC format, meaning that first name cannot be a subfacet in a search on surname.

Bibliographic data is becoming increasingly important in the management of information. Information resources that have been placed in off-site storage would require extensive bibliographic data due to lack of accessibility, while mass digitization projects do not necessarily require such extensive data because the resources are readily available for visual comparison. Current book digitization projects by Google and Microsoft include author, title, publisher, pagination, and an image of the book cover. However, as digitization projects encompass copyrighted material, additional bibliographic data may be needed to identify use restrictions.

Library consortia also have information management issues that require thoughtful use of bibliographic data. For consortia with shared catalogs, some bibliographic data may be applicable to one library, but not another. For example, if all libraries in a consortium own the print edition of a journal, but only one library also subscribes to the electronic version, should the OpenURL for that e-journal be included in the shared MARC record?

Cross-compatibility of bibliographic data with other data formats will be important to allow sharing of bibliographic data across other services. User added content is becoming part of our common information landscape. How will social tagging and user reviews be integrated with bibliographic data?

The publishing industry has a data sharing initiative, Online Information Exchange (ONIX), which promotes book sales by providing booksellers with an online version of the information available on book covers and jackets, as well as related audio and video files. The data provided by ONIX is usable across various information platforms in the book trade. Recently, ONIX worked with the Joint Steering Committee for the Revision of AACR to find mutual agreement on categorization terms to be used in Resource Description and Access, allowing categorization terms to be consistent and interoperable between libraries and the book trade industry.(Delsey, 2006) What other data elements are used in ONIX or similar initiatives that could be exchanged productively with bibliographic data?

This brief background paper is not a comprehensive survey of the users and uses of bibliographic data for consideration in the future of bibliographic control. Rather, it is a starting point from which to generate discussion that may elicit additional information and perspectives.

#### References

- Calhoun, K. (2006). *The Changing Nature of the Catalog and its Integration with Other Discovery Tools* (Final Report): Library of Congress. Available at <http://www.loc.gov/catdir/calhoun-report-final.pdf>
- De Rosa, C., et al. (2005). *Perceptions of Libraries and Information Resources*. Dublin, OH: OCLC Online Computer Library Center, Inc. Available at <http://www.oclc.org/reports/2005perceptions.htm>
- Delsey, T. (2006). *Categorization of Content and Carrier* (Recommendations submitted for comment): Joint Steering Committee for the Revision of AACR. Available at <http://www.collectionscanada.ca/jsc/docs/5rda-parta-categorization.pdf>
- International Federation of Library Associations Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional Requirements for Bibliographic Records: Final Report*. Available at <http://www.ifla.org/VII/s13/frbr/frbr.pdf>
- Markey, K. (2007). The Online Library Catalog: Paradise Lost and Paradise Regained? *D-Lib Magazine*, 13(1/2). Available at <http://www.dlib.org/dlib/january07/markey/01markey.html>

O'Reilly, T. (2005). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Available at <http://www.oreilynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1>