Reexamining Content-Enriched Access: Its Effect on Usage and Discovery

Yuji Tosaka and Cathy Weng

Content-enriched metadata in bibliographic records is considered helpful to library users in identifying and selecting library materials for their needs. The paper presents a study, using circulation data from a medium-sized academic library, of the effect of content-enriched records on library materials usage. The study also examines OPAC search transactions of circulated items to learn how enriched metadata is used. The findings show that enhanced records were overall associated with higher circulation rates and that keyword search was the most frequently used search option directly associated with circulation. Contents data can play a key role in discovery. Libraries should continue to provide and exploit content-enriched metadata. The combination of optimal library system data mining capability, postsearching evaluation, and OPAC display are crucial to achieve content-enriched access.

Introduction

Content-enriched metadata in bibliographic records is helpful to library users in identifying and selecting library materials for their needs. The title and subject have long been the two basic elements that users consult to learn the content of a bibliographic item. Content-enriched data go beyond the title and subject of a bibliographic resource to include additional components, such as contents notes, summaries, links to tables of contents, sample text, and publication-related information. For decades, libraries have employed various methods in an effort to enhance bibliographic data in the belief that users benefit from content-enriched data. Indeed, many recent user studies published in the library literature have shown that library users today—influenced by Internet search engines, online bookstores, and seamless access to full-text resources—are more than ever demanding enhanced content and functionality in library catalogs to assist their discovery of relevant search results and resources.[1]

Content-enriched metadata is valuable in many respects. Bibliographic records provided with essential, content-enhanced descriptive data "can serve to increase the descriptive quality of the bibliographic record."[2] Content-enriched metadata contains many searchable, subject-related unique terms that enhance the retrieval of relevant titles. The presence of content-enriched data in bibliographic

Yuji Tosaka is Cataloging/Metadata Librarian and Cathy Weng is Head of Cataloging at The College of New Jersey Library; e-mail: tosaka@tcnj.edu, weng@tcnj.edu. *We would like to thank Jeffrey Weng for his editorial assistance and the anonymous reviewers for their invaluable comments and suggestions.*

displays of online public access catalogs (OPAC) helps users determine a particular item's relevance to their needs without having to go to the library and examine the item itself.

How content-enriched metadata benefits library users and enhances access to library materials was not clear until Van Orden introduced the concept of "content-enriched access" in an article published in 1990:[3] "Well-selected content components and full-text materials in electronic systems must be linked with improved search methodologies, better computer interfaces, and greater understanding of the structure and use of knowledge."[4] Most important, "determining which content components contribute the most value to initial searching and post-retrieval evaluation is a key to planning cost-effective systems."[5] In essence, merely adding content-bearing elements to bibliographic records is not sufficient. To achieve content-enriched access, it is necessary to have a well-designed data-mining mechanism to dig out content-enriched components and to connect those components to system retrieval ability and postsearch evaluation. This combination of a data-mining mechanism and its connection to retrieval and evaluation enables logical relevance ranking of retrieval results. Truly relevant titles can then be retrieved and delivered to end users.

In this paper, we examine and analyze circulation and bibliographic data to determine if any correlation exists between titles with content-enriched data and circulation rate. We also look at OPAC search history to see how bibliographic data are used during retrieval and how content-enriched access can be achieved. As shown in the recent report by the Library of Congress Working Group on the Future of Bibliographic Control, the library profession has not produced a "persuasive body of evidence that indicated what parts of the record are key to user access success."[6] Such studies would enable cataloging professionals and library managers to "make informed judgments about how best to direct efforts to improve record quality" to better support user needs and expectations in the evolving information environment.[7] Toward that end, we hope to contribute to a better understanding of what and how data elements in bibliographic records are used by academic library users in the OPAC environment. We also hope that this evidence-based study can shed some light on the effect of additional content in bibliographic records on information retrieval and use patterns.

Brief Overview of Previous Studies

The inclusion of contents and summary notes in bibliographic records in the pre-MARC era was mainly for description rather than for access, because elements in the note area were not accessible in the card catalog environment. Contents notes, if included, were usually limited to multivolume titles with individual volumes bearing distinct titles. In the early age of online catalogs, when computer technology was not as advanced, system retrieval functionality was focused on known item searches. Keyword free-text searching of the whole catalog, which usually produced high recall along with a non–user-friendly arrangement of retrieved entries, was considered ineffective and was rather discouraged. In spite of this, the value of having content-enriched bibliographic data in the catalog was noted in a number of studies.

In their 1987 paper, Markey and Calhoun examined 1,010 records with a summary and/or contents notes.[8] The purpose of the study was to determine the average number of unique subject words in bibliographic records that successively contributed to the search process. They found that, among the records examined, the contents and/or summary notes contributed an average of 15.5 unique terms per record (45 percent)—the largest number compared to other access points. Other related studies also revealed that content-enhanced

records resulted in higher retrieval rates of relevant items. The landmark study of the correlation between enhanced record and retrieval was Cochrane's SAP (Subject Access Project) study conducted at the University of Toronto library.[9] The study analyzed searches of more than 2,000 content-enhanced records in the social sciences and humanities. Ninety controlled searches were performed in the enhanced database, and these searches resulted in both higher recall and precision rates compared to searches in the unenhanced file. In their studies published during the 1990s, Dillon and Wenzel, Michalak, and Poulsen also found that adding content-enriched information to bibliographic records resulted in a significant increase in the number of items retrieved.[10]

The potential benefits to users from the presence of content-enriched data have also been explored. In their 1983 article, Cochrane and Markey pointed out that users wanted "the ability to search books' tables of contents, summaries, or indexes."[11] More than two decades later, a study performed by OCLC found that "end users rely on and expect enhanced content including summaries/abstracts and tables of contents"[12] and that "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."[13] In their 2006 article, Dinkins and Kirkland also noted that a table of contents enhances the descriptive quality of the bibliographic record.[14] Additionally, "the presence of additional access points (beyond author, title, and subject headings) improves the likelihood of retrieving that record, and also increases the patron's success at determining the book's relevance."[15]

Studies Related to Usage

Knutson's 1991 study was the first to empirically connect enhanced records with circulation.[16] He examined 291 selected records in the social sciences area (Library of Congress Classification schedule H) at the University of Illinois at Chicago Library. The sampled records were divided into three groups: the Enhanced Group with added subjects and contents notes; the Control Group with no added subjects and contents; and the Contents group with no added subjects but with full contents notes. The study revealed that of the 98 titles that were circulated, "the data all point towards the likelihood that the added subjects for the Enhanced group did influence circulation,"[17] whereas no significant difference in circulation was detected in the Control Group and the Contents Group. Knutson also pointed out other factors that might have had an impact on usage: system keyword searching functionality and OPAC display.

Conversely, Morris's study conducted at the University of New Mexico Health Sciences Center Library in 1998 yielded positive and encouraging results.[18] The study found that titles with enhanced data (tables of contents) showed an increase in usage. "Online tables of contents in book records increases [*sic*] the likelihood of in-house use by 43%; the presence of online TOC increases the likelihood of circulation by 33%."[19] The study findings also suggested that the currency of the titles and the previous usage history were two other factors that affected circulation.

In 2004, Madarash-Hill and Hill performed a use study at Southeastern Louisiana University Library.[20] The purpose was to find out if records with URL enhancement links experienced higher usage. The study sampled and examined two sets of online catalog records, both those with and those without URL enhancement links. Circulation data of the two sets of records were compared. The average records with enhanced URL links had a higher circulation rate than those without, with 93 percent and 79 percent respectively. However, the study dealt with a relatively small size of sample records (180 records), and its criteria of extracting tested records (based on subject terms) were not designed to sample titles from all subject fields. Furthermore,

the results did not specify the publication date of circulated books, which is considered an influential factor in usage rates. Madarash-Hill and Hill also found that the inclusion of searchable elements of tables of contents and summary data might also have contributed to higher usage.

Hill and Madarash-Hill had also conducted a similar study at the University of Akron Libraries in 2002.[21] The study's purpose was to find usage data of circulating IEEE conference proceedings records with full-text links to electronic resources. The study revealed that the usage of content-enriched records with full-text links was approximately four times higher than the usage of records that did not have content-enriched data. The study suggested that "add[ing] IEEE Xplore full-text links and TOC enhancements to bibliographic records in the online catalog can greatly increase the accessibility of IEEE conference proceedings."[22] The key was that searchable tables of contents enhanced materials' discoverability.[23]

Dinkins and Kirkland's 2006 study at Stetson University was a result of a record enhancement project (that is, adding tables of contents to bibliographic records).[24] For the purpose of the project, over 2,500 records were selected for enhancement. Circulation statistics before and after the project were then compared. The study found a 5 percent increase in the circulation of titles after their records had been enhanced (32 percent of titles circulated before the study, whereas 37 percent of titles circulated after the study). However, the authors stated that it was not clear if the added tables of contents contributed as much to the increase in circulation as other variables, such as publication date, acquisition date, location, and OPAC search results display. These variables needed to be controlled for a full study to be conducted.

In their 2007 article, Faiks, Radermacher, and Sheehan focused more on catalog accessibility and discoverability.[25] The authors believed that, by adding tables of contents, which were considered additional access points, the library catalog ultimately would result in greater retrieval. In their usage study at the Cooperating Libraries in Consortium, tables of contents and summary notes were added to bibliographic records, and circulation statistics were compared before and after the change. While no detailed numbers and study procedures were given for this study, "the percentage increase in circulation after TOCs/Summary Notes were added was 20.40%."[26] The authors also mentioned that more data gathering and analysis needed to be done to explore additional usage factors, such as subject areas and OPAC functionalities.

The Current Study

The present study seeks to expand on the previous studies of the effect of content enrichment on circulation. The authors wanted to see if the effects would be similar or different in different fields. Additionally, we were interested in the way users did searches and how they used content-enriched data. We focused on two questions:

1. Do content-enriched records have an impact on circulation in various subjects? And how?
2. Are the content-enriched metadata of circulated titles being used during OPAC retrieval? And how?

The study comprises two parts. The first part focuses on the correlation between content-enriched records and circulation rates. The second part analyzes individual OPAC search transactions to determine the role that enhanced data plays in OPAC retrieval.

The Environment

The College of New Jersey (TCNJ) is a four-year residential college located in Ewing, New Jersey. The college has a full-time enrollment of over 6,000 undergraduate students, along with several small graduate programs. It has a wide variety of degree programs, with courses offered through seven schools covering various

disciplines: Arts and Communication, Business, Culture and Society, Education, Science, Nursing, Health, and Exercise Science, and Engineering. The library is a mid-sized academic library designed to support campus learning, teaching, and research, and it holds more than 600,000 volumes in its collection.

Methodology—Part I

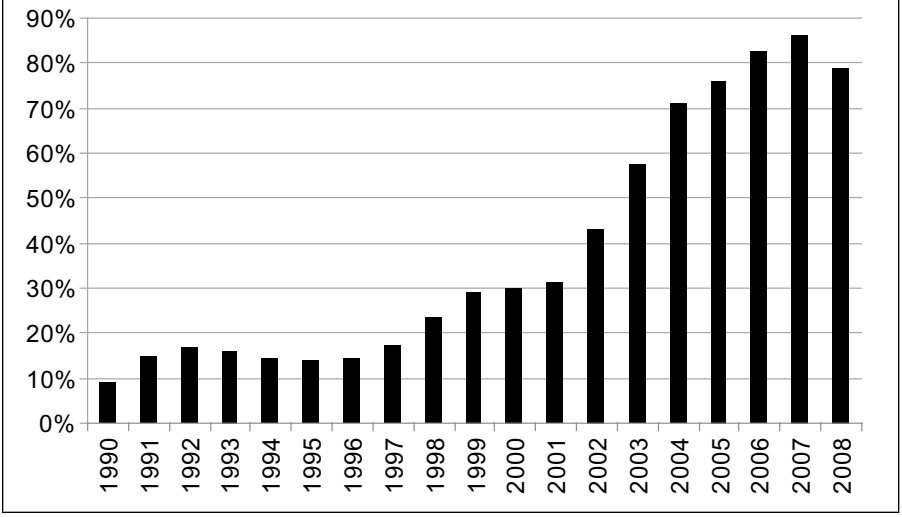
For this study, we decided to investigate the usage of titles with content-enriched records across four subject areas in our circulating print collections—history, social sciences, language and literature, and science and technology. Data for the study were taken from circulation transaction logs from January to May 2009. As a basis for calculating circulation rates of checked-out titles, the latest data were extracted from the local Voyager system. As a way to compare circulation rates of titles with and without enhanced records, the circulation and collection data were examined from the DEF, HJL, P, and QRST classes of the Library of Congress Classification System, classes that matched the four subject areas chosen for this study. The units of analysis in this study were (1) bibliographic records in the collection data and (2) circulation transactions of a bibliographic record in the circulation data.

The extracted data were organized into a set of two raw data files—bibliographic records and circulation transactions—for each group of LC classes. Each file had data values for publication dates and MARC 505, 520, and 856 fields (respectively: tables of contents, summaries, and URL links to remote tables of contents, summaries, and other publication-related information). E-books were not included in this study because tracking their electronic usage was outside the scope of the current study. We decided to limit our analysis in this paper to titles published since 1990 because content-enriched records, as we will see later, did not reach levels appropriate for meaningful comparison with nonenhanced records until the 1990s. With the coding rules applied to the raw data, the final data had 88,538 titles and 7,782 circulation transactions.[27]

Then we took two steps to create a table summarizing each data file. First, we determined the total number of titles containing each content-enriched field. We arranged both the collection and circulation data by publication date to chart over time the percentage of records containing enhanced data. Second, we created aggregate variables to see if the effect of one field on the circulation rate might have depended on the presence of another field. For example, if records with MARC 856 fields correlated with an overall positive effect on collection use, closer examination might reveal that circulation was higher only when those records also contained 505 fields, while no effect was found when records were enhanced with 856 fields only. In that case, we can be reasonably confident that the presence of 856 fields itself was not associated with higher circulation.

The intent of our study was to determine the effect of enhanced metadata on circulation rates. To do this, we treated books without any enhanced records as a baseline and calculated their circulation rates for comparison. We then looked at books with enhanced records and considered how their circulation rates deviated from that baseline. For instance, let us imagine that there were 1,000 books without enhanced metadata in one broad subject area at TCNJ Library. A total of 100 books were checked out from January to May 2009. The circulation rate would then be 10 percent. Then let us imagine that there were another 1,000 books in that subject area with enhanced data added to their records. If 130 books were checked out over the same 5-month period, their circulation rate would be 13 percent. Although it may be natural to think that the difference between 13 percent and 10 percent was 3 percent, a more appropriate measurement for comparison is to treat 10 percent as the "baseline" circulation rate, and that thirty more books repre-

FIGURE 1
Percentages of Content-Enriched Records, 1990–2008



sented a 30-percent *relative difference* in circulation. The relative difference allows a more meaningful comparative analysis of categories of circulated materials, because they enable us to calculate the effect of enhanced records on circulation in a consistent, comparable manner.

Study Findings—Part I

Based on our data analysis, there was a marked increase in content-enriched records starting in the early 1990s (see figure 1). For titles published in 2008, nearly 80 percent of their records contained at least one content-enriched field. Further breaking down the data (see figure 2), we found that the presence of MARC 505, 520, and 856 fields all increased between 1990 and 2008, but at different rates. Records containing 505 fields comprised the overwhelming majority of enhanced records until 2000. Thereafter, records containing

FIGURE 2
Percentages of Content-Enriched Fields, 1990–2008

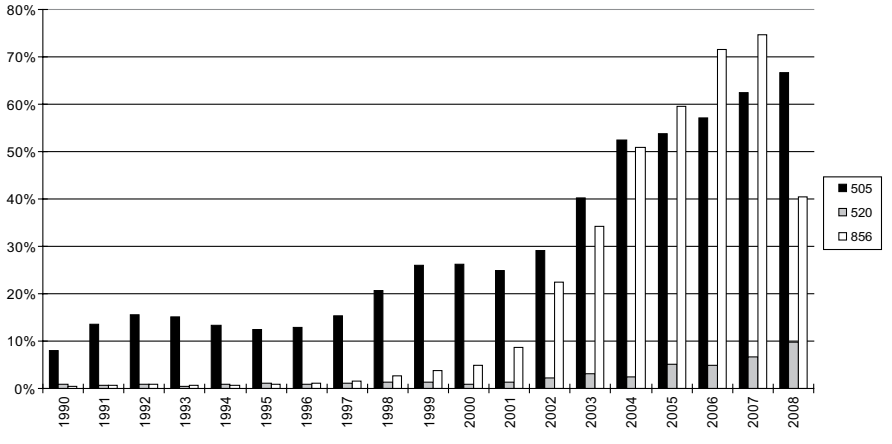


TABLE 1
Effect of Enhanced Records on Circulation Rates, by Discipline, 1990–2008

Publication Dates	% of Enhanced Records	Effect (Relative Percentage Difference)			
		History	Social Sciences	Language & Literature	Science & Technology
1990–94	14.3%	18.0%	50.0%	60.7%	25.0%
1995–99	19.3%	32.7%	99.0%	27.0%	50.3%
2000–04	45.8%	34.0%	25.7%	30.6%	36.9%
2005–08	80.9%	21.5%	–16.2%	18.6%	–10.6%

856 fields increased dramatically. Those with 520 fields started to increase in the mid-2000s, but they still accounted for no more than 10 percent of titles published in 2008.

In our analysis (see table 1), we found that, for the 1990–2004 titles, the effect of enhanced data on circulation rates were uniformly positive across all four fields. The 1990–1999 data were highly variable, however. The data for titles published between 2005 and 2008 were mixed. The effect of enhanced records was as highly variable as the 1990–1999 data but ranged from a circulation rate in Social Sciences that was 16.2 percent *lower* to a circulation rate in History that was 21.5 percent *higher*.

The high degree of variability in the 1990–1999 and 2005–2008 data is expected in a study comparing percentages using real-world library data. As shown in table 1, enhanced records comprised less than 20 percent of all bibliographic records in the 1990–1999 period, while nonenhanced records constituted a similarly small data subset in the 2005–2008 period. In such situations, a small difference in the number of circulated titles, especially when further divided by subject field, were apt to make an overstated difference in the relative circulation rates. In contrast, the 2000–2004 data, which show relatively similar positive figures (25.7 to 36.9 percent) across the four subject fields, should provide the most accurate measure of the overall effect of enhanced records on circulation vis-à-vis nonenhanced records. The proportion of enhanced and nonenhanced records was almost equal, with the result that each title circulated had roughly the same effect and made their circulation rates more comparable.

To deal with the problem of considerable gaps in the number of enhanced and nonenhanced records in the 1990–1999 and 2005–2008 data, we aggregated the circulation rate data in each subject area and examined the effect of enhanced records on circulation across all four subject

TABLE 2
Effect of Enhanced Records on Circulation Rates, Aggregate Data, 1990–2008

Publication Dates	% of Enhanced Records	Circulation Rate			
		Enhanced Records	Nonenhanced Records	Percent Differences	Effect (Relative Percentage Difference)
1990–94	14.3%	8.8%	6.0%	2.7%	45.5%
1995–99	19.3%	10.1%	6.5%	3.6%	55.6%
2000–04	45.8%	12.3%	9.4%	2.9%	30.7%
2005–08	80.9%	13.8%	14.8%	–1.0%	–4.0%

TABLE 3
Effect of MARC 520 Fields on Circulation Rates, 2000–2008

Publication Dates	Records with 520 Fields (%)	Circulation Rate	Effect (Relative Percentage Difference)	Records Enhanced with 520 Fields Only (%)
2000–04	2.0%	14.6%	54.7%	0.8%
2005–08	6.3%	17.0%	14.7%	0.9%

areas (see table 2). We found that, for titles published between 1990 and 2004, the effect of enhanced records was overall positive, with approximately 30–55 percent higher circulation than those with nonenhanced records. No notable effect was observed for titles with enhanced records published between 2005 and 2008, however.[28]

Another question we explored is how each content-enriched field contributes to higher library materials usage. The results suggest that the presence of MARC 520 or 856 fields had little notable effect on circulation. As shown in table 3, recent titles with records containing 520 fields had higher circulation as a whole. However, because the number of records containing only 520 fields was too small (less than 1 percent), we cannot yet make a meaningful analysis of the circulation effect of 520 fields themselves. For titles with records containing 856 fields (see table 4), we found that their circulation rate was 11.7 percent higher than those with nonenhanced records for the period 2000–2004. However, it was not the presence of 856 fields alone that was correlated with an elevated circulation rate. Only when we looked at all records with 856 fields (including records with other enhanced fields) did we see a noticeable positive effect on circulation. When we likewise ruled out the effect of the other enhanced fields, we even found a negative effect for titles published between 2005 and 2008. One possible reason for this result is that both nonenhanced records—baseline for comparison—and records enhanced with only 856 fields comprised less than 20 percent of all bibliographic records in this period (see also table 2). The data might have been skewed by the problem of comparing percentages between the two small subsets in our real-world circulation data.

Because, as the current study suggests, MARC 520 and 856 fields did not contribute to higher circulation, we can be reasonably confident that the 505 field was the major factor leading to higher materials usage in general. In addition, it should be evident that, for earlier publications, the positive effect of enhanced records resulted largely from the presence of 505 fields anyway, because the number of records with the other enhanced fields did not start to increase until after 2000 (see figure 2). As a result, we focused our analysis on the effect of 505 fields on circulation in greater detail (see table 5). As expected, records containing 505 fields had a higher effect overall on

TABLE 4
Effect of MARC 856 Fields on Circulation Rates, 2000–2008

Publication Dates	Records with 856 Fields (%)	Circulation Rate	Effect (Relative Percentage Difference)	Records Enhanced with 856 Fields Only (%)	Circulation Rate	Effect (Relative Percentage Difference)
2000–04	23.3%	11.7%	24.0%	10.7%	9.4%	0.3%
2005–08	63.6%	13.3%	–6.2%	20.0%	11.8%	–14.5%

TABLE 5
Effect of MARC 505 Fields on Circulation Rates, 1990–2008

Publication Dates	Records with 505 Fields (%)	Circulation Rates			
		Records with 505 Fields	Nonenhanced Records	Percent Differences	Effect (Relative Percentage Difference)
1990–94	13.2%	8.6%	6.0%	2.6%	42.6%
1995–99	17.1%	10.1%	6.5%	3.6%	55.3%
2000–04	34.1%	13.3%	9.4%	3.9%	41.3%
2005–08	59.1%	14.5%	14.8%	–0.3%	–0.8%

circulation for the 1990–2004 period, a result that matched the aggregate effect of content-enriched records (see table 2). Likewise, no notable effect on circulation was found for the most recent 2005–2008 publications.

These results raise obvious questions about why the effect of content-enriched records appears to level off when their percentage has increased considerably for the latest publications. There are a couple of plausible explanations for these associations. One simple explanation is that while type of record (that is, enhanced/nonenhanced) is the only variable examined in this study, there may be another variable, for the most recent publications in particular, that better explains collection use. As Manning, Raghaven, and Schütze rightly pointed out in their recent work on information retrieval, "relevance is assessed relative to an information need, *not* a query."[29] In other words, additional content in a bibliographic record returned by a query might be of much lower value for users in determining the item's relevance for their particular purposes. One such variable that likely far outweighs the potential effect of content-enriched records might be the age of material. The users might only want to consider the latest publications.

That library users prioritize recentness seems to fit many classic studies that have found a high positive correlation between publication date and collection use.[30] Our circulation figures as shown in table 6 also provide strong evidence about their correlation. In Social Sciences and Science and Technology, the titles published since 2000 indeed accounted for nearly half of the total library materials usage. In History and Language and Literature, such titles accounted for a lower percentage of total collection use (approximately 25 percent each), suggesting that users in these fields prioritize recentness less. And yet, the importance of the age of material

TABLE 6
Collection Use in TCNJ Library, by Publication Date

Publication Dates	History	Social Sciences	Language & Literature	Science & Technology
–1979	36.2%	18.1%	34.3%	17.8%
1980–89	15.0%	10.5%	14.7%	12.3%
1990–99	22.0% (9.3%)	26.0% (5.8%)	27.0% (9.2%)	24.3% (5.1%)
2000–	26.8% (14.1%)	45.3% (11.1%)	24.0% (12.4%)	45.6% (11.0%)

Note: The circulation rates of titles published in the 1990–1999 and 2000–2008 periods are within parentheses. Percentages may not add up to 100% due to rounding.

FIGURE 3
Sample Circulation Log for September 22, 2009

Q	Vertebrate hard tissues / L. B. Halstea	1	22-Sep-09	47274	1974	QM569 .H32

is seen clearly in the fact that, even in these humanistic disciplines, there was still a marked increase in circulation rates for titles published since 2000 (for instance, History: 14.1 percent vs. 9.3 percent [1990–1999]; Language and Literature: 12.4 percent vs 9.2 percent [1990–1999]). When users come to the library intending only to use newer publications, it is therefore possible that publication date might simply be used as the main evaluative filter for choosing among their search results, rather than additional content in a bibliographic record or the OPAC relevance rankings based on the occurrence of searched terms.

Another compounding factor that could be affecting our results is that it might not be possible to push circulation rates beyond a particular "saturation" level. TCNJ Library has a finite user base of students and faculty who use library materials mostly for their highly specific learning and research purposes. Also, the most recent publications, as discussed above, are already the most heavily used items. In such settings, it might be unrealistic to expect that collection use would be substantially higher than the current rate if we increased the percentage of enhanced records. This might be particularly true when a large majority of MARC records for newer publications already contain enhanced data anyway. Additional content in bibliographic records might be much less salient as a factor that might increase the accessibility of such records and thus contribute to higher circulation.

Methodology—Part II

For the purpose of the second part of the study, the circulation data of one randomly chosen day (September 22, 2009) were extracted and grouped into the same four subject categories (history; social sciences; language and literature; science and technology.) The OPAC transaction log of September 16–22 was also generated. That date range was chosen to capture searches that might have been conducted in advance of the checkout date. A total of 133 titles were checked out that day. Approximately 26,000 OPAC transactions were examined and analyzed. Three things were done to find possible search option(s) and search strings used to find circulated items:

We examined Voyager cataloging records of circulated titles for indications of access points possibly used to retrieve the circulated titles (see figures 3 and 4).

We checked against the OPAC transaction log using the access points obtained from the cataloging record for probable search options and strings (see figure 5).

We replicated the OPAC search using selected search strings recorded in the

FIGURE 4
Sample MARC Cataloging Record for the Circulated Title

245	1	0	‡a Vertebrate hard tissues / ‡c L. B. Halstead [assisted by R. Hill]
260			‡a London : ‡b Wykeham Publications ; ‡a Andover : ‡b distributed by Chapman & Hall, ‡c 1974.
300			‡a xi, 179 p. : ‡b ill. ; ‡c 22 cm.
490	1		‡a Wykeham science series
500			‡a Distributed in the U.S. by Springer-Verlag New York Inc., New York.
500			‡a Includes index.
504			‡a Bibliography: p. 170.
650		0	‡a Bone.

FIGURE 5
Probable Search Log Entry Identified for the Circulated Title

22-Sep-09	Keyword	(GKEY bone)	K	359	1	159.91.122.103

system to confirm that the titles checked out corresponded with those strings (see figure 6).

For the purpose of the study, the search options and strings that were likely to be directly associated with circulation were then recorded and analyzed. Findings are illustrated in tables 7 and 8. We also kept track of what types of searches were issued. It should be noted that, if a keyword search option was issued but obviously meant for a known item search (for instance, author, title search), we considered it a known item search and grouped into the appropriate category (author or title search, or any other categories). The keyword search category was strictly given to searches that were clearly issued intentionally as keyword searches. Furthermore, of the books in circulation, we counted the number of books with records containing data in the MARC 505 field, as well as the type of search that was used for such books. Out of the 130 circulated titles, three titles were placed in the "Other" category, as no associated searches could be detected.

In spite of their limited date range, the data still provided a fascinating picture of how users search for and find resources in an academic library. For this study, it was not possible to identify or interview the users who checked out the titles on the selected day to determine their specific intentions. In most cases, however, it was possible to analyze the search options and terms used from OPAC transaction records and circulation logs, as well as the times at which the searches were performed and the corresponding items charged, to see what those users were looking for and connect the searches they issued to the items checked out soon after. For instance, if a user had entered the term "beisner" as a keyword search, for the purposes of the present study, that search was classified as an author search (in other words, a search for the author "beisner") and it could be deduced that the person who issued that search was also the one who checked out his title

FIGURE 6
Sample OPAC Display of Replicated Search Results

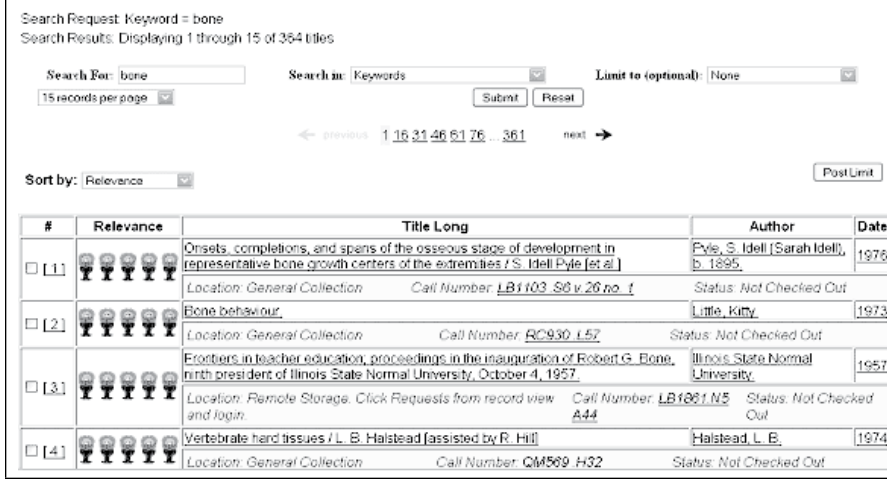


TABLE 7
Identified Search Options Used for Circulated Titles and the Presence of MARC 505 Fields, September 22, 2009

	Author	Title	Keyword	Other	Total
No. of Records	7 (5.3%)	54 (40.6%)	66 (49.6%)	6 (4.5%)	133 (100%)
Records with 505 Fields	3 (6.3%)	21 (43.8%)	23 (47.9%)	1 (2.1%)	48 (100%)

Note: Percentages may not add up to 100% due to rounding.

TABLE 8
Identified Search Option for Circulated Titles, by Subject, September 22, 2009

Search Option	History	Social Sciences	Language & Literature	Science & Technology
Author	1 (3.2%)	1 (3.7%)	5 (14.7%)	
Title	7 (22.6%)	12 (44.4%)	13 (38.2%)	23 (60.5%)
Keyword	23 (74.2%)	12 (44.4%)	15 (44.1%)	15 (39.5%)
Other		2 (7.4%)	1 (2.8%)	
Total	31 (100%)	27 (100%)	34 (100%)	38 (100%)

Note: Percentages may not add up to 100% due to rounding.

Twelve against Empire a few hours later the same day (see figures 7 and 8).

Study Findings—Part II

As can be seen in table 7, keyword search was the most frequently used search option, comprising about 51 percent of all searches. This finding coincides with the findings of the 2005 OCLC report that users today are significantly influenced by online search engines, where natural-language (keyword) search is constantly performed.[31] Additionally, as Yu and Young have noted, "the menu sequence for search options plays a significant role in user selection."[32] Since "Keyword" search was the default search option in the TCNJ OPAC, users naturally were led to begin their search with keyword search. Approximately half of the searches (46.9%) were known item searches, with the majority being title search. This suggests that a solid percentage of library users come to the library with a targeted item in mind. One interesting finding is that title search trumped other search options in the Science and Technology fields, whereas keyword search contributed to

FIGURE 7
OPAC Transaction Log of a Search Session (performed on 1:44 pm, Sept. 22, 2009)

2009092213440	22-Sep-09	Keyword	(GKEY beisner)	K	6	1	159.91.66.227

FIGURE 8
Circulation Log for the Corresponding Title (checked out on 4:41 pm, Sept. 22, 2009)

E	Twelve against empire; the anti-imperialists, 1898-1900, by Robert L. Beisner.	1	9/22/2009 4:41:46 PM	1968	E713 .B47

nearly 75 percent of the circulated items in the History field (see table 8).

To further understand how content-enriched data might be used during OPAC retrieval, it was essential to learn the connection between keyword search strings and retrieved records, as contents notes will only be searched when a keyword search is issued. For the purpose of the present study, contents notes in bibliographic records and searched keywords were examined to see if searched keywords appeared in the contents note field of the record of the circulated item. The data from our quick examination showed that the majority of searched terms appeared either in the title, subject, or contents note fields. However, more search terms appeared in the title and subject fields than in the contents note field, which suggests that subject and title data elements play a key role in OPAC relevance ranking, more so than the contents note element. This finding leads us to believe that enriched content data were not effectively and sufficiently used in searches. One record in particular was retrieved because the searched word only appeared in the contents note field, however, which was encouraging evidence of the usefulness of content-enriched data.

The underutilization of contents data in bibliographic records might have been the result of two factors. First, the TCNJ OPAC display was set to brief view default, thus making enhanced data less visible at first sight. Second, in the field weight system (designed for postsearch evaluation and relevance ranking), the table of contents field was given a relatively low weight.

The first potential factor is that the default TCNJ OPAC bibliographic display at the time of this study was a brief display that did not include content-enriched data. Users would have no way of knowing the presence of the table of contents in the bibliographic record from the brief display alone. It is likely that many users did not switch to full record display to learn more about the contents of a specific title, since such a process can be quite cumbersome. In that sense, content data are most likely used to retrieve a resource rather than to assess the relevance of that particular source.

The second factor contributing to the underutilization of contents data is that relatively low weight was given to the contents note field in the settings of the TCNJ online catalog when the study data was extracted. This might be key, because content data affects the relevance ranking of a bibliographic record. TCNJ library OPAC's keyword relevance ranking system is governed by, along with other factors, locally adjustable MARC field weights. When a search transaction occurs, a score is calculated based on the field containing the searched terms. (Other factors such as the frequency of words appearing in one record and the uniqueness of search terms in the whole database are hard-coded and cannot be modified locally.) During the study period, the table of contents field was given a relatively low weight in the field weight table. This could potentially affect the ranking of search results, as records with or without searched terms in the table of contents field could have been weighted similarly. Because of this similar weighting, more relevant titles might be buried low in the results, having been ranked among less relevant titles.

Conclusion

To summarize, this study suggests that content-enriched metadata overall contributes to higher circulation across the four subject fields. Content-enriched data also play an important role in OPAC discovery. Many libraries have incorporated content-enriched metadata into their workflow by either systematically entering them into their catalog or by purchasing vendors' record enhancement services. This can be seen from the higher percentage of content-enriched records input in bibliographic utilities and local catalogs in recent years.

As mentioned earlier, for content-enriched access to succeed to a great extent,

the combination of optimal library system data mining capability, postsearching evaluation, and OPAC display are crucial. Furthermore, displaying content-enriched data in OPAC with matching keywords highlighted is essential in helping users identify the resources they need. Many libraries still do not have this functionality enabled in their OPAC system. Some libraries still sort their keyword search results by publication date or system ID number. As a result, users are forced to work around the limitations of the system to find the resources they need.

While OPAC display of content-enriched data is highly desirable, it should be handled with caution. Lengthy content-enriched data in record displays can easily overwhelm users and generate a negative effect. Designing a more effective display of lengthy tables of contents should be a priority, which can be easily achieved using today's technology. Resolving the issue by not displaying it, as implemented by many libraries, should be discouraged. Another enhancement suggestion is to have tables of contents or summary data displayed as snippets with matching keywords highlighted in context in initial search results displays, as found in many Internet search engines. This enhancement will enable users to perform preliminary filtering of relevant items without going into individual record displays to find more content-related information.

The results of this study suggest some questions that point to possible avenues of further research. We found that publication date had a significant effect on material use, perhaps more significant than content-enriched metadata. One possible future study is to test whether any correlation exists between publication date and circulation when the majority of catalog records are content-enriched records. In addition, it will be important to see how library materials usage might change when desired enhancements are made in OPAC displays of content-enriched data.

Another possible future study is determining how our findings are applicable to other library settings, such as larger academic libraries or public libraries. Examining larger circulation datasets apart from our mid-sized academic library is necessary to achieve a broader view. Also, users of public libraries may employ a usage pattern that is different from that in academic libraries. It would be interesting to learn how different types of library users take advantage of content-enriched data during the retrieval process. User studies of different segments of the academic community, such as comparing the enhanced data usage patterns of graduate students to undergraduates or faculty members, would be another interesting avenue of research.

Libraries have invested resources and efforts in making content-enriched data available for end users. This seems to have been an encouraging trend. We would like to see such data being used to full advantage to support users' information needs in a robust and creative way. As we learned from our study, however, there is still a great distance between where content-enriched access is today to where it can be tomorrow. Only by continuing to provide content-enriched metadata and content-enriched access can users easily retrieve the library resources they need. The library community should commit to keeping the OPAC relevant in an evolving scholarly information landscape, where the quantity and variety of resources have proliferated on a massive scale. Enabling content-enriched access is a crucial part of keeping library catalogs relevant and making library materials easily discoverable among the myriad other resources in the larger digital environment.

Notes

1. Lynn Silipigni Connaway and Timothy J. Dickey, *The Digital Information Seeker: Report of Findings from Selected OCLC, RIN, and JISC User Behaviour Projects*, 35. Available online at www.

jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf. [Accessed 21 July 2010].

2. Debbi Dinkins and Laura N. Kirkland, "It's What's Inside That Counts: Adding Contents Notes to Bibliographic Records and Its Impact on Circulation," *College & Undergraduate Libraries* 13 (2006): 61.

3. Richard Van Orden, "Content-Enriched Access to Electronic Information: Summaries of Selected Research," *Library Hi-Tech* 8, no. 3 (1990): 27–32.

4. Ibid., 27.

5. Ibid., 29.

6. *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*, 14. Available online at www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf. [Accessed 28 April 2010].

7. Ibid., 14.

8. Karen Markey and Karen Calhoun, "Unique Words Contributed by MARC Records with Summary and/or Contents Notes," *Proceedings of the 50th Annual Meeting of the American Society for Information Science* (1987): 153–62.

9. Ruth C. Morris, "Online Tables of Contents for Books: Effect on Usage," *Bulletin of the Medical Libraries Association* 89 (2001): 29–36; Evan Pappas and Ann Herendeen, "Enhancing Bibliographic Records with Tables of Contents Derived from OCR Technologies at the American Museum of Natural History Library," *Cataloging & Classification Quarterly* 29, no. 4 (2001): 61–72; Van Orden, "Content-Enriched Access," 27–32.

10. Martin Dillon and Patrick Wenzel, "Retrieval Effectiveness of Enhanced Bibliographic Records," *Library Hi-Tech* 8, no. 3 (1990): 43–47; Thomas J. Michalak, "An Experiment in Enhancing Catalog Records at Carnegie Mellon University," *Library Hi-Tech* 8, no. 3 (1990): 33–42; Claus Poulsen, "Tables of Contents in Library Catalogs: A Quantitative Examination of Analytic Catalogs," *Library Resources & Technical Services* 40 (1996): 133–38.

11. Pauline A. Cochrane and Karen Markey, "Catalog Use Studies—Since the Introduction of Online Interactive Catalogs: Impact on Design for Subject Access," *Library & Information Science Research* 5 (1983): 337–63.

12. Karen Calhoun, Joanne Cantrell, Peggy Gallagher, and Janet Hawk, *Online Catalogs: What Users and Librarians Want, an OCLC Report* (Dublin, Ohio: OCLC, 2009), v.

13. Ibid., 11.

14. Dinkins and Kirkland, "It's What's Inside That Counts," 59–71.

15. Ibid., 61.

16. Gunnar Knutson, "Subject Enhancement: Report on an Experiment," *College & Research Libraries* 52 (1991): 65–79.

17. Ibid., 73.

18. Morris, "Online Tables of Contents for Books," 29–36.

19. Ibid., 34.

20. Cherie Madarash-Hill 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, no. 2 (2005): 19–31.

21. J.B. Hill and C. Madarash-Hill, "Enhancing Access to IEEE Conference Proceedings: A Case Study in the Application of IEEE Xplore Full Text and Table of Contents Enhancements," *Science & Technology Libraries* 24 (2004): 389–99.

22. Ibid., 398.

23. Ibid., 389–99.

24. Dinkins and Kirkland, "It's What's Inside That Counts," 59–71.

25. Angi Faiks, Amy Radermacher, and Amy Sheehan, "What about the Book? Google-izing the Catalog with Tables of Contents," *Library Philosophy and Practice* (2007). Available online at http://unllib.unl.edu/LPP/faiks.htm. [Accessed 14 January 2010].

26. Ibid.

27. To obtain an accurate count of bibliographic records, the following coding rules were set. First, item records from multivolume titles or duplicate copies were treated as duplicate records and counted as one bibliographic record. Second, for the circulation data, when duplicate copies were checked out on the same day, such circulation transactions were treated as separate cases. For the purpose of the study, a large outlier data in the P class published in 2006 was detected and removed from the analysis. In the category in question, circulation statistics increased approximately 350 percent over the previous year. This unusual observation was accounted for by a French-language learning CD-ROM set, which was checked out 318 times over the five-month period. Since we found no other similarly unusual multiple checkout case, we decided to remove the 318 checkouts from the circulation transactions data for 2006 so that they would fall within the possible range of values normally expected from the circulation data in the preceding and

following years.

28. For the circulation rate data, the 2005–2007 data are substituted for the 2005–2008 data in all tables, because circulation would be calculated at lower rates if the number of checkouts in the first five months of 2009 were simply divided by the latest volume of our library collection, which includes a large number of subsequent acquisitions published in 2008. In contrast, the effect (relative percentage difference) columns for the year 2005–2008 simply compare the 2005–2008 circulation rate data of nonenhanced records and enhanced records (or their subsets).

29. Christopher D. Manning, Prabhakar Raghaven, and Hinrich Schütze, *Introduction to Information Retrieval* (Cambridge, U.K.: Cambridge University Press, 2008), 152.

30. F.W. Lancaster, *The Measurement and Evaluation of Library Services* (Washington, D.C.: Information Resources Press, 1977).

31. Cathy De Rosa, Joanne Cantrell, Diane Cellentani, Janet Hawk, Lillie Jenkins, and Alane Wilson, *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership* (Dublin, Ohio: OCLC, 2005), 17. See also Connaway and Dickey, *The Digital Information Seeker*, 27–28.

32. Holly Yu and Margo Young, "The Impact of Web Search Engines on Subject Searching in OPAC," *Information Technology and Libraries* 23 (2004): 176.