Online Catalogs: What Users and Librarians Want

An OCLC Report

OCLC™

The world's libraries.
Connected.

Online Catalogs:
What Users and Librarians Want

An OCLC Report

Principal contributors

Karen Calhoun, Vice President, WorldCat and Metadata Services

Joanne Cantrell, Marketing Analyst

Peggy Gallagher, Market Analysis Manager

Janet Hawk, Director, Market Analysis and Sales Programs

Graphics, layout and editing

Brad Gauder, Creative Services Writer

Rick Limes, Art Director

Sam Smith, Art Director

Contributor

Diane Cellentani, Market Research Consultant to OCLC

Table of Contents

Executive Summary

An end user's expectations and work practices on the Web influence his or her decision to use a library online catalog. Catalog interfaces matter, but catalog *data quality* is also a driving factor of the catalog's perceived utility—and not only for end users, but also for librarians and library staff. To gain a rounded, evidence-based understanding of what constitutes "quality" in catalog data, OCLC formed a research team to:

- Identify and compare the data quality expectations of catalog end users and librarians

- Compare the catalog data quality expectations of types of librarians

- Recommend catalog data quality priorities, taking into account the perspectives of both end users and librarians.

Readers who are seeking to define requirements for improved catalog data (exposed in both end-user and staff interfaces) may find this report helpful as a source of ideas. The same is true for readers who have a part to play in contributing, ingesting, syndicating, synchronizing or linking data from multiple sources in next-generation library catalogs and integrated library systems.

Selected key research findings:

- The end user's experience of the delivery of wanted items is as important, if not more important, than his or her discovery experience.

- End users rely on and expect enhanced content including summaries/abstracts and tables of contents.

- An advanced search option (supporting fielded searching) and facets help end users refine searches, navigate, browse and manage large result sets.

- Important differences exist between the catalog data quality priorities of end users and those who work in libraries.

- Librarians and library staff, like end users, approach catalogs and catalog data purposefully. End users generally want to find and obtain needed information; librarians and library staff generally have work responsibilities to carry out. The

work roles of librarians and staff influence their data quality preferences.

• Librarians' choice of data quality enhancements reflects their understanding of the importance of accurate, structured data in the catalog.

The findings suggest two traditions of information organization at work—one from librarianship and the other from the Web. Librarians' perspectives about data quality remain highly influenced by their profession's classical principles of information organization, while end users' expectations of data quality arise largely from their experiences of how information is organized on popular Web sites. What is needed now is to integrate the best of both worlds in new, expanded definitions of what "quality" means in library online catalogs.

The report concludes with recommendations for a data quality program that balances what end users and librarians want and need from online catalogs, plus a few suggestions for further research.

Introduction

"A persistent shortcoming in the decision-making process [about library database quality] that needs to be addressed is the lack of serious research into user needs and benefits, and the actual impact on users of database quality decisions."[1]

What constitutes "quality" in catalog data has been reasonably well-understood by library professionals. The informed librarian's definition of catalog quality can be traced directly to Charles Cutter's statement in 1876 of the objectives of a library catalog;[2] these objectives have guided librarians' preferences for catalog design for over a hundred years. Thanks to Cutter and the theorists who followed,[3] today's library catalogs are founded on predictable and consistent record and heading structures, which facilitate serendipitous discovery, efficient known-item retrieval and many ways to browse. Library catalogs typically contain good metadata, in the sense that they use authority control, classification and content standards to describe and collocate related materials—all practices founded on Cutter's objectives for catalog searching by author, title or subject, or for distinguishing editions.

A study conducted by the OCLC Online Data Quality Control Section in 1989 confirmed that librarians' consensus about quality in their own library catalogs carried over to their expectations of WorldCat as a source of shared catalog records.[4] Carol Davis, then head of the Online Data Quality Control Section, found that librarians' top three quality concerns with WorldCat at the time were duplicate records (more than one record describing the same edition), incorrect (unauthorized) forms of name headings and incorrect (unauthorized) forms of subject headings. Today, OCLC's WorldCat quality program continues to center on managing these top three database quality priorities. Further, a glance into the contents of *OCLC Bibliographic Formats and Standards*[5] demonstrates one way in which Cutter's classical principles underpin present-day best practices in standardized description, consistent record encoding and authorized forms of names and subject headings.

Many writers have affirmed that Cutter designed his objectives of the catalog with the convenience of the user in mind. A similar motivation (the convenience of the user) underlies the "Functional Requirements for Bibliographic Records" (FRBR), a conceptual model based on the user tasks of finding, identifying, selecting and obtaining wanted information.[6] However, an examination of the literature turns up little evidence that Cutter, the distinguished theorists who followed him, or those who framed FRBR and "Resource Description and Access" (RDA)[7] rigorously tested their conceptual frameworks with information users. Fran Miksa, a well-known professor and librarian, has noted of Cutter's and early librarians' work to establish

traditions of information organization that "in the field of librarianship, user studies of a serious kind only begin in the 1920s."[8] Similarly, in the quote that begins this introduction, Janet Swan Hill enjoins librarians to clearly understand how and which library catalog data quality decisions help users find and obtain the information they need.

Information-seeking Behavior

In 2003, the *OCLC Environmental Scan* identified self-service, satisfaction and seamlessness as definitive of information seekers' expectations.[9] That report documented ease of use, convenience and availability as equally important to information seekers as information quality and trustworthiness. In 2005, the report *Perceptions of Libraries and Information Resources* looked further into people's information-seeking behaviors and preferences with respect to libraries, most notably revealing the trend of information seekers to begin a search for information with a search engine (84%) rather than on a library Web site (1%).[10] In addition to these examples of serious research into end-user information needs, a large body of research is available from the fields of communications, learning theory, sociology, psychology, consumer research, human-computer interaction and elsewhere. Social science researchers have investigated many paradigms in information-seeking research;[11] the "Principle of Least Effort," attributed to philologist George Zipf, is probably the best-known in libraries. A report from Marcia Bates to the Library of Congress (on improving user access to catalogs and portals) also contains a helpful review of the information-seeking literature.[12]

Donald Case, in his book on information seeking, points out that much research focuses on information sources (e.g., books or newspapers) and systems (e.g., catalogs) rather than on the needs, motivations and behavior of information users.[13] In other words, much research has emphasized information objects and systems over people. In contrast, usability experts have recognized the importance of designing systems contextually—that is, conducting "work practice" studies and using that information to drive information system design.[14] Librarians at the University of Rochester River Campus Libraries have taken the lead in applying work practice studies to library research questions.[15] An example is studying faculty research work practices to identify how scholars might use institutional repositories.

Catalog Use, Users and Data Quality

The recent library literature contains numerous articles on the need for change in online catalogs to better satisfy the expectations of information seekers who are accustomed to Web search engines, online bookstores and seamless linking to full text. Increasingly it is understood that an end user's expectations and work practices on the Web matter a good deal to whether he or she will use or revisit a library online catalog. In his August 2005 paper for the International Federation of Library Associations and Institutions (IFLA), John Byrum, Library of Congress, wrote of the need for library catalogs to provide access to more content and to offer significantly enhanced functionality based on the features of popular search engines.[16] Speaking of the limited scope of the catalog and its emphasis on print, Norm Medeiros,

Associate Librarian and Coordinator for Bibliographic and Digital Services at Haverford (PA) College, wrote, "More and more, users want, expect, and pursue full text. In increasing numbers they look past the catalog when searching for e-journals, databases and Web sites."[17]

At the same time, many of the users of library online catalogs are librarians and staff—these individuals form an important catalog user community themselves. Therefore, just as catalog end users (e.g., citizens, students and faculty) have information needs, preferences and expectations that need to be supported by catalog data, so do librarians who get their work done using the data underpinning the catalog.

Designing a New Database Quality Program

As next-generation library online catalogs emerge, and as OCLC makes decisions about how to shape its next-generation WorldCat data quality program, it is essential to gain an evidence-based, user-centered understanding of what catalog data "quality" is to the various communities (both end users and librarians) for whom library online catalogs are or can be an important, frequently visited information resource. To this end, OCLC formed a research team of Karen Calhoun, a cataloging/metadata expert; a professional from an independent market research firm; and members of OCLC's Market Analysis team to:

- Identify and compare the data quality expectations of catalog end users and librarians

- Compare the catalog data quality expectations of types of librarians

- Recommend catalog data quality priorities, taking into account the perspectives of both end users and librarians.

While many of the research team's findings relate specifically to defining a WorldCat data quality program, many findings are generalizable to the data supporting the current generation of library online catalogs and integrated library systems (ILSs). This report describes those generalizable findings. Readers who are seeking to define requirements for improved catalog data (exposed in both end-user and staff interfaces) may find this report helpful as a source of ideas. The same is true for readers who have a part to play in contributing, ingesting, syndicating, synchronizing or linking data from multiple sources in next-generation library catalogs and integrated library systems.

Notes

1. Hill, Janet Swan, "Is it worth it? Management decisions related to database quality," *Cataloging and Classification Quarterly* 46 (1) (2008): 5–26.

2. Miksa, Francis L., *Charles Ammi Cutter: Nineteenth-Century Systematizer of Libraries*, (dissertation, University of Chicago, 1974) p. 373, http://www.ischool.utexas.edu/~miksa/ dissertation.html. Cutter's 1876 objectives of the catalog are quoted in Miksa's comprehensive study of Cutter's contributions to librarianship:

- To enable a person to find a book of which the author, the title or the subject is known.
- To show what the library has by a given author, on a given subject or in a given kind of literature.
- To assist in the choice of a book as to its edition (bibliographically) or as to its character (literary or topical).

3. A comprehensive review of past and present cataloging theory may be found in *The Future of Cataloging: Insights from the Lubetzky Symposium*, April 18, 1998, University of California, Los Angeles, edited by Tschera Harkness Connell and Robert L. Maxwell, Chicago: American Library Association, 2000.

4. Davis, Carol C., "Results of a survey on record quality in the OCLC database," *Technical Services Quarterly* 7 (2) (1989): 43–53.

5. *OCLC Bibliographic Formats and Standards,* 4th edition, http://www.oclc.org/bibformats/en/.

6. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional requirements for bibliographic records: final report,* UBCIM publications (München: K.G. Saur, 1998) new ser., v. 19.

7. Joint Steering Committee for the Development of RDA, *Full Draft of RDA,* 2008, http://www. collectionscanada.gc.ca/jsc/rdafulldraft.html.

8. Miksa, Fran, "Information organization and the mysterious information user" (lecture delivered at the School of Information, University of Texas—Austin), forthcoming in *Libraries and the Cultural Record* 4 (4) (Fall 2009).

9. De Rosa, Cathy, Lorcan Dempsey and Alane Wilson, *The 2003 OCLC Environmental Scan: Pattern Recognition: A Report to the OCLC Membership* (Dublin, OH: OCLC Online, 2003).

10. De Rosa, Cathy, *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership* (Dublin, OH: OCLC, 2005).

11. Case, Donald O., *Looking for information: a survey of research on information seeking, needs, and behavior (*San Diego CA: Academic Press, 2002) p. 140. The Principle of Least Effort asserts that information seekers will use information sources that are the easiest and most convenient, even when better or more authoritative sources are available, but not as easy or convenient to use or find.

12. Bates, Marcia J., "Improving user access to library catalog and portal information: final report," prepared for the Library of Congress, 2003. http://www.loc.gov/catdir/ bibcontrol/2.3BatesReport6-03.doc.pdf. The section on information-seeking behavior is on pages 4–6.

13. Case, p. 6.

14. See for example Holtzblatt, Karen, and Hugh Beyer, "Contextual design: using customer work models to drive systems design," in *Conference companion on human factors in computing systems (*New York NY: ACM Press, 1996) p. 373–374.

15. Foster, Nancy Fried, and Susan Gibbons, "Understanding faculty to improve content recruitment for institutional repositories," *D-Lib Magazine* 11 (1) (2005), http://www.dlib.org/dlib/january05/ foster/01foster.html.

16. Byrum, John D., "Recommendations for urgently needed improvement of OPAC and the role of the National Bibliographic Agency in achieving it" (paper prepared for the 71st IFLA General Conference and Council, Oslo, Norway, August 14–18, 2005), http://www.ifla.org/IV/ifla71/ papers/124e-Byrum.pdf.

17. Medeiros, Norm, "Driving with eyes closed: the perils of traditional catalogs and cataloging in the Internet age," *Library Computing* 18 (4) (1999): 300–305.

Methodology

The OCLC research team employed three methods to identify the catalog data quality expectations of catalog end users and library staff: focus groups, a pop-up survey on WorldCat.org and a Web-based survey. WorldCat.org, OCLC's freely available end-user interface on the Web, provided the means for the team to study focus group and pop-up survey participants' reactions to the data elements underlying a recently designed library online catalog. The third method, the Web-based survey, targeted librarians and library staff accessing WorldCat via both end-user and staff interfaces—that is, not only WorldCat.org and WorldCat on FirstSearch, but also Connexion and Z39.50 access to OCLC Cataloging (chiefly used by catalogers), and WorldCat Resource Sharing (generally used by interlibrary loan staff).

In the study of librarian and library staff expectations, it was appropriate to consider all interfaces to WorldCat data, since the primary research interest centered not on a particular interface but on librarian/staff preferences around catalog *data quality*. The team's interest in the particular interface used to access the data was secondary; instead the goal was to gain an evidence-based, user-centered understanding of the data quality needs of a variety of subcommunities of librarians and library staff. To the extent that WorldCat and library catalogs are comparable, the methodology's focus on data quality, rather than interface, permits insights gained about WorldCat to be germane to the data quality requirements of library online catalogs and integrated library systems in general.

End-user Focus Groups

OCLC commissioned Blue Bear LLC to facilitate three qualitative focus groups during May 2008 in Columbus, Ohio. Blue Bear's facilitator conducted one session each for the following groups: undergraduates ages 18 to 24, casual searchers ages 25 to 59, and faculty and graduate students (referred to as scholars in this report). The three groups included the following participants:

Undergraduates	Casual Searchers	Scholars
Participant demographics	**Participant demographics**	**Participant demographics**
• 8 participants	• 8 participants	• 8 participants
• 18 to 24 years of age	• 25 to 59 years of age	• 25 to 59 years of age
• 7 undergraduates; 1 recently graduated from a university with a bachelor's degree	• General Internet users	• 3 graduate students and 5 faculty
• 6 have used a library's Web site	• Occupations included basketball coach, public speaker, teacher, disc jockey, business professional and author	• Graduate student disciplines included computer science & linguistics, school counseling and education
• 6 have a current library card	• 5 have used a library's Web site	• Faculty disciplines included German studies, political science, communications, economics & finance and sociology & anthropology
	• 5 have a current library card	• All had published work or assigned a paper with 10+ citations
Online usage	**Online usage**	**Online usage**
• 6 participants spend 6–10 hours per week and 2 spend 1–5 hours per week online (excluding e-mail)	• 6 participants spend 10+ hours per week online (excluding e-mail); 2 spend 6–10 hours per week	• Half spend 10+ hours per week online (excluding e-mail); 3 spend 6–10 hours per week, and 1 spends 1–5 hours per week online
• All have used an eCommerce site; textbooks.com, Amazon, Yahoo!, Bestbuy.com and eBay were most frequently mentioned sites	• All have used an eCommerce site; Amazon, eBay, Barnes & Noble and half.com were most frequently mentioned sites	• All have used an eCommerce site; Amazon and Barnes & Noble were mentioned by all participants

Research Protocol

Participants sat at a networked laptop using software and connections that allowed the research team to capture Web metrics and comments/page mark-ups for immediate viewing and for post-group analysis. Images of screen shots with some of these comments are presented in the findings.

Rating symbols (see the image below) associated with all comments were captured during the focus group session.

The green plus sign means they liked what they saw and the comment is positive; the red minus sign means they did not like what they saw and the comment is negative.

The groups themselves were modular with online activity followed by offline discussion. The sessions focused on:

- **Task 1:** Online exploration of WorldCat.org with users creating searches of their own design
- **Task 2:** Online search on WorldCat.org for a specific book, *Guns, Germs, and Steel*
- Group discussion of their experiences with Tasks 1 & 2, including walk-through of WorldCat.org
- **Task 3:** Online search on WorldCat.org for material specific to their key interest (undergraduates and casual searchers) or specialty (scholars)
- Group discussion of their experiences with Task 3
- Closing discussion to gauge overall reactions

End-user Pop-up Survey

OCLC commissioned ForeSee Results to conduct an online pop-up survey targeting worldwide end users of WorldCat.org, OCLC's freely available interface to WorldCat on the Web. ForeSee collected a total of 11,151 total responses between May 12, 2008 and July 9, 2008, with a 4% response rate based on the number of times the survey displayed to users.

The online survey popped up when a respondent clicked through to the Detailed Record in WorldCat.org. The survey included 25 closed-end questions and one open-end question. One of the questions asked respondents to describe themselves ("Which best describes you?"). Answers to this question enabled division of the responses into groups self-identified as students, teachers/professors, other general searchers and librarians/library staff. To capture the perceptions of end users only, responses from the librarians/library staff group were excluded from the end-user data. Of the 11,151 total responses, 68% were from end users.

The end users described themselves as follows:

Age	
18 or below	5%
19–30	24%
31–40	17%
41–50	20%
51–60	20%
61 or over	13%
Prefer not to answer	1%

Total End Users	7,583
Students*	28%
Teacher or professor	22%
Business professional	19%
Other	16%
Retired	9%
Healthcare professional	6%

*Type of Student	
Graduate college/Post-graduate	54%
Undergraduate college/Post-secondary	30%
High/Secondary school	10%
Middle/Intermediate/Junior high school	3%
Other	2%
Elementary/Primary school	1%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey)

Library Survey

OCLC commissioned Marketing Backup, an independent marketing research firm, to conduct a Web-based survey targeting librarians and library staff who use the WorldCat database from a variety of interfaces.

OCLC mailed an invitation to participate in the survey to 3,522 OCLC members from academic, public and special libraries around the world. In order to collect the perspectives of all who use WorldCat data, the research team divided the mailing into three equal subsets. Library directors, heads of cataloging and heads of public services received invitations. OCLC asked recipients to forward the invitation to all who use WorldCat data, whether for collection development, acquisitions, cataloging, resource sharing, reference or other work in their libraries.

A total of 1,397 librarians and library staff responded to the survey between September 10, 2008 and November 30, 2008. Since multiple respondents from one institution were invited to participate in the survey, a response rate cannot be calculated.

Respondents self-identified their job responsibilities from a list of roles—library director/dean, public services/reference, cataloging, acquisitions and interlibrary loan. They could choose more than one role, and the majority (58%) did; 42% reported that they play only one role in their libraries. Throughout the report the term librarians is used to represent librarians and library staff.

Library Type	
Academic	52%
Public	25%
Special	18%
Other	5%
Job Responsibilities (Selected all that apply)	
Cataloging	64%
Public Services/Reference	44%
Collection Development	31%
Acquisitions	28%
Interlibrary Loan	24%
Library Director/Dean	21%
Region	
Inside North America	74%
Outside North America	26%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Throughout the report significant differences are reported. A statistically significant difference is a difference in responses that, based on statistical tests, are not due to chance.

Data Quality:
What End Users Want

Key Findings

The end user's experience of the delivery of wanted items is as important, if not more important, than his or her discovery experience. Appropriate, accurate and reliable data elements supporting the transitional experience from discovery through delivery are critical.

End users expect a seamless flow from discovery to delivery; end users want to know immediately if the item is available and if so, how to get the item. For online materials, end users want more direct links or easier access to the online content, both text and media.

Discovery-related information elements beyond author and title, such as summaries, excerpts and tables of contents, are essential aspects connecting the stages of an end user's discovery-to-delivery experience.

While conducting their searches, end users value and expect evaluative information to assist their discovery, and ultimately, the delivery of materials.

Libraries need to make it easier for end users to quickly ascertain whether items meet their needs; for analog items, the available data needs to help users decide if it is worth their time to obtain the items—most often by going to the library.

Search results must be relevant and the relevance must be obvious.

End users have expectations of the types of results they should get when they conduct searches. They want the library catalog to return relevant results based on those expectations. Behind the scenes, the catalog needs supporting data elements and ranking algorithms that make it obvious to end users why search results are returned and the order in which they are returned.

Keyword searching is king, but an advanced search option (supporting fielded searching) and facets help end users refine searches, navigate, browse and manage large results sets.

End users want to be able to do a simple Google-like search and get results that exactly match what they expect to find. At the same time, they appreciate the ability to conduct advanced, guided searches so that they do not have to scroll through pages of results to find the items they need. They view faceted narrowing of results as an effective way to reduce what may be unmanageable results lists.

Detailed Observations

Overview

The research team collected end-user findings via two methods: focus groups and an online pop-up survey. Three different segments of end users participated in focus groups: undergraduates ages 18–24, casual searchers ages 25–59 years old, and faculty and graduate students (called scholars throughout this report). The online survey randomly popped up as end users were searching in WorldCat.org.

Most essential data elements

When asked what information in the records was most essential in identifying the items they need, end users responding to the pop-up survey chose a blend of data elements that support the user's delivery experience, as well as some data elements that support the discovery experience. The elements related to delivery are represented by the horizontal striped bars in the chart below; the elements related to discovery are represented by the solid bars.

Most Essential Data Elements*

*What information is **most** essential in helping you identify the item that you need?*

Base: End-user pop-up survey respondents

Data element	Percentage
A list of libraries that own the item	24%
The ability to see what is immediately available	14%
Author	12%
Item details	7%
Links to online content/full text	7%

Legend: Delivery-related element (striped) — Discovery-related element (solid)
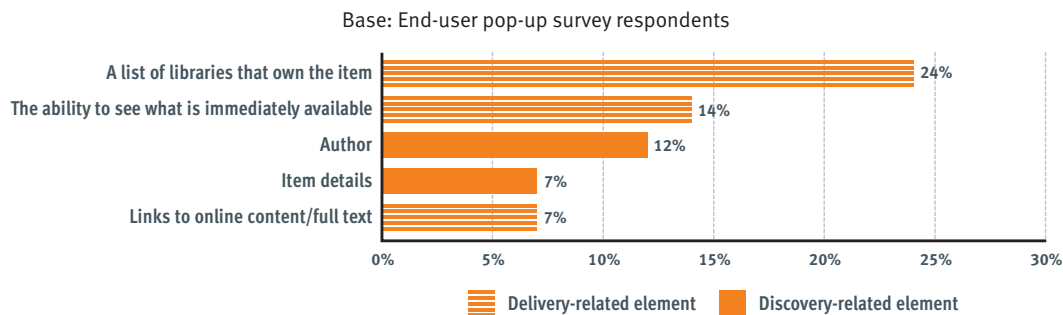
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey)

Three of the end users' top five choices included data elements supporting delivery: 24% of the respondents said they want to see the *list of libraries that own the needed item;* 14% said they want to know *what's immediately available*; and 7% want *links to online content/full text.*

*Note: Title—the ubiquitous choice—was excluded in order to focus more attention on other data elements.

Two of the end users' top five choices were discovery-related data elements: *author* (12%) and *item details* (7%).*
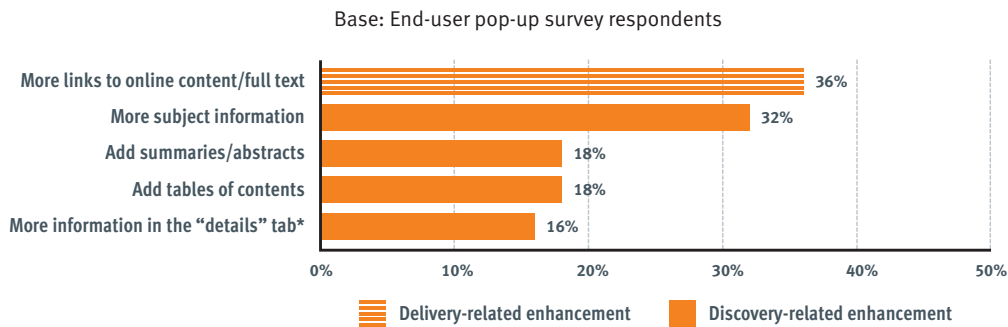
When specifically asked what data elements were most useful in finding the items they need, focus group participants primarily mentioned discovery-related data elements such as the ability to preview the book, the cover art, the ability to refine a search by various criteria and summaries/abstracts. Throughout the discussion group sessions, however, participants spoke of delivery-related data elements as well. One participant noted that we are living in a "buy it now, get it now" world of instant access to electronic materials. This is the reality that end users expect from libraries: the links that connect them from the metadata *describing* online content to the content itself.

Desired data quality enhancements

When asked what enhancements to a catalog could help them consistently find wanted items, more than one-third (36%) of the pop-up survey respondents answered that having *more links to online content/full text* would be the enhancement they would want to see. As was evident in the responses to the earlier question about most essential data elements, this question about suggested enhancements leads to the same finding: both discovery-related and delivery-related information elements are important to end users.

Top Five Desired Data Quality Enhancements

*What changes would be **most** helpful to you in identifying the item that you need?*

Base: End-user pop-up survey respondents

More links to online content/full text	36%
More subject information	32%
Add summaries/abstracts	18%
Add tables of contents	18%
More information in the "details" tab*	16%

Delivery-related enhancement   Discovery-related enhancement

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey)

As for the rest of the most frequently chosen enhancements—from *more subject information* to *more information in the details tab*—respondents appear to be expressing the desire for data elements not generally included in a standard catalog description. The desired data elements may be structured or unstructured, but they need to help end users assess the utility of items in a results set and decide which ones merit taking the time to obtain.

*At the time of the pop-up survey, the WorldCat.org "details" tab contained basic bibliographic information plus enriched data such as table of contents and summary/abstract, if available.

Discovery

The next two sections present more detailed observations from the end users about their discovery and delivery experiences. This section will discuss discovery. Interspersed in the reporting of findings are discussions of the differences among subgroups of end users—casual searcher, undergraduate and scholar focus group participants and pop-up survey respondents.

Searching

Focus group participants discussed their experiences in searching, from the relevancy of the search results to the tools used to assist in their search process, such as advanced and faceted search features.

Relevancy of search results. The findings suggest that end users' preferences and expectations are increasingly driven by their experiences with search engines like Google and online bookstores like Amazon. When end users conduct a search in a library catalog, they expect their searches to find materials on exactly what they are looking for; they want relevant results.

During the focus groups, when participants searched for known items, their search strategies were often successful in returning relevant search results. Thus, they could easily find the items they wanted. One participant noted that the exact item she was looking for was the first search result. "It was my #1 choice," said a scholar participant, referring to the search for a known item. Another participant excitedly explained how the search not only resulted in what was the correct item but also found new things: "I looked for something familiar and found a more comprehensive list than from Amazon ..."

When searching for known items, focus group participants were usually successful in finding them; the search strategy was often as simple as entering the title and/ or author. It was when participants conducted general searches on a topic (i.e., searches for unknown items) that they expressed dissatisfaction when items unrelated to what they were looking for were returned in the results list. End users may not understand how to best craft an appropriate search strategy for topic searches. The findings indicate that end users expect to type in a few words and have the catalog return the items they want; they expect the catalog to "know" what they are looking for based on the terms they type in the search box. Additionally, if the words they use in their searches have multiple meanings depending on the context, they still expect their searches to return appropriate materials on exactly what they want. As noted above, their experience with search engines and even the use of natural language in searching appears to have influenced their expectations. As one end-user survey respondent requested, "Make it as easy as a Google Book Search ..."

During the focus groups, participants discussed their expectations that catalogs would use weighting in the search algorithm toward common uses of popular terms to ensure that the catalog would identify needed items. In situations when this weighting is not obvious, the users explained that they feel that the catalog has

failed to meet their expectations. One user discussed a search he had conducted for "running" and the search returned results on running a business and several novels with the word "running" in their titles. He was expecting results for books and materials on the sport or exercise of running, so these results surprised him. He felt that the catalog should have understood the common use of the term "running" and have been weighted toward the sport or exercise.

Improving the relevance of search results is an interesting data quality problem whose solution goes well beyond the boundaries of the types of metadata that catalogers have been responsible for supplying, obtaining, managing or mining. Yet it is clear that end-user expectations are likely to (or at least, should) drive new developments in library catalog search relevance capabilities.

Advanced search. Supporting an advanced search feature involves indexing bibliographic data elements separately—title, author, subject, format, publication date, and so on. The effective operation of advanced search features may also involve controlled vocabularies to ensure, for example, that all of the materials authored by Ernest Hemingway are retrieved by a query on Hemingway's name.

End-user focus group participants used the advanced search feature of WorldCat.org particularly when searching for known items. During the focus groups, a single search box appeared helpful for users conducting general searches (i.e., unknown items). However, when a searcher was looking for a known item, the ability to be more precise in the search was more likely to ensure successful search results. Advanced search functionality allowed the users to more quickly identify the needed item in the search results because the results were more precise.

> **"Always nice to be able to specify—or include all info to quicken search."**

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Focus groups)

During the focus groups, some participants in the scholar group initially looked for more functionality beyond the basic keyword search box. The facilitator discussed with these participants their expectations in building a search strategy and learned from them that prominently placed advanced search functionality is desirable as an optional search method, preferably with some guidance on how to use it. The participants understood the basics of an advanced search feature but seemed uncertain about its full capabilities.

Refine your search. Provided the data are appropriately indexed, preexisting fields in a catalog database such as author, publication year and format can be extracted and presented (usually in the left frame of the Web page) to permit faceted navigation, also called faceted search or faceted browsing. Faceted search supplements (and obviates the imprecision associated with) the popular direct search method, i.e., typing keywords in a search box, that is characteristic of keyword searches. Facets enable searchers to progressively narrow their choices from a large and often unwieldy retrieval set. Librarians have embraced faceted search in next-generation catalog interfaces, as the underlying data elements are readily available in databases created using library cataloging standards and traditions. These findings suggest that facets are another effective application of the controlled terminologies and authorized heading forms that have been built and maintained by librarians.

Focus groups participants found faceted browsing in WorldCat.org to be a "very comprehensive way to refine" their search results. While reviewing a short results list is manageable, keyword searches often return hundreds of results, making faceted browsing a useful tool to quickly narrow the choices—"You need [to] refine your search when you get hundreds of results, but not when you get 15," said a participant from the undergraduate group.

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Focus groups)

Some participants in the focus groups were less accustomed to using faceted browsing and, therefore, did not immediately understand the functionality. Others, however, immediately understood how to use the faceted browsing option once they saw it, and what results it should return to them. This was especially apparent among the participants in the undergraduate and casual searcher groups. Additionally, they indicated a strong intent to use the faceted browsing in the future.

Discovery-related information elements

Focus group participants discussed information elements that are particularly valuable in determining which item(s) they want from a search results list. Pop-up survey responses and focus group findings were consistent as to what information elements are helpful in identifying the items needed. Both groups of respondents rely on and expect enhanced content, including summaries/abstracts and tables of contents.

In addition to enhanced content, the findings suggest that social features are important information elements to some end users. In particular, reviews and ratings, depending on the credibility of the author of the reviews and ratings, are desirable discovery-related elements to certain segments of user populations.

Summaries/abstracts. Summary information was well-received by focus group participants in all three groups and was also frequently the topic of suggestions among pop-up survey respondents as a way to improve the catalog. Most focus group participants felt that the actual summaries need to be highly visible when searching, ideally in the search results, as well as on the single-record display page. One survey respondent made the following suggestion: "I wish the results page would list a short blurb (one line) about the book similar to the way Google shows you a tiny bit about what a site link is about."

The findings suggest that summaries are most important in searches for unknown items; focus group participants and pop-up survey respondents expressed interest in wanting a quick verification to determine if it is worth their time to even look at an item. When probed about what information was most essential in finding the wanted item, a participant from the scholar focus group commented, "Honestly, nothing really jumps out at me but probably I think the summary. Then I can go from there to decide to get it." One survey respondent echoed that sentiment: "Please include a description of what the item is about so viewers can know if it is worth their time to get this book out of the library before they go."

Other discovery-related results

Focus group and pop-up survey responses mirrored one another with respect to their choice of essential data elements for discovery—including item details, summaries/abstracts and tables of contents. Because the focus group sessions were interactive, it was possible to also explore the role of cover art, excerpts, reviews, ratings and user-contributed lists as aids to the discovery and selection of wanted items.

Cover art. Focus group participants from all three groups reacted positively to cover art displays. Positive comments such as "I like to see a picture of the book," "I like the cover image being prominent" and "Like to see what the book looks like" were made in all focus group sessions. The participants in the undergraduate group were the most vocal on their interest in seeing an image of the book's cover. Participants discussed that they see value in having cover art appear in the search results as well as in the individual item's record.

Excerpts. Focus group participants from all three groups indicated a desire to sample the items returned in the search results. They appeared to want to review sample content online; as they discussed this desire, they did not limit the samples to only books. They also discussed the desire to sample content for other media, such as audio and video files.

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Focus groups)

Related subjects and subject categories. Participants from the undergraduate focus group liked the related-subject information as a way of determining the usefulness of the item. In a WorldCat.org single record display, the "related subjects" label precedes the controlled subject terms assigned to the title. When displayed in the left frame as facets, many focus group participants found the subject categories listed there helpful for refining searches and for browsing.

Social features. The focus group participants offered a mixed reaction to social features, such as the ability to create reviews and share lists. Participants in the undergraduate focus group were more favorable toward user-contributed content, and they were quite discerning in their ability to distinguish authoritative from non-authoritative reviews.

Overall, editorial reviews were considered more valuable than end-user-contributed reviews; this opinion was particularly evident when the focus group participant's information need was for academic or professional purposes.

User-contributed reviews were at least somewhat interesting to some participants in all of the focus groups when they were not conducting academic searches. Scholars were more interested in professional reviews and were less interested in other users' opinions on the materials unless they were "experts."

While some participants in the casual searchers group felt the reviews were "boring," one participant mentioned he loves to read reviews due to his profession (basketball coach); "… we learn from each other."

Participants in the undergraduate group said reviews would be helpful in deciding what item is needed. Not all felt it was necessary to have "professional" reviews,

though they would like some indication of who wrote the review. They felt everyone has different tastes/opinions, and these opinions are valuable. While not all agreed that professional, editorial reviews were *necessary*, they all agreed that they could be useful.

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Focus groups)

The participants in the undergraduate group also found value in user-created lists, which provide a way for end users to keep track of and share items using WorldCat.org. These lists allow end users to refer to items of interest, and they can share the lists with friends and colleagues. These participants felt it would be beneficial to be able to see which items others are using and for what purpose; they felt that it would add credibility to an item to know who was using it.
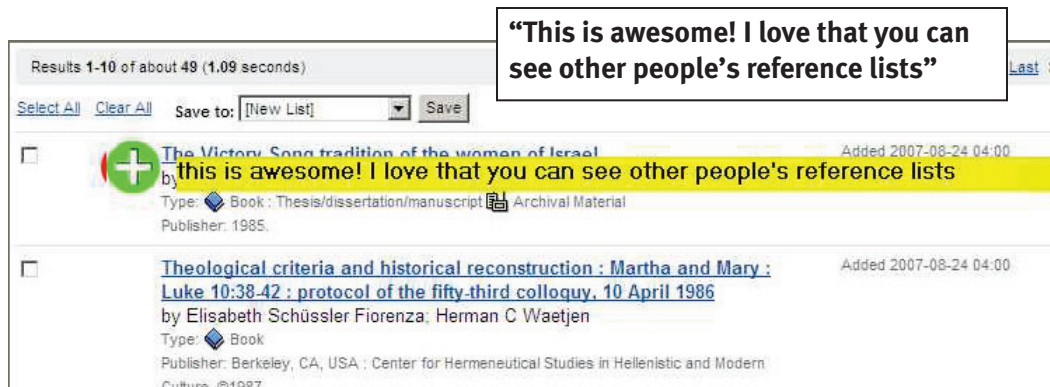
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Focus groups)

Participants in the casual searchers and scholar groups did not perceive user-created lists to be useful; one participant in the casual searchers group described user-created lists as "… a bunch of random people who read the book and put it on a list somewhere … not the least bit helpful in conducting research…" These participants would like to be more informed about the author/creator of the lists.

Delivery

Discovery is important, but delivery is as important, if not more important, than discovery. The findings suggest that a seamless, easy flow from discovery through delivery is critical to end users. This point may seem obvious, but it is important to remember that for many end users, without the delivery of something he or she wants or needs, discovery alone is a waste of time.

The pop-up survey results confirmed the importance of delivery as the goal of most searches. Presented with a list of what they would do with the information found by searching WorldCat.org, most pop-up survey respondents chose options related to delivery; 30% would *request the item from a library* and 21% would *visit a library listed*.

What Will Be Done with the Information

What will you do based on the information you found?

Base: End-user pop-up survey respondents

Response	Percentage
Request the item from a library	30%
Visit a library listed here	21%
Cite the item	12%
I did not find the item I was looking for	8%
Other	7%
Save to my favorites	6%
Nothing	5%
Share the item with a friend	4%
Purchase the item	4%
Write a review	3%
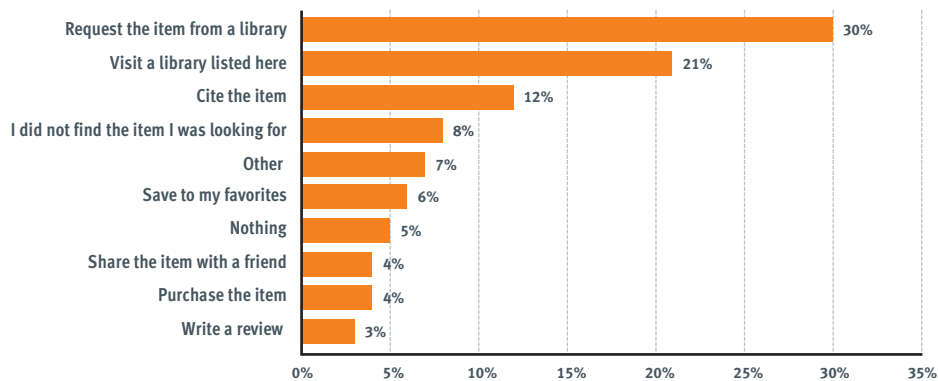
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey)

Item ownership and availability. The findings indicate that end users want to know if an item is available, preferably at the point of sorting through the search results. Beyond just being helpful, the delivery-related information was the most critical for some focus group participants. A couple of the participants in the casual searchers group went as far as to suggest that if an item is not available, it should not even be included in the results list. As one participant put it, "Why put a result there… if you cannot get it?" A participant from the undergraduate group wanted the ability to set up a "preferred library affiliation" account and have search results display accordingly.

These research results suggest that much effort could be usefully invested in determining how and what data elements might be improved to make the delivery experience for all types of materials—print, licensed and digital—as positive as possible for end users.

Access to online content. The findings emphasize the need for more direct links and easier access to online content from library catalogs, both for reading online and downloading. Delivery, and in fact, immediate delivery is what users have come to expect based on their experience with various downloadable media: "… great concept but in 2008, there has to be an immediate fix!! When you buy it now, get it now. eBooks or MP3 downloads are a must," summed up the need for immediate gratification for information by one participant in the casual searchers focus group.

From a data quality perspective, the need for easy access to online content from library catalogs is likely to require increased investment in linking metadata management and interoperability with third-party data.

Data Quality:
What Librarians Want

Key Findings

Librarians and library staff, like end users, approach catalogs and catalog data purposefully. End users generally want to find and obtain needed information; librarians and library staff generally have work responsibilities to carry out using catalog data. The data quality preferences of librarians and staff are driven primarily by their work assignments.

At a glance…

What do librarians want?

- Duplicate records merged
- Typographical errors fixed
- Brief records upgraded
- Evaluative content, such as tables of contents, summaries/ abstracts and cover art

Librarian and library staff work assignments generally fall into two broad types: those assignments involving direct contact with end users of the library, and those that predominantly support library operations, such as selecting or acquiring new materials for the library. The type of work role appears to have a good deal of influence on a librarian's or library staff member's priorities for catalog data quality.

Duplicate records (multiple records for the same edition or manifestation) impede the work of librarians and staff.

The merging of duplicate records in WorldCat was the top priority of all types of librarians, from all types of libraries, inside or outside North America. One respondent captured the general sentiment saying, "I would get rid of all the duplicate records! I spend a lot of time trying to determine why records are different and which one I should use. If the duplicates were gone, this would make my job a lot easier." Merging, removing or better managing duplicate records would appear to be a top priority for improving the effectiveness and efficiency of library catalogs for the work carried out by librarians and staff.

Librarians and staff place priority on enriching catalog records with tables of contents data.

From a long list of potential data quality enhancements, respondents consistently selected *adding tables of contents to bibliographic records* as their second choice (behind merging duplicates). This consistency held across nearly all categories of respondents.

Except for the priorities to merge duplicate records and add tables of contents, significant differences exist in the data quality priorities of librarians by work role, type of library and region.

Inside the library community, expectations and preferences for catalog data quality enhancements appear to be dependent on a variety of factors. There was no common third or subsequent choice for catalog data enhancement across respondents with different work roles, from different types of libraries, or from libraries inside or outside North America. For example, the third most-chosen enhancement for respondents from libraries outside North America was *adding more records for non-English materials;* this enhancement did not make it into the top ten list of North American library respondents. *Upgrading brief bibliographic records* was the fourth most-chosen enhancement by respondents from academic and public libraries, but this enhancement did not make it into the top ten list for special libraries.

Detailed Observations

Overview

The research team collected data from 1,397 librarians and library staff via a Web-based survey; 74% of the responses came from from North American institutions and 26% came from institutions in other countries. The Web-based survey targeted librarians and library staff accessing WorldCat via both end-user and staff interfaces—that is, not only WorldCat.org and WorldCat on FirstSearch, but also Connexion and Z39.50 (chiefly used for cataloging on WorldCat), and WorldCat Resource Sharing (generally used by interlibrary loan staff).

Librarian and staff survey respondents identified themselves by type of library, country and job role (director, acquisitions, cataloging, collection development, resource sharing and reference). Survey respondents could choose more than one job role. The analysis presented in this chapter compares responses across these categories. Significant differences in responses between these categories are noted throughout this report. A statistically significant difference is a difference in responses that, based on statistical tests, are not due to chance.

The library study's purpose was to evaluate what data elements librarians and library staff find most helpful for identifying items in WorldCat and to identify the enhancements that librarians and staff require for their work with WorldCat.

The survey research centered not on a particular WorldCat interface but on librarian/staff preferences around WorldCat *data quality*. This report presents the findings about WorldCat data quality enhancements. To the extent that WorldCat and library catalogs share similar data quality issues (duplicate records, for example), the findings of this study are applicable to the data quality requirements of library online catalogs and integrated library systems in general.

Desired Data Quality Enhancements

Desired data quality enhancements by total library survey respondents
To try to compile a comprehensive list of data quality enhancements, reflecting as

many points of view as possible, the survey designers looked into the results of many studies of catalog data. The exercise yielded 18 choices to present to survey respondents, who were asked to choose a favorite, plus other enhancements from the survey's list of choices. The chart below gives a ranked order of the choices from all library survey respondents. *Merging duplicate records* (52%), *adding tables of contents to detailed bibliographic records* (40%) and *adding summaries/abstracts to detailed bibliographic records* (28%) are the top three recommended data quality enhancements from all library survey respondents.

Desired Data Quality Enhancements

Which of the following enhancements would you recommend?

Base: Library survey respondents

Enhancement	%
Merge duplicate records	52%
Add tables of contents to records	40%
Add summaries to records	28%
Fix typos	27%
Upgrade brief records	25%
Add cover art to results	25%
Make it easier to correct records	25%
Fix MARC coding errors	24%
Add summaries to results	24%
Increase accuracy of library holding information	24%
More records for online resources	22%
Add more formats	22%
More records for non-English materials	21%
More clickable links to online content	18%
Enable more libraries to make corrections	15%
Add excerpts to the records	11%
Add support for multilingual searching/record displays	10%
Greater exposure of holdings on the Web	8%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Reasons for top desired enhancements by total library survey respondents

The survey prompted respondents for their reasons for choosing a favorite enhancement.

Respondents who would recommend *adding table of contents to detailed bibliographic records* stated it would **provide enough detail for librarians and users to identify the item** (53%) or it would **enhance searching** (24%), based on 133 comments.

More than two-thirds (68%) of those who would recommend *adding summaries/abstracts to detailed bibliographic records* stated the **summaries help librarians and end users identify the needed source**, based on 82 comments.

Desired Data Quality Enhancements by Category of Respondent

The ranked list of all respondents' data quality enhancement choices is weighted toward the demographic composition of the library survey respondents, over half of whom come from academic libraries. Further, nearly three-fourths of the respondents come from North American libraries, and nearly two-thirds have job duties associated with cataloging. An examination of the data quality enhancement choices by category of respondent yields some interesting differences across every category—type of library, job role and region.

Desired data quality enhancements among respondents by library type

The analysis revealed several significant differences among survey respondents by library type. The following table lists the enhancements that were among the top ten enhancements chosen by respondents by at least one of the three types of libraries included in this study: academic, public and special libraries.

Desired Data Quality Enhancements
Which of the following enhancements would you recommend?

By library type

Enhancement	Total		Academic		Public		Special	
	Rank	%	Rank	%	Rank	%	Rank	%
Merge duplicate records	1	52%	1	57%	1	48%	1	56%
Add tables of contents to records	2	40%	2	46%	5	27%	2	44%
Add summaries to records	3	28%	3	29%	6	27%	3	28%
Fix typos	4	27%	5	28%	3	29%	4	27%
Upgrade brief records	5	25%	4	29%	4	29%	14	17%
Add cover art to results	6	25%	11	23%	2	35%	12	18%
Make it easier to correct records	7	25%	6	26%	11	25%	6	24%
Fix MARC coding errors	8	24%	7	26%	9	25%	8	21%
Add summaries to results	9	24%	9	23%	7	27%	10	19%
Increase accuracy of library holding information	10	24%	13	22%	8	25%	5	25%
More records for online resources	11	22%	8	26%	14	14%	7	24%
Add more formats	12	22%	10	23%	10	25%	11	19%
More clickable links to online content	14	18%	14	20%	13	15%	9	21%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

In the following charts, significant differences are shown by the shading of the columns. For example, the striped columns highlight the enhancements chosen by significantly more academic and special library respondents than by public library respondents. The dotted columns highlight the enhancements chosen by significantly more academic and public library respondents than by special library

respondents. Finally, the checkered column shows where significantly more public library respondents recommended that enhancement when compared to academic and special library respondents.

Academic library respondents. Academic and public library respondents are significantly more likely than special library respondents to give priority to *upgrading level 3 or other brief records* (note the dotted column).

Academic and special library respondents are significantly more likely than public library respondents to recommend *adding tables of contents to the record* (note the striped columns). Significantly more academic and special library respondents ranked *more records for online/digital resources* in their top ten enhancement lists—an enhancement not in the top ten list for public libraries (note the striped columns). Although not in the top ten list for either group, significantly more academic and public library respondents recommended adding *more records for non-English materials* as an enhancement compared to special library respondents.

Top Ten Data Quality Enhancements

Which of the following enhancements would you recommend?

Base: Academic library respondents
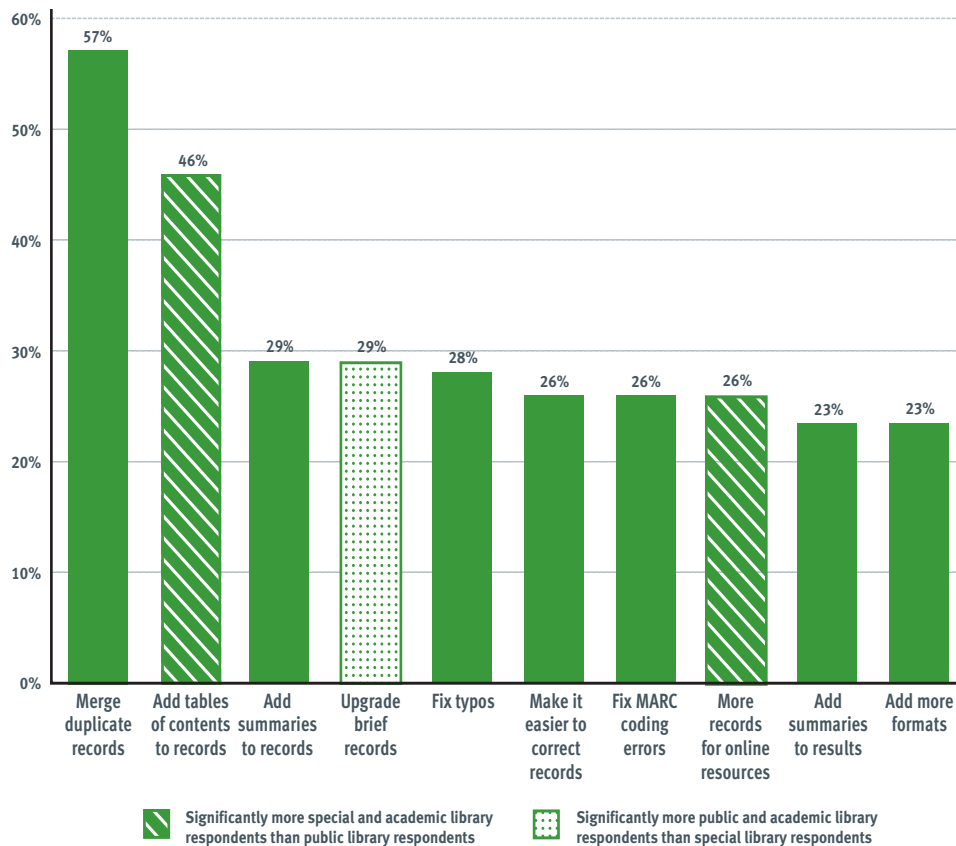
Enhancement	Percentage
Merge duplicate records	57%
Add tables of contents to records	46%
Add summaries to records	29%
Upgrade brief records	29%
Fix typos	28%
Make it easier to correct records	26%
Fix MARC coding errors	26%
More records for online resources	26%
Add summaries to results	23%
Add more formats	23%

Significantly more special and academic library respondents than public library respondents

Significantly more public and academic library respondents than special library respondents

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Public library respondents. Unlike every other respondent group, respondents from public libraries' second most-chosen data quality enhancement was to *add cover art to results* (35%), rather than *adding tables of contents to records*. Significantly more public library respondents selected to *add cover art to results* (35%) (note the checkered column) compared to academic and special library respondents, at 23% and 18%, respectively (not in their top ten list). As already noted, public and academic library respondents are more likely to give priority to *upgrading level 3 or other brief records* compared to special library respondents (note dotted column).

Top Ten Data Quality Enhancements

Which of the following enhancements would you recommend?

Base: Public library respondents

Category	%
Merge duplicate records	48%
Add cover art to results	35%
Fix typos	29%
Upgrade brief records	29%
Add tables of contents to records	27%
Add summaries to records	27%
Add summaries to results	27%
Increase accuracy of library holding information	25%
Fix MARC coding errors	25%
Add more formats	25%

▨ Significantly more public library respondents than academic and special library respondents

▨ Significantly more public and academic library respondents than special library respondents

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Special library respondents. As previously noted, special and academic library respondents are significantly more likely to give priority to *adding tables of contents to the records* and *more records for online/digital resources* compared to public library respondents (note the striped columns).

Top Ten Data Quality Enhancements

Which of the following enhancements would you recommend?

Base: Special library respondents
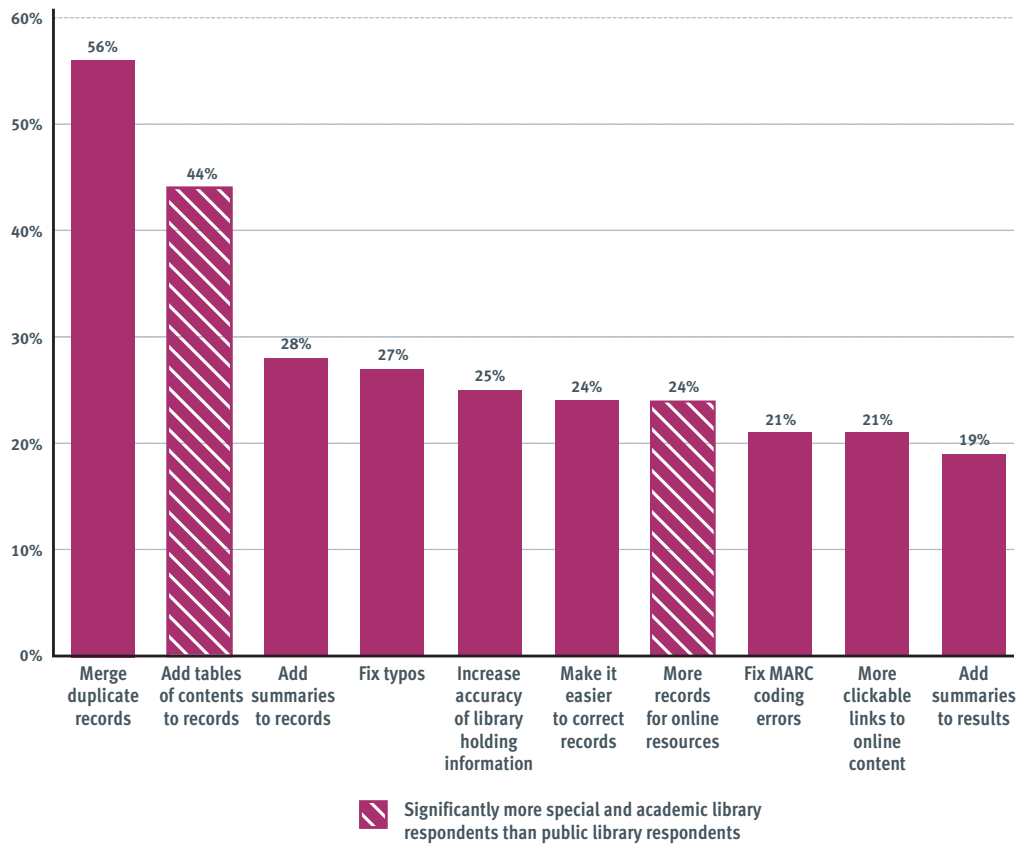
Merge duplicate records — 56%
Add tables of contents to records — 44%
Add summaries to records — 28%
Fix typos — 27%
Increase accuracy of library holding information — 25%
Make it easier to correct records — 24%
More records for online resources — 24%
Fix MARC coding errors — 21%
More clickable links to online content — 21%
Add summaries to results — 19%

Significantly more special and academic library respondents than public library respondents

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

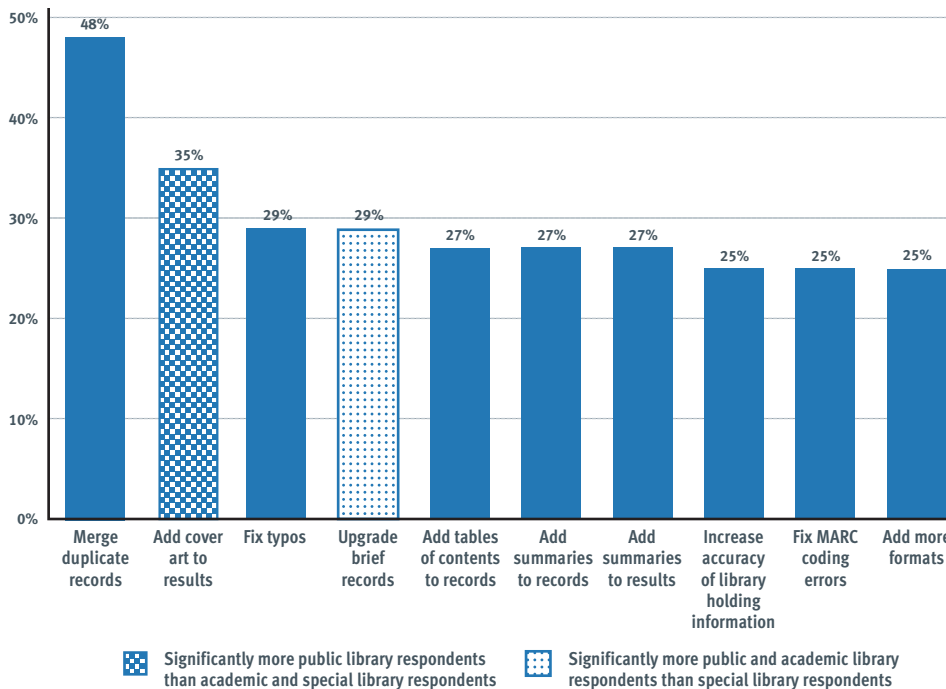**Desired data quality enhancements by job responsibility**

The table on the following page lists the data quality enhancements that were among the top ten enhancements chosen by at least one of the six librarian roles included in this study: cataloging, acquisitions, library directors, reference, collection development and resource sharing.

Merging duplicate records was the top enhancement for each job responsibility, while *adding tables of contents to records* was the second choice among each group.

More records for online/digital resources made it into the top ten list of all groups except for reference and resource sharing staff.

More records for non-English materials made the top ten list of catalogers, but not any other group.

Directors were the only respondent group whose top ten list did not include *making it easier for my library to correct or enrich records.*

Charts showing the top ten enhancements chosen by each role follow this table; significant differences in top ten enhancements among roles are highlighted immediately after these charts.

Desired Data Quality Enhancements

Which of the following enhancements would you recommend?

By job responsibility

Enhancement	Total Rank	%	Cataloging Rank	%	Acquisitions Rank	%	Library Directors Rank	%	Reference Rank	%	Collection Development Rank	%	Resource Sharing Rank	%
Merge duplicate records	1	52%	1	61%	1	55%	1	51%	1	52%	1	55%	1	49%
Add tables of contents to records	2	40%	2	38%	2	40%	2	47%	2	45%	2	46%	2	42%
Add summaries to records	3	28%	7	28%	7	27%	5	30%	3	30%	3	31%	4	27%
Fix typos	4	27%	4	35%	3	31%	8	24%	6	26%	4	29%	9	21%
Upgrade brief records	5	25%	3	37%	4	30%	14	18%	13	19%	12	23%	14	16%
Add cover art to results	6	25%	11	22%	8	26%	4	30%	7	25%	7	26%	8	23%
Make it easier to correct records	7	25%	6	31%	5	29%	12	21%	9	22%	8	25%	10	20%
Fix MARC coding errors	8	24%	5	33%	6	29%	11	21%	10	21%	9	25%	11	19%
Add summaries to results	9	24%	14	19%	12	21%	3	31%	4	29%	6	27%	5	25%
Increase accuracy of library holding information	10	24%	12	20%	9	26%	10	22%	5	29%	5	28%	3	38%
More records for online resources	11	22%	8	25%	10	26%	9	23%	12	19%	10	25%	12	19%
Add more formats	12	22%	10	22%	11	22%	7	26%	8	23%	11	24%	7	23%
Add more records for non-English materials	13	21%	9	23%	13	20%	13	21%	14	15%	14	16%	13	17%
More clickable links to online content	14	18%	15	14%	14	17%	6	28%	11	21%	13	22%	6	23%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

The charts on the following two pages detail the top ten enhancements for each job responsibility: cataloging, acquisitions, library directors, reference, collection development and resource sharing.

Cataloging—Top Ten Enhancements

Which of the following enhancements would you recommend?

Enhancement	%
Merge duplicate records	61%
Add tables of contents to records	38%
Upgrade brief records	37%
Fix typos	35%
Fix MARC coding errors	33%
Make it easier to correct records	31%
Add summaries to records	28%
More records for online resources	25%
More records for non-English materials	23%
Add more formats	22%
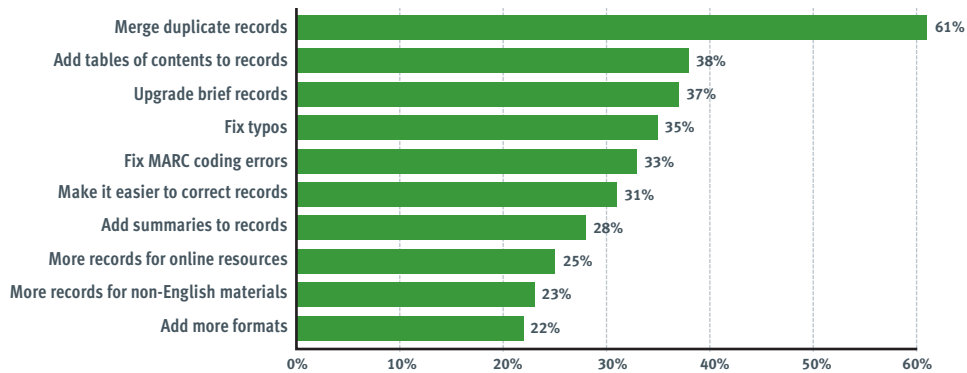
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Acquisitions—Top Ten Enhancements

Which of the following enhancements would you recommend?

Enhancement	%
Merge duplicate records	55%
Add tables of contents to records	40%
Fix typos	31%
Upgrade brief records	30%
Make it easier to correct records	29%
Fix MARC coding errors	29%
Add summaries to records	27%
Add cover art of results	26%
Increase accuracy of library holding information	26%
More records for online resources	26%
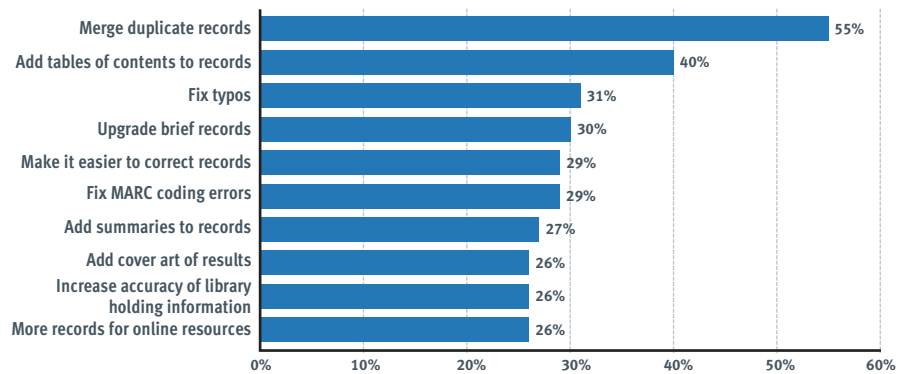
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Library Directors—Top Ten Enhancements

Which of the following enhancements would you recommend?

Enhancement	%
Merge duplicate records	51%
Add tables of contents to records	47%
Add summaries to results	31%
Add cover art to results	30%
Add summaries to records	30%
More clickable links to online content	28%
Add more formats	26%
Fix typos	24%
More records for online resources	23%
Increase accuracy of library holding information	22%
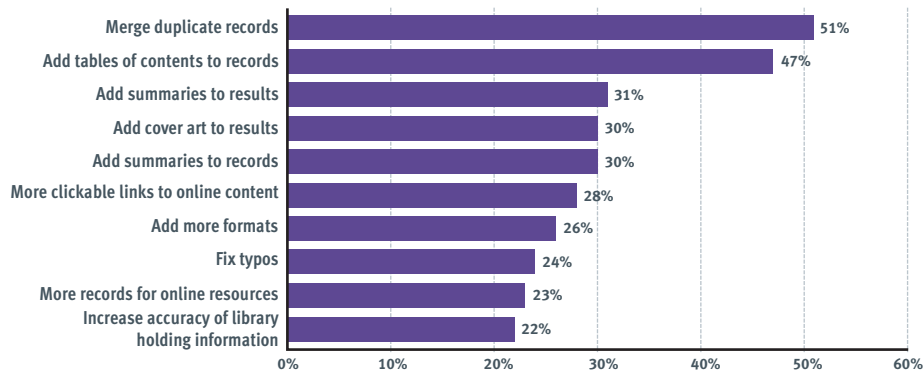
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Reference—Top Ten Enhancements

Which of the following enhancements would you recommend?

Enhancement	Percent
Merge duplicate records	52%
Add tables of contents to records	45%
Add summaries to records	30%
Add summaries to results	29%
Increase accuracy of library holding information	29%
Fix typos	26%
Add cover art to results	25%
Add more formats	23%
Make it easier to correct records	22%
Fix MARC coding errors	21%
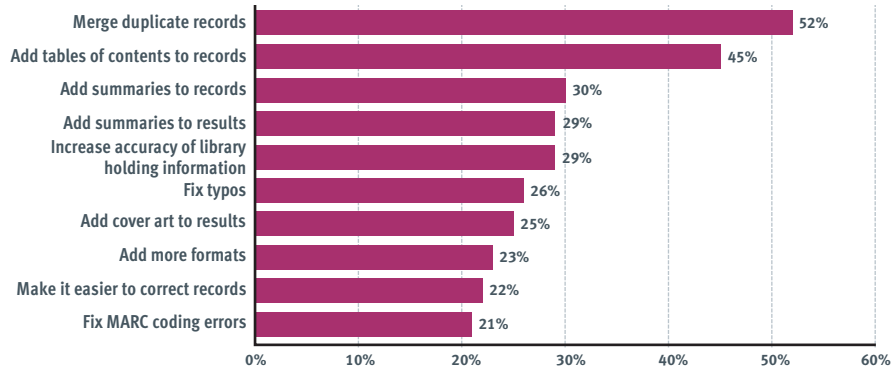
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Collection Development—Top Ten Enhancements

Which of the following enhancements would you recommend?

Enhancement	Percent
Merge duplicate records	55%
Add tables of contents to records	46%
Add summaries to records	31%
Fix typos	29%
Increase accuracy of library holding information	28%
Add summaries to results	27%
Add cover art to results	26%
Make it easier to correct records	25%
Fix MARC coding errors	25%
More records for online resources	25%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Resource Sharing—Top Ten Enhancements

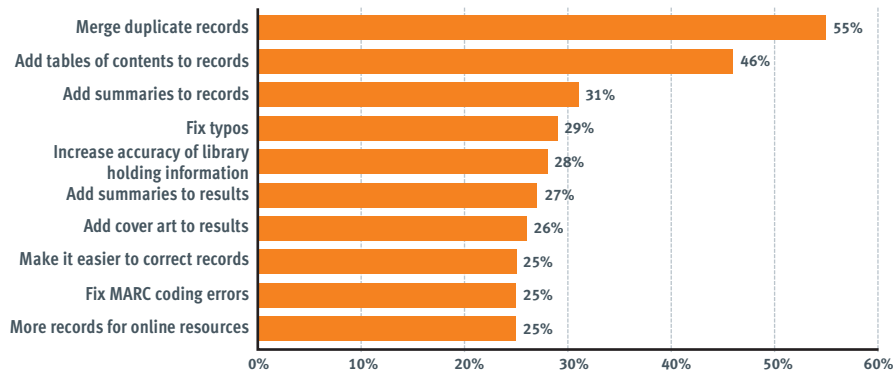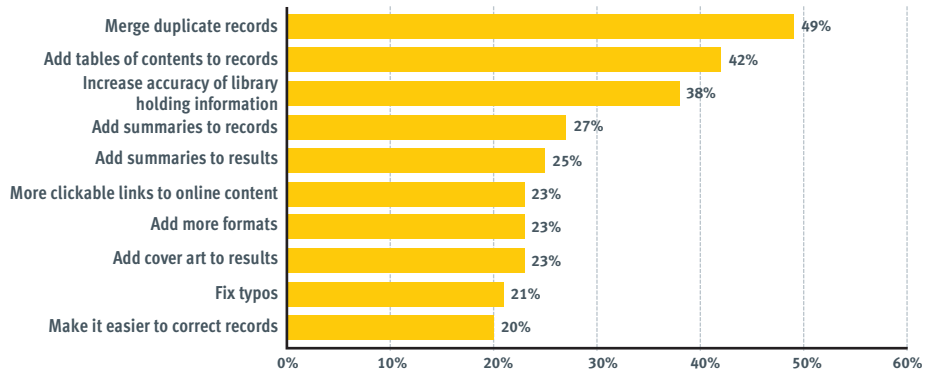*Which of the following enhancements would you recommend?*

Enhancement	Percent
Merge duplicate records	49%
Add tables of contents to records	42%
Increase accuracy of library holding information	38%
Add summaries to records	27%
Add summaries to results	25%
More clickable links to online content	23%
Add more formats	23%
Add cover art to results	23%
Fix typos	21%
Make it easier to correct records	20%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Respondents who reported roles in cataloging

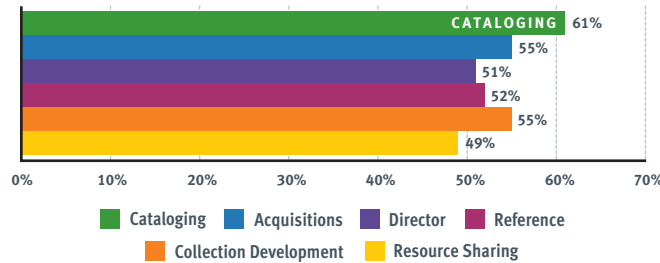Significantly more respondents with job responsibilities in cataloging selected *merging duplicate records* as their top choice compared to directors, reference or resource sharing staff.

Merge Duplicate Records
Which of the following enhancements would you recommend?

CATALOGING 61%
55%
51%
52%
55%
49%

0% 10% 20% 30% 40% 50% 60% 70%

■ Cataloging   ■ Acquisitions   ■ Director   ■ Reference
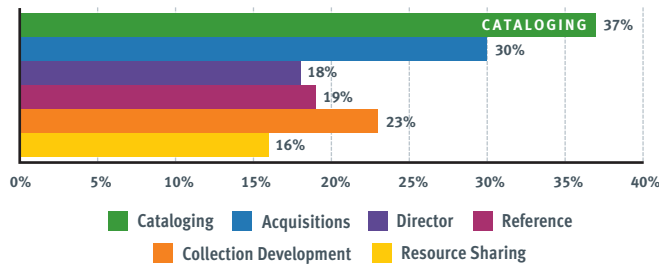■ Collection Development   ■ Resource Sharing

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Similarly, cataloging respondents were significantly more likely to recommend *upgrading level 3 or other brief records* compared to directors, reference, collection development and resource sharing staff.

Upgrade Brief Records
Which of the following enhancements would you recommend?

CATALOGING 37%
30%
18%
19%
23%
16%

0% 5% 10% 15% 20% 25% 30% 35% 40%

■ Cataloging   ■ Acquisitions   ■ Director   ■ Reference
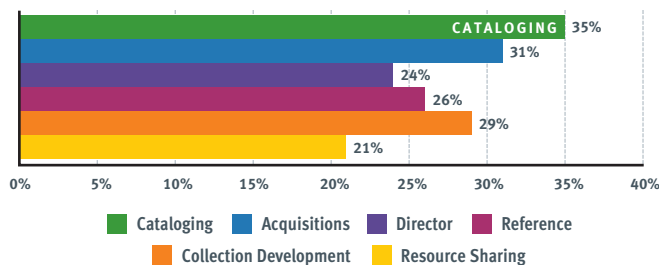■ Collection Development   ■ Resource Sharing

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Cataloging respondents were significantly more likely to recommend *fixing typographical errors* compared to resource sharing, reference and director respondents.

Fix Typos
Which of the following enhancements would you recommend?

CATALOGING 35%
31%
24%
26%
29%
21%

0% 5% 10% 15% 20% 25% 30% 35% 40%

■ Cataloging   ■ Acquisitions   ■ Director   ■ Reference
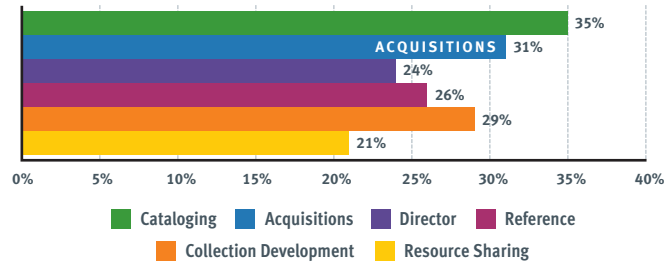■ Collection Development   ■ Resource Sharing

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Respondents who reported roles in acquisitions

Acquisitions respondents were significantly more likely to recommend *fixing typographical errors* compared to resource sharing respondents.

Fix Typos

Which of the following enhancements would you recommend?

Cataloging	35%
ACQUISITIONS	31%
Director	24%
Reference	26%
Collection Development	29%
Resource Sharing	21%

Legend: Cataloging, Acquisitions, Director, Reference, Collection Development, Resource Sharing
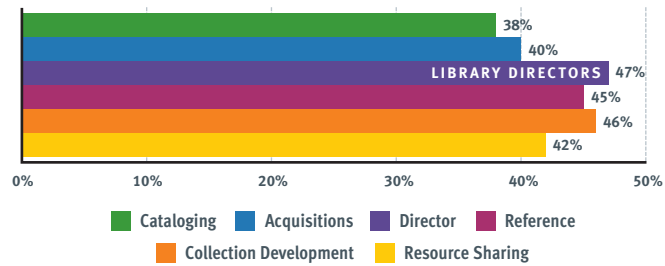
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Respondents who reported roles as library directors

Even though respondents in all job roles chose *adding tables of contents to records* within their top ten priorities, significantly more library director respondents recommended this enhancement compared to cataloging respondents.

Add Tables of Contents to Records

Which of the following enhancements would you recommend?

Cataloging	38%
Acquisitions	40%
LIBRARY DIRECTORS	47%
Reference	45%
Collection Development	46%
Resource Sharing	42%

Legend: Cataloging, Acquisitions, Director, Reference, Collection Development, Resource Sharing
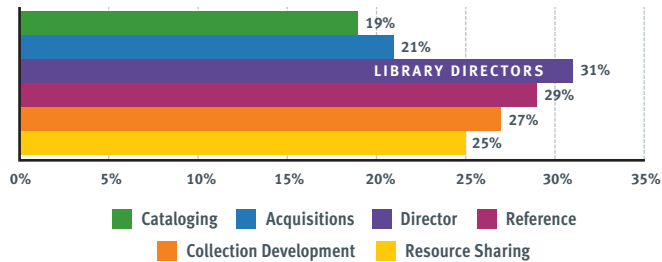
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Director respondents were significantly more likely than acquisitions and cataloging respondents to give priority to *adding summaries/abstracts to the entries on results lists* and *more clickable links to online content.*

Add Summaries to Results
Which of the following enhancements would you recommend?

- 19%
- 21%
- **LIBRARY DIRECTORS** 31%
- 29%
- 27%
- 25%

■ Cataloging ■ Acquisitions ■ Director ■ Reference
■ Collection Development ■ Resource Sharing

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

More Clickable Links to Online Content
Which of the following enhancements would you recommend?

- 14%
- 17%
- **LIBRARY DIRECTORS** 28%
- 21%
- 22%
- 23%

■ Cataloging ■ Acquisitions ■ Director ■ Reference
■ Collection Development ■ Resource Sharing

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Respondents who reported roles in reference

Reference librarians were significantly more likely than cataloging respondents to give priority to *adding summaries to results lists* and *increasing the accuracy and currency of library holdings information.*

Add Summaries to Results
Which of the following enhancements would you recommend?

- 19%
- 21%
- 31%
- **REFERENCE** 29%
- 27%
- 25%

■ Cataloging ■ Acquisitions ■ Director ■ Reference
■ Collection Development ■ Resource Sharing

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Increase Accuracy of Library Holding Information

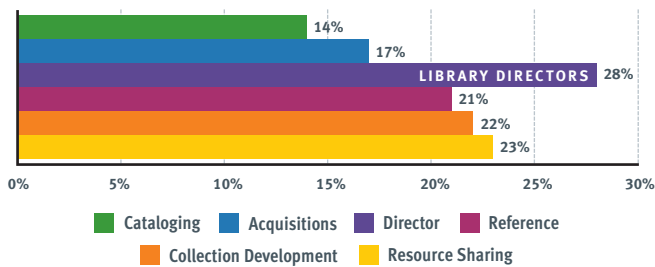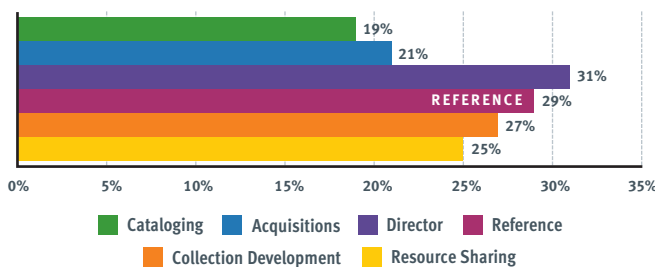*Which of the following enhancements would you recommend?*

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Respondents who reported roles in resource sharing

Resource sharing respondents were significantly more likely than any other group to recommend *increasing the accuracy and currency of library holdings information.*
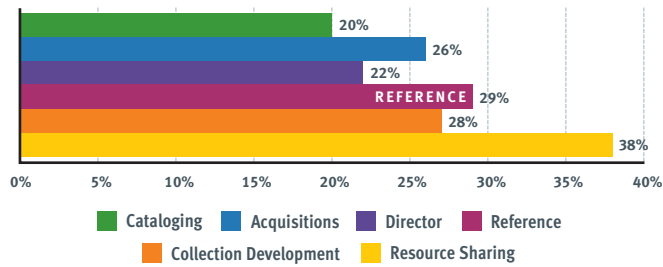
Increase Accuracy of Library Holding Infomation

Which of the following enhancements would you recommend?
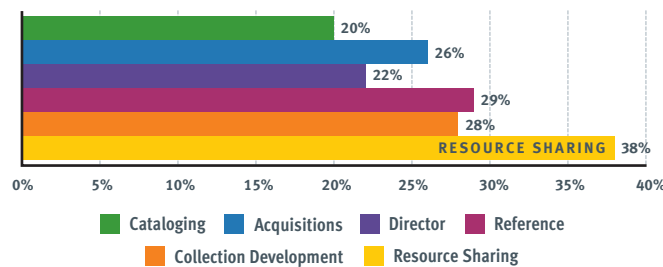
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

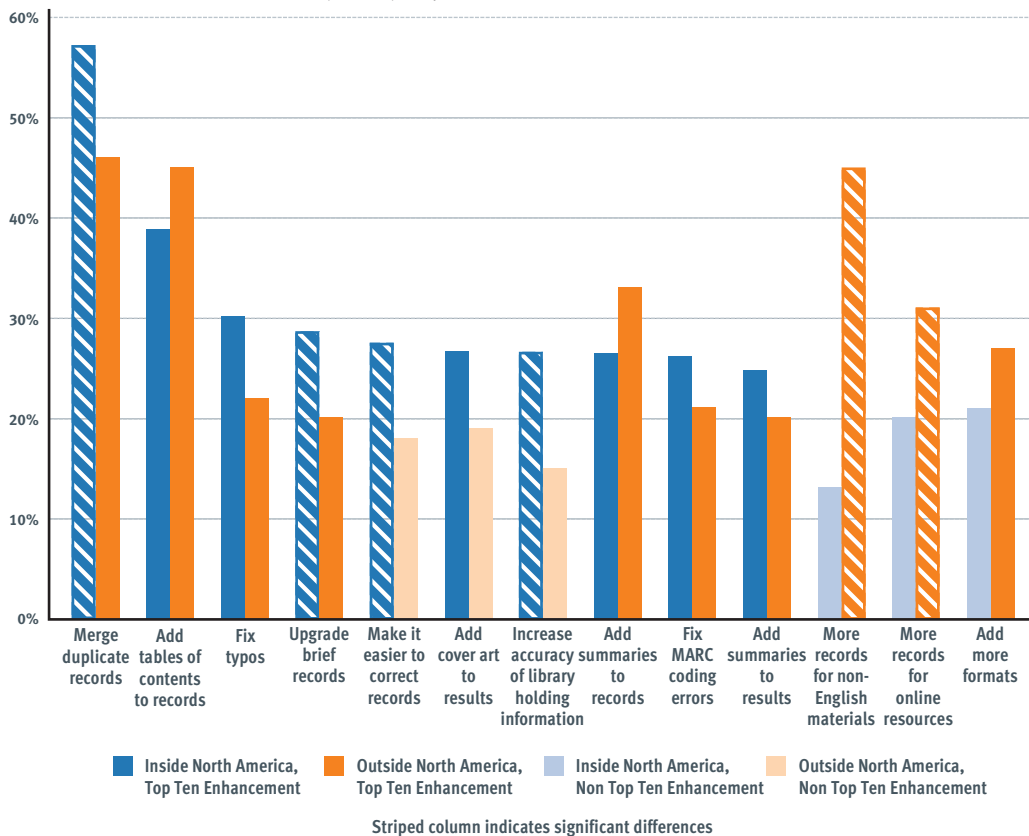**Desired enhancements among respondents by region**

The analysis revealed many differences among survey respondents from libraries inside and outside North America. Respondents from outside North America were significantly more likely to recommend *more records for non-English materials* and *more records for online materials*—neither of which made it into the top ten priorities among North American respondents. Another enhancement included in the top ten of respondents outside North America but not in the top ten for North American respondents was *adding more formats (e.g., articles, images, maps, audio, video, etc.)*.

Even though respondents from inside and outside North America chose *merging duplicate records* as their top priority, significantly more North American respondents recommended this enhancement when compared to respondents outside North America. North American respondents were also significantly more likely to recommend *upgrading level 3 or other brief records, making it easier for my library to correct or enrich records,* and *increasing the accuracy and currency of library holdings information. Making it easier for my library to correct or enrich records, adding cover art to entries on results lists,* and *increasing the accuracy and currency of library holdings information* made it into the top ten priorities of North American respondents but not of respondents outside of North America.

Top Ten Data Quality Enhancements
Which of the following enhancements would you recommend?

Base: Library survey respondents Inside and Outside North America

Inside North America, Top Ten Enhancement — Outside North America, Top Ten Enhancement — Inside North America, Non Top Ten Enhancement — Outside North America, Non Top Ten Enhancement

Striped column indicates significant differences

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Data Quality: Librarians and End Users

Key Findings

- **Both end users and librarians place a high priority on delivery-related data elements for printed or other analog materials—that is, knowing where the items are held and which are available immediately.**

- **End users, but not librarians, give the highest priority to enhancing the catalog with *more links to online content*. Librarians give the highest priority to *merging duplicate records*.**

Librarians and library staff ranked *more links to online content* in the bottom third of their enhancement list. However, when asked to describe the preferences of end users, many librarians who work directly with end users can accurately predict the importance that end users place on linking to online content from the catalog.

While their own data quality priorities tend to be markedly different from those of end users, many librarians who work directly with end users also appear to be highly aware of the importance that end users place on enriched content, more records for formats other than print (e.g., audio and video), and the availability of sample text/excerpts from the catalog.

- **Catalog data elements selected by librarians and library staff as *essential* tend to relate to known item searching. The presence of an ISBN, for example, is a high priority for librarians' identification of needed items.**

- **End users' choices of data elements that are *essential* for identifying wanted items suggest they value and expect evaluative information (such as summaries) to assist their discovery and choice of wanted items.**

- **The enhancement of catalogs with tables of contents and summaries is equally important to end users and librarians.**

These data elements are not generally included in a standard catalog record.

- **Except for tables of contents and summaries, the catalog data quality requirements of end users and librarians tend to be different.**

Differences appear to be driven primarily by the individual's purpose in using the catalog—to find and obtain needed information (end users) or to carry out a work assignment (librarians—for example, to order a new book for the library).

- **Librarians' choice of data quality enhancements reflects their understanding of the importance of accurate, structured data in the catalog.**

End users benefit from this structured data, for example when refining their searches with facets; however, end users tend to be unaware of the data quality requirements that support the functionality they rely on for precise, consistent results. This phenomenon is another driver of the differences in data quality priorities between librarians and end users.

Detailed Observations

Overview

The first section of this chapter compares the responses of the end-user pop-up survey to the responses of the library survey on two dimensions:

- Data considered most essential for identifying wanted items
- Recommended data quality enhancements.

The end-user findings reported here are based on end-user reactions to WorldCat.org, OCLC's freely available end-user interface to WorldCat. For comparing responses about *essential data for identifying items,* the research team extracted that subset of library survey respondents who reported using WorldCat.org.

The findings about desired *data quality enhancements* represent end-user reactions to WorldCat.org and librarian and staff reactions to WorldCat data from any OCLC interface.

To the extent that WorldCat and a library's online catalog are comparable and share data quality issues, the findings of this study are generalizable to individual library online catalogs.

As previously noted, the study seeks to assess data quality (rather than interface) needs, based on the searcher's purpose in using the data; for end users, the purpose is generally finding and obtaining items to meet an information need, while for librarians and staff the purpose is generally carrying out a work assignment.

A second section of this chapter analyzes library survey respondents' perceptions of what enhancements would be most helpful for *end users*. The analysis compares librarian and staff respondents' perceptions of what end users want to what end-user survey respondents actually reported.

Comparing What Librarians and End Users Want
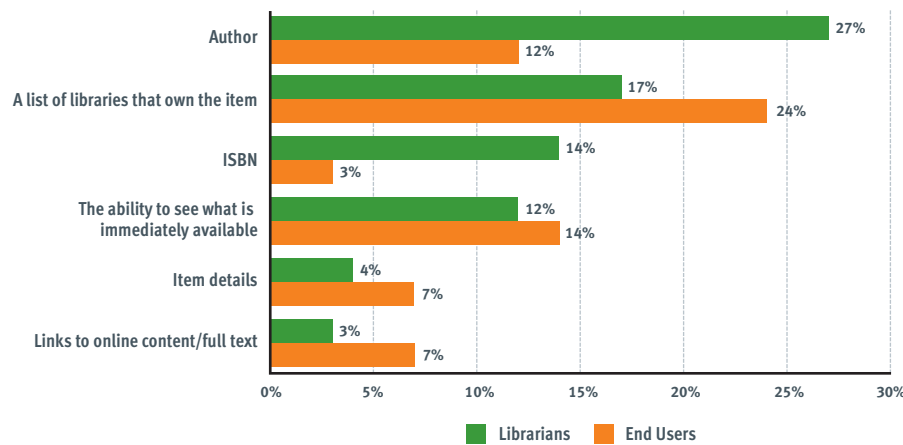
Most essential data elements

Both library and end-user survey respondents were asked what information in the records was most essential in identifying the items they need. The following chart

details the responses among library survey respondents and also end users who use WorldCat.org.

Comparing the librarian/staff to end-user results, both library and end-user respondents find *author, a list of libraries that own an item* and *item availability* to be essential. Notably, ISBN—the third-most selected data element by librarian and staff respondents—was selected by only 3% of end users as essential.*

Most Essential Information

*What information is the **most** essential in helping identify the item needed? (Library survey)*
*What information is **most** essential in helping you identify the item that you need?*
(End-user pop-up survey)

	Librarians	End Users
Author	27%	12%
A list of libraries that own the item	17%	24%
ISBN	14%	3%
The ability to see what is immediately available	12%	14%
Item details	4%	7%
Links to online content/full text	3%	7%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey and end-user pop-up survey)

One part of the library survey prompted participants to write how they would choose to enhance the database if they could wave a magic wand over it. A written comment perhaps explains why library survey respondents find the presence of an ISBN so important: this cataloger/resource sharing staff member from a public library said, "It would make my job easier if all current editions of a bibliographic record had either an ISBN number or a publisher number." Other respondents who provided comments related to ISBN wished for all records to have accurate ISBNs and for improvement in the accuracy of ISBNs for older materials.

*Note: Title—the ubiquitous choice—was excluded in order to focus more attention on other data elements.
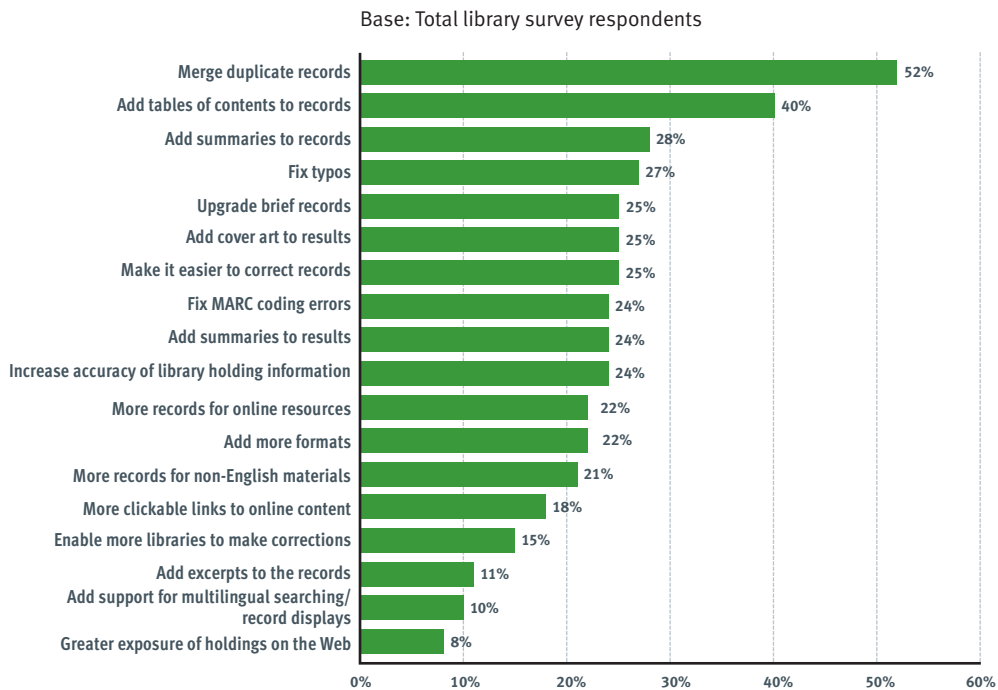
Desired data quality enhancements

The library survey and the end-user pop-up survey results suggest there are important differences in the data quality requirements of those who work with the catalog (librarians and staff) and those who use it to find and get information (end users). The following chart lists from most to least chosen all of the desired data quality enhancements from library and staff respondents (they could choose multiple items from the list of enhancements in the library survey).

Merging duplicate records (52%), *adding tables of contents to detailed bibliographic records* (40%) and *adding summaries/abstracts to detailed records* (28%) are the top three recommended data quality enhancements among respondents to the library survey.

Desired Data Quality Enhancements

Which of the following enhancements would you recommend?

Base: Total library survey respondents

Enhancement	Percentage
Merge duplicate records	52%
Add tables of contents to records	40%
Add summaries to records	28%
Fix typos	27%
Upgrade brief records	25%
Add cover art to results	25%
Make it easier to correct records	25%
Fix MARC coding errors	24%
Add summaries to results	24%
Increase accuracy of library holding information	24%
More records for online resources	22%
Add more formats	22%
More records for non-English materials	21%
More clickable links to online content	18%
Enable more libraries to make corrections	15%
Add excerpts to the records	11%
Add support for multilingual searching/record displays	10%
Greater exposure of holdings on the Web	8%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)
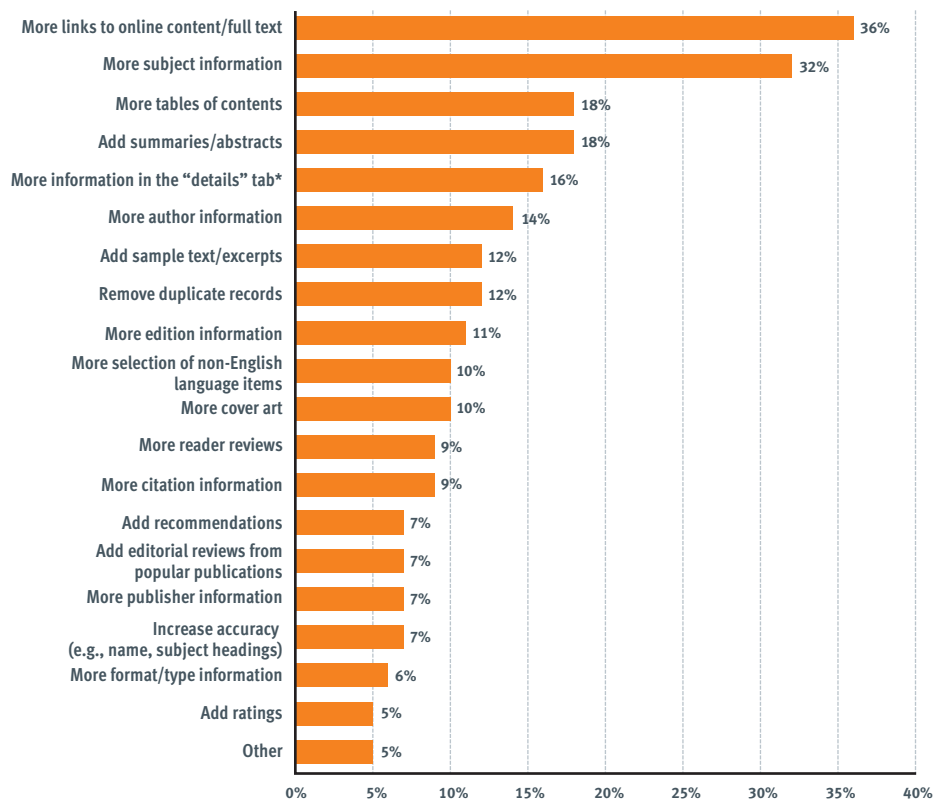
The next chart lists from most to least chosen all of the desired data quality enhancements from respondents to the end-user pop-up survey. The multiple-choice list of enhancements presented in the survey was (necessarily) not the same as the list presented in the library survey, but there were some comparable choices.

End users' top two enhancements were to add *more links to online content/full text* and *more subject information*. Tied for third are *more tables of contents* and *add summaries/abstracts*. From more *subject information* through *add sample text/excerpts*, end users appear to be asking for data elements that are not generally included in a standard catalog description. The end-user responses ranked the enhancements related to correcting data *(remove duplicate records* and *increase accuracy (e.g., name, subject headings)* as 8th and 17th, respectively.

Desired Data Quality Enhancements

*What changes would be **most** helpful to you in identifying the item that you need?*

Base: End-user pop-up survey respondents

Enhancement	Percentage
More links to online content/full text	36%
More subject information	32%
More tables of contents	18%
Add summaries/abstracts	18%
More information in the "details" tab*	16%
More author information	14%
Add sample text/excerpts	12%
Remove duplicate records	12%
More edition information	11%
More selection of non-English language items	10%
More cover art	10%
More reader reviews	9%
More citation information	9%
Add recommendations	7%
Add editorial reviews from popular publications	7%
More publisher information	7%
Increase accuracy (e.g., name, subject headings)	7%
More format/type information	6%
Add ratings	5%
Other	5%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey)

*At the time of the pop-up survey, the WorldCat.org "details" tab contained basic bibliographic information plus enriched data such as table of contents and summary/abstract, if available.

The multiple-choice lists of enhancements differed, but sufficient overlap exists to make some comparisons useful. Library survey and end-user survey respondents seem to agree on the importance of *adding tables of contents* and *adding summaries/abstracts;* beyond that, various differences emerge. Tabulating the overlapping choices on the two survey lists, and sorting each ranked choice into the top, middle or bottom third of the two lists yields the results detailed in the following chart. In these results, the difference between library and end-user survey respondents' choices pertaining to *adding more links to online content* seems the most significant; while this enhancement was the first choice of end users, library and staff respondents ranked it in the bottom third of their choices. End-user interest in *adding sample text/excerpts* also seems considerably keener than library and staff respondents' interest in this enhancement.

Relative Ranking of Data Quality Enhancements in Library and End-user Surveys

Which of the following enhancements would you recommend? (Library survey)
*What changes would be **most** helpful to you in identifying the item that you need?*
(End-user pop-up survey)

	Relative Ranking		
	● *Library Survey Respondents* ● *End-user Survey Respondents*		
Comparable Enhancements Choice*	***Top Third***	***Middle Third***	***Bottom Third***
Merge duplicate records	● #1 Choice	●	
More links to online content	● #1 Choice		●
Add tables of contents to records	● ●		
Add summaries to records	● ●		
Add cover art to results	●	●	
Add more formats		●	●
More records for non-English materials		●	●
Add excerpts to the records	●		●

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey and end-user pop-up survey)

*Note: This is a subset of responses from both the library survey and the end-user pop-up survey.

Advanced searching, refining searches and faceted search

As noted in an earlier chapter, supporting advanced searching, browsing and faceted navigation of search results involves fielded indexing of bibliographic data elements and usually the establishment and maintenance of controlled forms of headings (for names and subjects) to ensure the consistent and reliable collocation of search results in the catalog.

Library survey respondents' enhancement choices placed a good deal more emphasis than end users' choices did on correcting errors (like merging duplicates and fixing typographical and MARC coding errors) that compromise the ability of catalog data elements to effectively support advanced and faceted searching. The end users in this study's focus groups found advanced and faceted searching helpful in certain circumstances—especially when working with large retrieval sets—but end users tend to be unaware of the data structures and practices required to support this functionality. Librarians and staff are aware of what it takes to support more sophisticated search features, and so it makes sense that library survey respondents—especially those who reported cataloging responsibilities—gave higher priority than end users did to database correction activities.

Librarians' Perceptions of What End Users Want

Library survey participants who work directly with users were asked to predict what enhancements their end users would recommend (their perceptions are labeled as "Librarians' Perception of End Users' View" when presented in charts). This section provides the results of the perceived end-user views and where applicable, compares the perceived user views to the actual views of end-user respondents.

From a list of 18 enhancements (see chart on next page), library survey respondents who work directly with users were asked to select one, then other enhancements to WorldCat that would be the most helpful for their libraries' users. Nearly half of the library and staff respondents (48%) felt that *adding cover art to search result lists* would be helpful to their users, followed by 44% who felt *adding tables of contents to records* and 43% who felt *adding summaries/abstracts to search result lists* would be helpful.

An analysis of the library survey results segmented by type of library revealed that respondents from public libraries were significantly more likely than academic and special library librarians to feel that *adding cover art to search result lists* would be helpful to their users (Public = 69%, compared to Academic = 45% and Special = 31%).
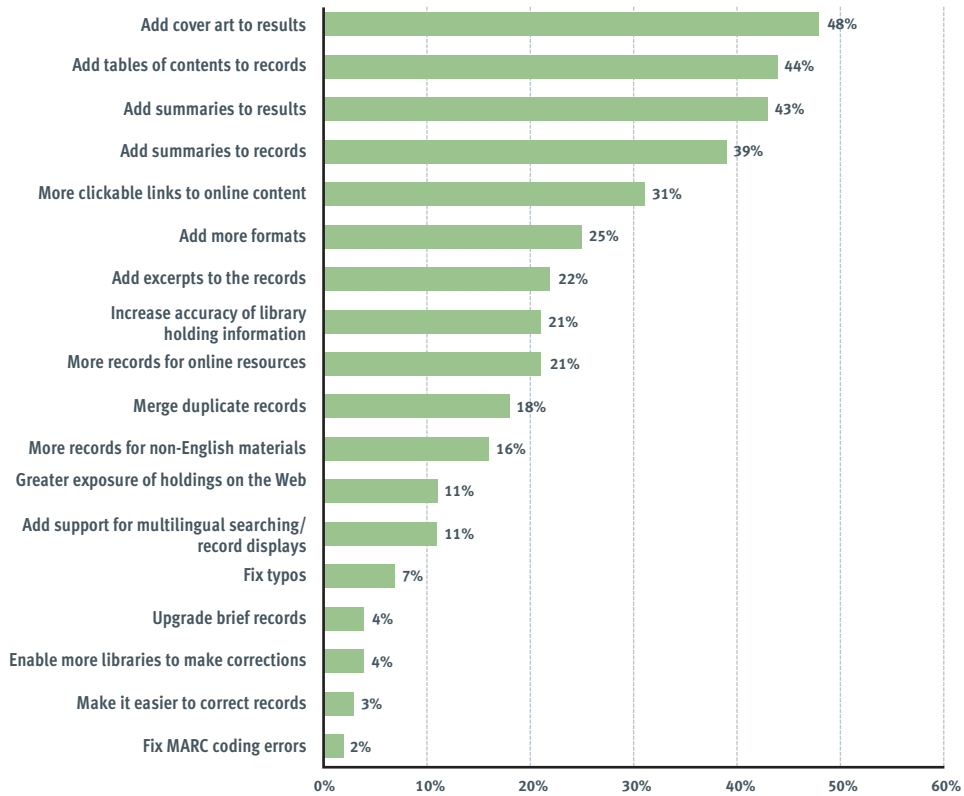
A quarter or more of total library survey respondents who work directly with users selected *adding summaries/abstracts to detailed bibliographic records* (39%), *adding more clickable links to online content* (31%) and *adding more formats* (25%) as enhancements that would be helpful to their users.

Librarians' Perception of End Users' View: All Desired Data Quality Enhancements

*Which of the following enhancements would be most helpful for your **patrons?***

Base: Librarians who work directly with users

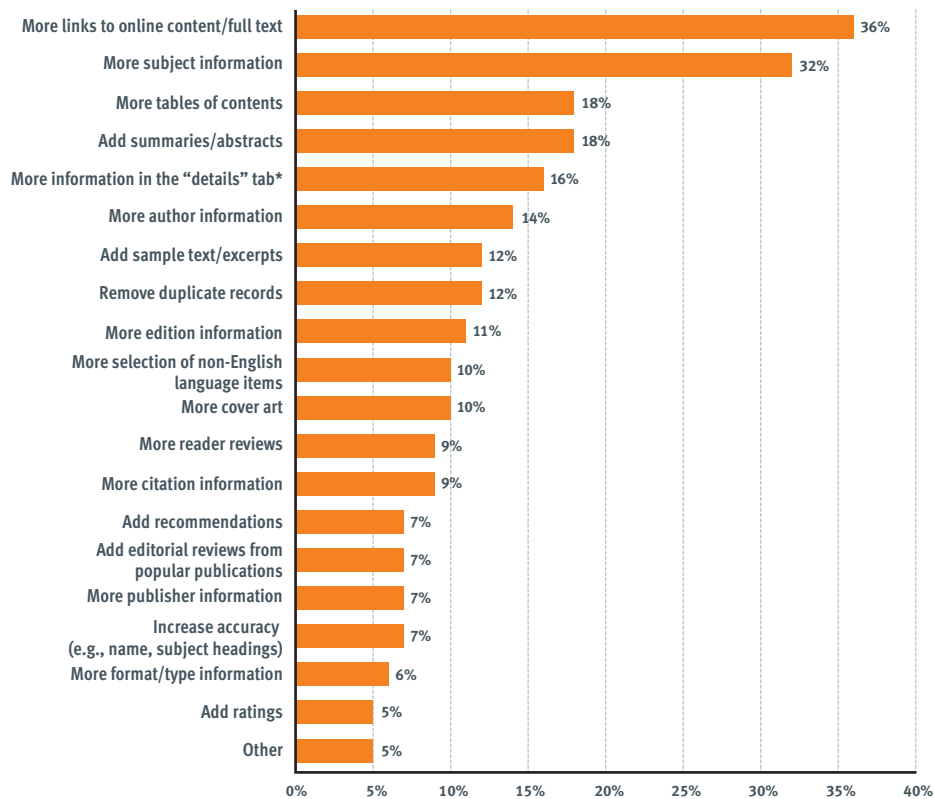| Enhancement	Percentage
Add cover art to results	48%
Add tables of contents to records	44%
Add summaries to results	43%
Add summaries to records	39%
More clickable links to online content	31%
Add more formats	25%
Add excerpts to the records	22%
Increase accuracy of library holding information	21%
More records for online resources	21%
Merge duplicate records	18%
More records for non-English materials	16%
Greater exposure of holdings on the Web	11%
Add support for multilingual searching/record displays	11%
Fix typos	7%
Upgrade brief records	4%
Enable more libraries to make corrections	4%
Make it easier to correct records	3%
Fix MARC coding errors	2%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (Library survey)

Thus, librarians and library staff are aware of the importance that end users place on enriched content such as summaries/abstracts, of their desire for more clickable links to online content/full text, and of the priority end users place on sample text/excerpts. These priorities are mirrored in the choices end users made from the enhancement list that was presented to them in the end user survey, shown below.

Desired Data Quality Enhancements

*What changes would be **most** helpful to you in identifying the item that you need?*

Base: End-user pop-up survey respondents

Enhancement	Percentage
More links to online content/full text	36%
More subject information	32%
More tables of contents	18%
Add summaries/abstracts	18%
More information in the "details" tab*	16%
More author information	14%
Add sample text/excerpts	12%
Remove duplicate records	12%
More edition information	11%
More selection of non-English language items	10%
More cover art	10%
More reader reviews	9%
More citation information	9%
Add recommendations	7%
Add editorial reviews from popular publications	7%
More publisher information	7%
Increase accuracy (e.g., name, subject headings)	7%
More format/type information	6%
Add ratings	5%
Other	5%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey)

*At the time of the pop-up survey, the WorldCat.org "details" tab contained basic bibliographic information plus enriched data such as table of contents and summary/abstract, if available.
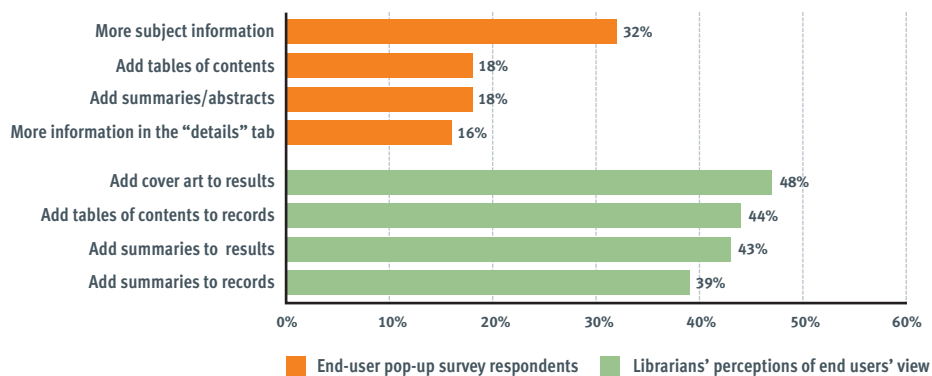
Adding cover art to results, *adding tables of contents to records*, *adding summaries to results* and *adding summaries to the bibliographic records* top the list of discovery-related enhancements that librarians and staff feel would be helpful to their users. To aid in discovery, end users reported that they want *more subject information*, followed by the addition of evaluative information similar to what librarians predicted—*adding tables of contents* and *summaries/abstracts*.

Top Discovery-related Data Quality Enhancements

*What changes would be **most** helpful to you in identifying the item that you need? (End-user pop-up survey)*
*Which of the following enhancements would be most helpful for your **patrons**? (Library survey)*

More subject information	32%
Add tables of contents	18%
Add summaries/abstracts	18%
More information in the "details" tab	16%
Add cover art to results	48%
Add tables of contents to records	44%
Add summaries to results	43%
Add summaries to records	39%

■ End-user pop-up survey respondents   ■ Librarians' perceptions of end users' view

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey and library survey)

On the delivery side, *more links to online content* topped both librarian (perception of end users' views) and end-user survey respondents' enhancement choices. Librarians and staff also predicted that end users would be helped by *increasing the accuracy of library holdings information*—an enhancement that undoubtedly would be helpful in improving the end-user's delivery experience, but of which end users would likely be unaware.

Top Delivery-related Data Quality Enhancements

*What changes would be **most** helpful to you in identifying the item that you need? (End-user pop-up survey)*
*Which of the following enhancements would be most helpful for your **patrons**? (Library survey)*

More links to online content/full text	36%
More clickable links to online content	31%
Increase accuracy of library holding information	21%

■ End-user pop-up survey respondents   ■ Librarians' perceptions of end users' view
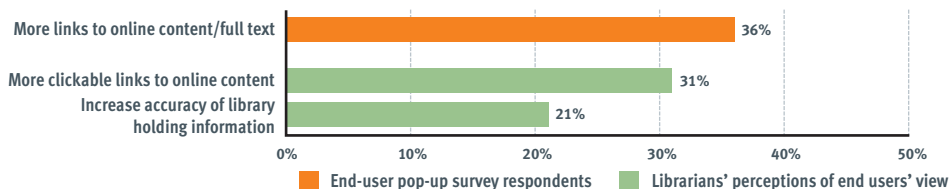
Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey and library survey)

Conclusions

Librarians and End Users

The study found important differences between the catalog data quality priorities of end users and those who work in libraries. End users, librarians and library staff approach library catalogs purposefully. When end users approach catalogs, they generally want to satisfy information needs; librarians and library staff generally have work assignments to carry out. The purpose for which catalog data is being used seems to be an important driver of differences in data quality priorities.

The study also found significant differences between data quality priorities of librarians/staff by work role, type of library and region. Work role and region seem to be the primary drivers of differences among types of library survey respondents in this study.

The understanding that those who work with catalog data have of how its data structures work—and the lack of that understanding among end users—is another driver of different data quality priorities among end users and some librarians. End users are largely unaware of the catalog's infrastructure, although the end-user focus group participants in this study responded favorably to some of the features that rely on it (e.g., facets, advanced search).

The fact that end-user participants found the ability to conduct an advanced search helpful suggests that the investment in separate indexes and controlled forms of names and subjects can benefit the end user's discovery experience in next-generation library catalogs.

To the extent that librarians and staff approach their own catalogs and WorldCat in comparable ways, this study supports the view that those with different roles in their libraries have somewhat different data quality priorities. In particular, those with roles in cataloging and acquisitions certainly place higher priority on database error correction activities than end users, but also compared to librarians and staff in other job roles. Cataloging and acquisitions staff place a high value on—and recognize the importance of—the catalog's formally structured data, for example, fielded indexes and authorized forms of headings that support advanced searching, search limits, faceted browsing and other features of the catalog that rely on the catalog's infrastructure. They also place the highest priority on merging duplicates (more than one record for the same edition) and fixing inaccuracies in structured data.

Balancing What End Users and Librarians Want

Catalogs have many constituencies, both inside and outside the library. This study's results suggest that end users place a high priority on enrichment data (tables of contents, summaries, etc.) and on links to online content, both text and media. Librarians and library staff are also important catalog users, and their data quality priorities tend to be different than end users' priorities. If their data quality needs are met, they can do their jobs more efficiently and effectively. And, some of librarians' data quality requirements (to be able to fix mistakes and control heading forms, for example), while not shared by end users, play a role in fulfilling end users' needs.

In a world of unlimited human and financial resources, a data quality program for a library's online catalog could meet all end users' needs and all librarian and staff needs. In a world of limited resources, library leaders must make choices, creatively deploy the resources they have, and balance competing data quality requirements. A data quality program that strikes a balance between what end users and librarians/staff want and need, but gives an edge to the desires of end users, seems more likely to assure the library will continue to thrive in the end-user communities it serves.

Usability, User-centered Design and the Principles of Information Organization

An interesting insight arises from reading the chapter in Fran Miksa's dissertation that covers the development and thinking behind Charles Cutter's nineteenth century *Rules for a Dictionary Catalog*;[1] then immediately reading the first chapter of the first full draft of *Resource Description and Access* (RDA),[2] which carries forward the concepts of *Functional Requirements for Bibliographic Description (FRBR)*.[3] While Cutter's rules and RDA were or will be applied at very different points in history, reading the two texts suggests the principles of information organization underlying them are consistent. One can discern an unbroken thread of development from Cutter to RDA.

Yet with a few exceptions of rigorous user studies that may take shape around FRBR concepts,[4] there is no evidence that Cutter or the framers of FRBR or RDA systematically tested their assumptions with end users of information systems. Fran Miksa, who has over the course of his long career publicly noted librarianship's relatively separate and parallel developments of the principles of information organization on the one hand, and use and user studies on the other, has said, "I conclude that the idea of information users and use remains rather mysterious in its overall sense—rather like the images we see while driving in a fog."[5]

As Web information services have taken off, commanding a great deal of attention from all segments of the information-seeking public, those who built them appeared to pay little heed (if they were aware at all), of the conceptual frameworks for information organization embraced by librarianship. Instead a good deal of experimentation, trial and error seemed to be the rule.

As it became clearer to Web developers what worked and what didn't, many also learned to take advantage of new opportunities in the Web's virtual world—lessons that have emerged as some very different ways of organizing large volumes of information, for example, on Flickr or Facebook. On the Web, the principles of usability and user-centered design might be said to have displaced the traditional principles of information organization, at least as librarians have practiced them. David Weinberger, writing about new ways of bringing order to masses of digital information, notes that "everything is connected and therefore everything is metadata."[6] Based on their experiences with popular Web sites, Internet-savvy end users expect to be able to search on a rich store of metadata from many sources and easily find and get just what they want.

Given these two different traditions for bringing order to information on end users' behalf—one from librarianship and the other from the Web—this study's results are not surprising. Librarians' perspectives about data quality remain highly influenced by classical principles of information organization, while end users' expectations of data quality arise from their experiences of how information is organized on popular Web sites. What is needed now is to integrate the best of both worlds in new, expanded definitions of what "quality" means in library online catalogs, as well as who is responsible for providing it.

Metadata and Content

Delivery, Links and More Online Content in the Catalog

In this study, end user respondents, but not library survey respondents, gave the highest priority to enhancing the catalog with more links to online content. End users appear to perceive the process of moving through discovery and selection to access as one continuous flow, while librarians tend to think about user tasks as separate and distinct. In a blog post on FRBR, Karen Coyle goes so far as to say, "The FRBR user tasks [find, identify, select, obtain] are limited in scope, and as such they limit how we think about users and catalogs."[7]

Because end users come from an information world where a huge amount of content is online, it is natural for them to expect to be able to access content—not just discover, select and be directed how to get it (the *modus operandi* of the library catalog). As Google Book Search gains the attention of those who use library collections, an end user's appetite for linking immediately to the digitized content of books, or at least to snippets, can be expected to increase even more. The end-user expectation to link to content extends beyond text and includes the expressed desire to link to samples of music and video.

Discovery, Delivery and Enrichment Data

These results suggest that end users want to reduce the difference between the description of an item (its catalog record) and the item itself by enriching the basic catalog record with tables of contents, summaries/abstracts, cover art and excerpts

or sample content. For some end users (in this study, undergraduates), social content like user reviews, ratings and tags was also seen as helpful. Librarians seem to agree that at least tables of contents should be added to catalog records.

In keeping with the finding that delivery is as important, if not more important than discovery, end users generally don't see the point of finding things they can't get or spending time to get things that won't meet their information need. Not wanting to waste their time or energy, end users seem to want this enriched content to help them decide if it's worth their time to try to obtain the item.

"Physical delivery has long been the ignored stepchild of the library world."[8] These words begin Valerie Horton's 2009 report on the second 'Moving Mountains' conference on the physical delivery of library materials. Her article describing the state of library delivery (interlibrary loan, consortial borrowing, home delivery) through a variety of conference presentations is a good starting point for considering how to raise the priority of delivery in libraries.

Subject Headings and Subject Information

When end-user survey respondents selected "more subject information" as an enhancement priority, what did they mean? It is unlikely, given the relatively few unique subject-rich words contributed to a catalog description by controlled subject headings,[9] that they mean more controlled subject headings. Given end-user survey respondents' top choices for catalog enhancement and what end-user focus group participants reported, "more subject information" is more likely to be interpreted as subject-rich data elements not generally included in a standard catalog description.

At the same time, controlled subject terms and phrases serve end users in a number of ways: as subject-rich index terms; to support multilingual subject searching (when records contain subject headings in more than one language); as facets for refining or expanding searches; for browsing; as words or phrases linked to classification or other terminologies; as a factor in determining relevance ranking; and more. To support these features, today's catalogs rely on labor-intensive practices for producing controlled subject headings. Given the growing concern that these traditional methods are not sustainable going forward, it may be necessary for libraries to find more economical means to achieve the benefits to end users that controlled subject vocabularies provide.

Standard Numbers

The findings suggest that standard numbers like the ISBN are critical to support librarians' work. Despite what end users selected as essential elements for identifying wanted items, standard numbers can also be essential to support end-user tasks. The presence of standard numbers is essential, for example, for providing reliable links to (and movement of data between) the same item in multiple repositories, as well as supporting a variety of machine-to-machine interactions that improve the end user's discovery or delivery experience.

Where to Get the Data

Enrichment Data

This study is far from the first to find that enrichment data such as summaries and contents notes are important contributors to end-user searching;[10] that users want enriched records;[11] and that enriched records increase usage of library materials.[12] However, adding enrichment data like tables of contents and summaries by hand to individual records in a local catalog is an approach that will not scale to the task ahead. Libraries need to work together to find ways to share the costs of enriching discovery systems and keyword indexes with this content.

Existing catalog data is the first source to mine. FRBRizing catalog data is a promising approach for obtaining more evaluative content such as summaries, subject tag clouds, cover art and possibly tables of contents. The OCLC researchers who produced FictionFinder[13] were very successful in mining FRBR clusters for many of these data elements, as illustrated below. By exploiting information in the FRBR clusters, FictionFinder is able to provide a summary on the search page for many works. For example, the illustration of the FictionFinder search for "foundlings" provides summaries for famous works on the topic, such as *Wuthering Heights*. Data mining is the first approach to be exploited to its fullest potential for obtaining enriched content to add to basic bibliographic descriptions.

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009

Another approach, very common in the field today, is to purchase enriched content from a vendor or through the library's ILS supplier. Cherie Madarash-Hill and J.B. Hill's 2005 study of the results of adding enrichment data obtained through a library's ILS vendor is a good place to start looking into this approach.[14] A third method is to pull the enriched content into the catalog using an application programming interface (API) such as the one provided through Amazon Web Services.[15] Other APIs are available as well.

Relevance Ranking, Classification and User Transaction Data

The state of relevance ranking in library catalogs does not appear to meet the expectations of today's end users. It is somewhat surprising that more has not been accomplished, given the research and publications of many information scientists that followed on the appearance of Melvin Maron and John Kuhns' article on "probabilistic indexing" in 1960.[16] Maron and Kuhns were early explorers of how to automate information search and retrieval. Their experiments tested probabilistic indexing, an automated document retrieval technique they invented. Their experimental process ingested user query terms, matched them to weighted keyword index terms assigned to documents, made statistical inferences, and then selected and organized documents in order of their probability of being relevant to the user's query. The technique produced promising results and engendered a good deal of related research in information science. Some promising research by Ray R. Larson and others on the application of relevance ranking techniques to library catalogs appeared in the late 1990s[17] but gained little traction with ILS vendors. Larson's relevance ranking techniques were especially interesting because he experimented with "classification clustering" to help focus subject searches. In effect, Larson and his team were using the terms associated with classification numbers to assist in ranking search results. More recently, staff at the National Library of Australia explored the possibility of mining the structured data of the catalog to produce better ranking of search results for Libraries Australia catalog searchers.[18] Using the fielded data available in MARC records plus holdings information, the staff reported reasonable success in refining the relevance ranking of search results in Libraries Australia, implementing their catalog changes in November 2008.

This study's results suggest that what drives end-user expectations of relevancy ranking in library catalogs are end-user experiences on Google and sites like Amazon.com. Google's incredible success is largely based on its ability to rank search results based on a complex algorithm of what links to what on the Web. Amazon's excellent relevance ranking of search results appears to be based on patterns of searching and buying and on user-contributed data.

Recommendations

As noted in the executive summary, this report is written for those readers who are seeking to define requirements for improved catalog data and for those readers who have a part to play in contributing, ingesting, syndicating, synchronizing or linking data from multiple sources in next-generation library catalogs. It is for those readers that the following recommendations are made:

- Examine and compare the library's investments in bibliographic work, catalog management, linking functionality and enrichment content (tables of contents, summaries, etc.) and rebalance them as appropriate to better meet end users' catalog data quality requirements.

- Within the library community and with relevant organizations, explore how to obtain or produce enrichment content (tables of contents, summaries, etc.)

through data mining, the use of APIs, partnerships with publishers and vendors, and collaborative library projects.

- Encourage the appropriate organizations to complete research and development to improve relevance ranking in online catalogs. Explore the feasibility of redeploying classification data (and the terms associated with classification numbers) and other existing data to improve relevance ranking.

- Pay more attention to the library's delivery services and the data elements that support a positive experience for the end user.

  - Explore the feasibility of home or office delivery of materials, or other faster, more convenient options for the delivery of physical items. Be willing to change policies and workflows, and to change how library human and financial resources are deployed to make this happen.

  - Explore consortial borrowing and lending options. As appropriate, be willing to link from the catalog to purchase options and to digitized materials on demand. Consider setting a library standard (e.g., within 24 hours) for getting requested materials to users.

  - Invest effort in improving the library's linking metadata management and interoperability with licensed and open-access data.

  - Consider adding more digital content that end users can easily link to from the catalog, and associate print descriptions with their digitized counterparts.

  - Where possible, link from the catalog to excerpts or snippets, both text and media.

- Examine the local editorial changes being made to bibliographic records and analyze which ones directly assist end-user discovery and delivery, and which do not matter as much. Redesign procedures and workflows to focus human expertise on what matters most to end users and which must be done manually.

- Libraries will not be able to accomplish what is needed by going it alone. Collaboration and a coordinated approach involving many organizations (and even end users) will be required. As noted in section 1 of the Library of Congress Working Group final report, traditional library workflows, featuring the same corrections being done multiple times at multiple libraries, are costly and redundant.[19] The right mechanisms for collaboratively sharing the effort of data quality improvements could assure better experiences for end users of catalog data at less cost to libraries.

- Within the library professional community and to the extent possible, look for ways to automate the production and maintenance of the structured data that supports collocation, faceting and advanced search features in the catalog.

- Explore ways for the library community to collaborate with the appropriate organizations to link and cross-reference standard numbers to support both library and end-user tasks.

A Few Ideas for Further Research

Researchers who want to build upon the findings presented in this report may wish to consider the following ideas for further study. These suggestions range from rather simple and practical projects to more advanced research undertakings.

- Produce a guide for libraries that describes the options for adding more enrichment data (tables of contents, summaries, etc.) to bibliographic descriptions in library catalogs. Compare the costs and benefits of a variety of approaches (manual enrichment, data mining, vendor data, APIs, etc.).

- Explore how classification, subject terminologies and structured data in the catalog might be used to improve relevancy ranking in online catalogs.

- Experiment with and assess the outcome of using more economical means (automation, data mining, user-contributed data, etc.) to support faceting and other enhanced search options. Greta de Groat's 2009 Digital Library Federation-commissioned report is a rich source of ideas for data mining, clustering, mapping and other tools to automatically enhance metadata for finding and using digital collections.[20]

- This study focuses on what end users and librarians say they want. Appropriately designed transaction log analyses, which measure what a target population actually does, could usefully complement the research reported here.

Notes

1. Cutter, Charles A., W. P. Cutter, Worthington Chauncey Ford, Philip Lee Phillips, and Oscar George Theodore Sonneck, *Rules for a dictionary catalog (*Washington: Government Printing Office, 1904).

2. Joint Steering Committee for the Development of RDA, *Full Draft of RDA,* Chapter 1, 2008, http://www.collectionscanada.gc.ca/jsc/rdafulldraft.html.

3. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional requirements for bibliographic records: final report,* UBCIM publications, new ser., v. 19 (München: K.G. Saur, 1998).

4. Indiana University, "Variations/FRBR: Variations as a Testbed for the FRBR Conceptual Model," http://www.dlib.indiana.edu/projects/vfrbr/. FRBR project Web site, Kent State University School of Library and Information Science, http://frbr.slis.kent.edu/projectinfo.htm.

5. Miksa, Fran, "Information organization and the mysterious information user" (lecture delivered at the School of Information, University of Texas—Austin), forthcoming in *Libraries and the Cultural Record.*

6. Weinberger, David, *Everything Is Miscellaneous* (New York: Times Books, 2007), 105.

7. Coyle, Karen, blog post February 21, 2007, http://kcoyle.blogspot.com/2007/02/i-have-long-had-hard-time-with-frbr.html.

8. Horton, Valerie, "Moving Mountains and Crossing Rivers" (report from the second conference on library physical delivery, Cincinnati, Ohio, September 2008), *Collaborative Librarianship* 1 (1) (2009): 18–26, http://www.collaborativelibrarianship.org/index.php/jocl/article/viewArticle/3.

9. Markey, Karen, and Karen Calhoun, "Unique words contributed by MARC records with summary and/or contents notes," in *ASIS '87: Proceedings of the 50th ASIS Annual Meeting, Boston, Mass.* October 4–8, 1987 (Medford, NJ: Published for the American Society for Information Science by Learned Information, 1987).

10. Ibid.

11. See the literature review in Markey, Karen, "The online library catalog: paradise lost and paradise regained?" *D-Lib Magazine* 13 (12) (2007), www.dlib.org/dlib/january07/markey/01markey.html.

12. Morris, Ruth C., "Online tables of contents for books: effects on usage," *Bulletin of the Medical Library Association* 89 (1) (2001): 29–36.

13. Author correspondence with Diane Vizine-Goetz, Consulting Research Scientist, OCLC, January 9, 2009 and FictionFinder, http://www.oclc.org/research/projects/frbr/fictionfinder.htm.

14. Madarash-Hill, Cherie, and J. B. Hill,"Electronically enriched enhancements in catalog records: A use study of books described on records with URL enhancements versus those without," *Technical Services Quarterly* 23 (2) (2005): 19–32.

15. As described by Ferraro, Joshua, blog entry "Amazon.com Web Services and library catalogs," April 16, 2007, http://blogs.liblime.com/developers/2007/04/16/amazoncom-web-services-and-library-catalogs/.

16. Maron, M. E. and J. L. Kuhns, "On relevance, probabilistic indexing and information retrieval," *Journal of the ACM* 7 (3) (1960): 216–244.

17. Larson, Ray R., Jerome McDonough, Paul O'Leary, Lucy Kuntz, Ralph Moon, "Cheshire II: Designing a next-generation online catalog," *Journal of the American Society for Information Science* 47 (7) (1996): 555– .

18. Dellitt, Alison and Tony Boston, "Relevance ranking of search results from MARC-based catalogues: from guidelines to implementation, exploiting structured data," National Library of Australia staff paper, 2007, http://www.nla.gov.au/nla/staffpaper/2007/documents/Boston_Dellit-relevance-ranking.pdf.

19. Library of Congress, *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*. Washington, D.C.: Library of Congress, recommendation 1.1, "Eliminate redundancies," 2008, http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

20. de Groat, Greta, "Future directions in metadata remediation for metadata aggregators," Digital Library Federation, 2009, http://www.diglib.org/aquifer/dlf110.pdf.

OCLC™

6565 Kilgour Place
Dublin, Ohio 43017-3395
1-800-848-5878   +1-614-764-6000
Fax: +1-614-764-6096
www.oclc.org

ISO 9001 Certified