

# Enriching Traditional Cataloging for Improved Access to Information: Library of Congress Tables of Contents Projects

John D. Byrum Jr. and  
David W. Williamson

*Traditionally, standard catalog records have provided bibliographic data that mostly address the basic features of library resources. At the same time, catalogs have offered access to these records through a limited array of names, titles, series, subject headings, class numbers, and a relatively small number of keywords contained within descriptions. Today's catalog users expect access to information well beyond what can be offered by traditional approaches to bibliographic description and access. By pursuing a suite of projects, the Library of Congress (LC) has responded to the challenge of enticing patrons to continue to include the online catalog among the tools they use for information retrieval. Drawing extensively on the power of automation, staff of LC's Bibliographic Enrichment Advisory Team (BEAT) have created and implemented a variety of initiatives to link researchers, catalogs, and Web resources; increase the content of the catalog record; and link the catalog to electronic resources. BEAT's ongoing work demonstrates how, in the electronic era, it is possible to provide new and improved ways to capitalize on traditional services in the digital age. This paper will illustrate these points by focusing on BEAT's tables of contents projects to demonstrate how library automation can make significant bibliographic enhancement efforts quick, easy, and affordable to achieve.*

In 1992 the Library of Congress's (LC's) Director for Cataloging established the Bibliographic Enrichment Advisory Team (BEAT) to conduct research and undertake initiatives to enhance the utility of bibliographic records. Composed of voluntary staff from a variety of service units, the team was urged to work outside the box and exempted from the restraints of many policies and practices pertaining to traditional cataloging activities. BEAT was also mandated to create and use automated methods to accomplish its work due to the impact of shrinking staff resources in the bibliographic access divisions.

Among BEAT's earliest undertakings was the development of a series of projects to focus on enriching bibli-

ographic records to include tables of contents (TOCs) information. LC's cataloging policy had been stringent in this area because of the expense of keying such data into records. Indeed, when BEAT decided it needed a benchmark against which to gauge the cost of its TOCs projects, the team experimented with the traditional method of typing the data and concluded that the cost of adding a typical TOC would be about forty dollars per record (in 1992 dollars).

## TOC studies

The theoretical foundations for concentrating on TOCs had been established by research conducted since the early 1980s. Pappas and Herendeen have reviewed the literature and shared their findings, reporting as follows:

- A study at the University of Toronto involving two thousand books revealed that twice as many relevant items for the social sciences and three times as many for those in the humanities were retrieved when users consulted a database that had been enhanced with TOCs.
- Another study found that TOCs added 15.5 unique subject-rich words per record when included in bibliographic descriptions.
- Yet another study of thirty-one publications on the history of taxation in Great Britain found more than six hundred terms in the TOCs to be content-indicative for an average of 19.5 per publication.
- An investigation conducted in 1990 at Carnegie Mellon University using both TOCs and abstracts revealed that contents enhancements increased the number of records retrieved by 20 to 30 percent.<sup>1</sup>

In 1998 Winkle found that 93 percent of a sample of 648 current English language books had TOCs with an average length of 67.75 words that could be included in catalog records. However, only 1.12 percent of the bibliographic records produced by LC at that time included contents notes.<sup>2</sup>

Pappas and Herendeen have also distilled the major advantages of enhancing bibliographic records with TOCs to introduce subject-indicative keywords that otherwise would be excluded from descriptions of publications. Of these, three advantages are considered to be especially compelling: (1) TOCs help users to determine the relevancy of particular titles to their informational needs—a service of value, especially in a closed-stack or remote-storage environment; (2) in an online environment, words in TOCs greatly improve search effectiveness, measured by the ability to identify and retrieve relevant items; (3) by providing content-indicative information, TOCs complement subject cataloging that strives

---

**John D. Byrum Jr.** (johnbyrum@earthlink.net) is Chief, Regional and Cooperative Cataloging, and **David W. Williamson** (dawi@loc.gov) is Cataloging Automation Specialist, Acquisitions and Bibliographic Access Directorate, Library of Congress.

to summarize the content of a work overall in a few carefully crafted access points per record.<sup>3</sup> Apropos to the latter point, according to an eleven-year longitudinal study cited by Yu and Young, “subject searching [is] being replaced by keyword searching.”<sup>4</sup> They reference another study recommending “that subject searchers should select keyword rather than subject headings as their first access strategy.”<sup>5</sup>

Contemporary investigations have confirmed the finding that books represented by bibliographic records with TOCs circulate more often than those with corresponding records that do not feature such data. For example, a recent case control study found that “the odds of a title being used increased by 45 percent if the titles had online tables of contents.”<sup>6</sup> The Cataloging Enrichment Initiative (RichCat), conceived of and coordinated by Kieft (Haverford College), is being established to encourage production of TOC data for older publications—particularly, those targeted for remote storage—so that catalog users can make informed decisions before recalling particular titles for their research.<sup>7</sup>

## ■ Providing TOC information

As a result of such considerations, one of BEAT’s earliest efforts to enhance bibliographic records focused on ways and means of providing TOC information.<sup>8</sup> The first application in this area centered on publications being processed through LC’s Electronic Cataloging-in-Publication (E-CIP) program. In this program, publishers electronically submit texts for cataloging prior to their publication so that the printed monographs will contain appropriate cataloging information about them. Currently, 55 percent of all publications submitted for Cataloging-in-Publication (CIP) are submitted as part of the E-CIP initiative. In fiscal year 2005 (ending September 30), a total of nearly thirty-five thousand digitally formatted galleys were received.

From 1993 to 1994, an application titled Text Capture and Electronic Conversion (TCEC) was written that enabled cataloging staff to include TOC data programmatically in the bibliographic records they were creating for publications submitted for E-CIP handling. Using the TCEC software and the ASCII-text electronic manuscripts submitted by the publishers, the cataloger highlights the TOC; next, the program manipulates it and adds the result into the bibliographic record’s MARC 505 field. TCEC formats the contents information to follow the *Anglo-American Cataloging Rules* specifications for recording TOCs. This includes deleting chapter, section, or part terms, and numbering; eliminating pagination, and adding International Standard Bibliographic Description (ISBD) punctuation. Because TCEC converts all words except the

first word in each chapter title to lowercase, the cataloger only needs to highlight any proper nouns that need to be capitalized. The resulting transfer of information from the manuscript to the record is accomplished instantaneously, and data are recorded as accurately as they appear in the electronic manuscript, thus obviating the need for detailed proofreading. Consequently, the former cataloging policy limitation that contents could be given only for monographs that are collections was lifted for E-CIP works. Catalogers are encouraged to apply the TCEC procedure as often as possible, following four criteria:

1. Does adding the chapter titles to the record provide improved natural language keyword searching?
2. Does adding the chapter titles to the record provide a greater understanding of the contents of the item than what is conveyed in the title and statement of responsibility area?
3. Will the TOC data require extensive manual editing to prepare the notes for machine manipulation?
4. If TOC is long and contains many entries, does this dilute the value of the information once it is put into a 505 field?

Fortunately, most staff can make quick decisions in answering these questions.

An informal study suggested that about half of the E-CIP publications would qualify for TCEC-TOC treatment, but catalogers do not always elect to apply this application to TOCs when they should. However, as staff has gradually become more comfortable working with this automated tool, the percentage of catalogers using it to produce contents notes has steadily risen. In fiscal year 2005, 13,627 E-CIPs received TOC treatment, a figure that represents 38 percent of all E-CIP materials received by LC.

In a second E-CIP-TOC project, BEAT members are creating a Web-based TOC record for nearly all E-CIP records that contain TOCs. These Web TOC records are created programmatically; a hot-link in the TOC field to and from the underlying record in the LC bibliographic database is made for every item. The program has been improved recently to include most diacritical marks and to add assigned LC subject headings to the Web versions of TOCs. By the end of fiscal year 2005, approximately sixty thousand E-CIP-TOC records had been added to the Web server.

## ■ Entry to the bibliographic record

The net result of these two E-CIP approaches is entry to the bibliographic record in the online catalog through keywords indexed in the TOC field as well as access from the Web, when search engines index the HTML version

of TOCs. As of July 11, 2005, a Yahoo! search on the phrase “contents for library of congress control” produced a result set numbering 242,000 entries, all linked to BEAT’s Web-based TOC records. A quick glance at some of the links reveals various uses of TOCs. Some links lead to records within the online catalogs of institutions that had downloaded TOCs. Others lead to such Web sites as “Ethical Schools of Thought,” “Mongabay.com,” “Hotel Marketing Associates,” “Solar sites,” “www.on-linenicaragua.com,” and many others that cite publications cataloged by LC.

## Digital tables of contents

In addition to developing a cost-effective method for enriching records for many of the important publications that are processed through the E-CIP program, BEAT has pursued two other approaches for making TOC information more widely available. The first is its Digital Tables of Contents (D-TOCs) project, which began in the late 1990s. This project has resulted in the creation of machine-readable TOC data derived from photocopied surrogates of TOCs taken from printed publications. By using scanning and optical character recognition (OCR) software as well as original programs written by BEAT’s automation staff, the scanned TOCs are subsequently HTML-encoded and placed on one of LC’s servers. The techniques used by the project have been modified recently to place heavier emphasis on use of imaging software and on adherence to a highly automated process to convert the TOC data to text format. The D-TOCs project has also implemented more automated and regularized quality control procedures to ensure that links work properly. In the process of HTML encoding, the underlying MARC catalog records are also automatically modified to include links to the TOC data, thus making linkage reciprocal between the two sources of information. Both the MARC catalog records and the linked TOC data may be viewed through a Web browser by accessing LC’s online catalog. In addition, the pervasive availability of Web indexing and search software also makes the D-TOCs records available from almost anywhere, providing access to LC’s Online Public Access Catalog (OPAC), even for the vast majority of users who are not aware of this project.

Thus, once the Web user has followed a D-TOC link back to LC’s catalog, LC can then make the wealth of its collections available for structured searching in items of related interest. As with BEAT’s other Web-based projects, D-TOCs serves to help bring Web users back to the library.

The following examples illustrate the various search paths and displays that might be encountered by a user in seeking information both on the Web and in LC’s OPAC.

As seen in figure 1, if a Web user searches Yahoo!, for example, using the phrase “animal communication networks,” because of interest in a work on this topic, the work by P. K. McGregor would appear near the top of the search results.

If the user clicks on the search result, he or she would be taken to the TOC for that book, partially illustrated in figure 2.

By clicking on “Bibliographic Record,” the searcher is taken to LC’s OPAC, where he or she will be shown a full description of the work for which the TOC is displayed. The display of the full record as opposed to one of the other possible views is governed by coding in the underlying link, thus providing the maximum amount of information available to the user immediately. Users searching the OPAC with the usual basic search form are initially presented with the Brief Record Display and must subsequently navigate to see more information. This step is eliminated by the link in the Web TOC display used to enter the OPAC (see figure 3).

This record provides hot links to other related works by authors, editors, or others represented by added entries through the related names link(s), as well as to other books on the same topic(s) through the subject link(s). In addition, the searcher can virtually browse the



Figure 1. Yahoo! partial search results for “animal communication networks”



Figure 2. HTML TOC record for “animal communication networks” (partial view)

LC shelf by using the call number link to see other books similarly classified, thereby providing entry to other resources of their interests.

Searching and retrieval are improved by various nontraditional techniques, including displaying words from the title and statement of responsibility fields of the bibliographic record, given at the beginning of the TOC display. Also, the keyword metadata tag in the TOC HTML file contains words from the subject heading fields of the bibliographic record, and the subject headings appear in the visible portion of the HTML record. This allows text-based searches on the file (as with a “find” capability resident in most Web browsing programs) while improving delivery of LC’s cataloger-supplied vocabulary terms for subject content.

Figure 4 illustrates the D-TOCs project from the vantage of a catalog user. A keyword search of LC’s OPAC for the terms “settlers wayne county” would produce the LC record with a hot link to the TOC.

Clicking on the hot link brings up a display of the TOC for the book (see figure 5).

## Selection of titles for the D-TOCs project

By the end of fiscal year 2005, more than thirty-one thousand titles had been selected for and processed through the D-TOCs project, and the figure is growing at a rate of 250 to 350 TOCs per week. Most of the publications included are drawn from LC’s current receipts, according to the following criteria: those selected should represent items of research value, including anthologies, biographies, and reference materials. In addition, the TOC should contain meaningful words and phrases and not exceed five pages in length. Titles selected are first searched in the database to eliminate those that already have been enriched as a result of other BEAT projects. To date, TOCs have been selected from English language publications. In 2005, however, coverage of the D-TOCs project was broadened to include books in German. In addition, those in Romance languages will soon be eligible. Also underway is implementation of a plan to create D-TOCs files in most of LC’s overseas offices, beginning in late 2005.

As an exception to its focus on current receipts, BEAT staff have experimented with retrospective publications acquired by staff of LC’s reference rooms. Upon their recommendation, the team began with genealogical works, specifically those in CS71 of the LC Classification schedules, intending to process them alphabetically by family name. (Interestingly, up to 70 percent of the titles in this collection do not have TOCs, possibly due to the fact that the majority of them are self-published.)

## ONIX-TOC

The newest, largest, and cheapest of BEAT’s three TOC projects is the ONline Information eXchange (ONIX)-TOC application, which was initiated in 2000. This undertaking involves extracting TOC data from publisher-supplied ONIX files. ONIX is an XML (extensible markup language) DTD (document type definition).<sup>8</sup> Publishers use this standardized format to provide book



Figure 3. Bibliographic record for “animal communication networks” (partial view)



Figure 4. Bibliographic record for “pioneer settlers of Wayne County, (West) Virginia”

DEDICATION .....	iv
PREFACE .....	vii
ACKNOWLEDGMENTS .....	xi
EXPLANATORY .....	xviii
1842 TAX LIST OF WAYNE COUNTY .....	1
1842 TAXPAYERS OF WAYNE COUNTY ILLUSTRATED .....	17
BALLENGER/BALLENGER FAMILY .....	39
BLOSS FAMILY .....	43
BROWNFIELD FAMILY .....	83
DEAN FAMILIES .....	103
Benjamin Dean Family .....	105
Isaac Dean Family .....	108
John Dean of Halifax County, Virginia .....	111
Gideon Dean Family .....	116
Elihu Dean Family .....	117
Joshua P. Dean Family .....	118
Benjamin & Elizabeth Dean Family Puzzle .....	120
GARRETT FAMILY .....	121
HALLETT FAMILY .....	149
HUTCHINSON/HUTCHINSON FAMILIES .....	181
Samuel Hutchinson Family .....	181
David Hutchinson .....	198
Alexander Hutchinson .....	200
VARIOUS SMITH FAMILIES OF WAYNE COUNTY .....	202

Figure 5. TOC for bibliographic record for “pioneer settlers of Wayne County, (West) Virginia” (partial view)

dealers and retailers with information about their publications; in turn, the retailers can reuse the information for promotional or other sales needs (e.g., creating Web-retailing screens). Because data used are supplied from commercial sources, BEAT's program adds the following disclaimer to each record processed on the basis of ONIX files: "Information from electronic data provided by the publisher. May be incomplete or contain other coding." In reality, such problems are quite rare.

The ONIX-TOC project is based on a Visual Basic program developed by cataloging automation specialist David Williamson, which scans ONIX files to create digital TOCs. The ONIX files are received regularly from publishers who want to make these data available to LC. The program does not validate the integrity of the ONIX file against DTD, but does sequentially seek out each ONIX record to begin processing the data in that record. Depending on the version of ONIX that was used in creating the file (as of June 2005, three versions of ONIX are being received by LC), the first element to be extracted is the ISBN for the book. This is usually the publisher's main identifier for the book. If there is no ISBN found (not yet assigned), the record is skipped and the program goes on to the next record in the file. If the ISBN is found, the ONIX record is searched for TOC information. There are three sets of tags that must be found (each tag has a mnemonic and alphanumeric equivalent):

1. `<TextTypeCode>` followed by a value of "04" for TOC and `</TextTypeCode>` to end the information;
2. `<TextFormat>` with a value usually indicating HTML markup or plain ASCII text followed by `</TextFormat>`;
3. And, then the actual `<Text>` tag that starts the TOC to be followed by the `</Text>` tag signaling the end of the TOC.

If all three sets are found, the data between the `<Text>` and `</Text>` tags are extracted. Next, the ISBN is searched against the LC database to see if there is a record for this book that also includes this ISBN.

Three problems can occur at this point:

1. The ISBN may not be unique. While ISBNs are supposed to be unique identifiers, the fact is that publishers sometimes reuse them (intentionally or not). An office outside the United States may apply for CIP for another edition being published outside the United States and may use the same ISBN as the one previously used for the U.S. edition. Tracking within the publisher's office(s) may get jumbled, and numbers may be reused. Another publisher may also accidentally put the wrong number on a publication.
2. If there are multiple records in the LC database, older versions of the program would link the

TOC to the record that was entered first into the LC bibliographic database. Until LC started to receive error reports for items with incorrect TOCs linked, the idea of a nonunique ISBN had not been considered. Subsequent investigation found that less than one percent of ONIX records presented this problem. The current version of the program will skip records with duplicate entries in the LC database as manual intervention entails too much time and expense.

3. The book may be represented in the LC database, but the record for it does not contain this ISBN. Publishers create separate ONIX records for each type of binding, for each edition, for each volume in a multivolume monograph, and for associated accompanying materials. If a paperback edition is released well after the hardback edition, and the hardback edition was published before ONIX was received by LC (or was published by a division of the publishing house that does not provide ONIX to LC), then the LC record probably will not have an ISBN for the paperback version. There is no way to equate the record for the paperback edition to the LC record for the hardback edition.

Assuming there is a match in the LC database, the MARC record is further processed, extracting out the LC Control Number (LCCN), the title field, and the LC subject headings. The title field and subjects are then cleaned up for use in the header or footer and the LCCN is added to the link connecting the TOC file to the LC OPAC record.

Because publishers tend to treat their TOC data the same throughout the file (either providing HTML or ASCII), the program is told what the publisher will do. The software will then either accept the HTML coding or, in the case of ASCII text, will wrap "`<pre>`" and "`</pre>`" tags around the text in addition to adding an HTML header and footer to the TOC information. Finally, after a spot check for quality assurance, the finished file is saved on the local machine for uploading to the LC Web server. The program then moves on to the next ONIX record, and the process is repeated until the end of the file is reached.

Each of these ONIX-TOC records offers the user an option to visit the bibliographic records in the LC online catalog for further information, following the pattern of the D-TOCs project described above. Similarly, the bibliographic records for these publications are programmatically enhanced by links in the 856 field to the ONIX-TOC files. Some of these records are further enhanced through the addition of book-jacket images (see figure 6).

The ONIX approach has proven to be the most economical in that most of the processing can be started and left to run unattended. Thus, from an actual cost

perspective, the ONIX approach has proved to be very inexpensive.

## Cost comparisons

The cost of adding a typical TOC is about \$40 per record (in 1992 dollars) for manual keying. BEAT's early initiatives with D-TOCs were much less expensive, about ten dollars per record for the scanning and linking. With better equipment and much more powerful OCR software—BEAT is able to take advantage of LC's use of Prime OCR for performing the conversion to text—the cost-per-record for D-TOCs has fallen to approximately \$2 per record. The E-CIP process where the TOC is inserted into the bibliographic record costs about \$3 per record, based on guidelines that the cataloger spend no more than five minutes trying to get the TOC into the record.

In comparison, ONIX data cost \$0.80 or less per record. The ONIX cost varies depending on the size of the data file received and how many new matches can be extracted from that file. The costs to set up the processing are about eight dollars (for an existing publisher) to ten dollars (for a new publisher) for each run that has to be performed. Once the program is running unattended, the number of successful new TOC files created determines the cost. If ten new TOC files are created, that's about \$0.80; if one hundred are created, the cost drops to \$0.08; and if one thousand or more are processed, the cost is less than one cent per TOC for accomplishing extraction and linking.

## Harvesting back files

The back files received with new sources of data usually give rise to a one-time harvest resulting in the creation of thousands of new TOC files. For example, when the firm of John Wiley and Sons sent its ONIX back file, 10,090 TOCs were extracted and linked. Wiley was the test case for ONIX; the software to process ONIX files was developed based on this back file, so the costs were a bit higher, \$0.26 per record for the 10,090 TOC files. However, once the basic software was developed, it was easily adapted for new publishers, and the per-unit cost has dropped dramatically. For example, when data started to come from the Cambridge University Press DataShop, the software was able to extract and link 12,975 TOC files for \$0.0008 per record. More recently received data from Cambridge has far fewer new TOC files available, but on average the cost is about \$0.016 per record.

Publishers' ONIX files vary in the amount of information they contain. The information is not aimed at library use but is intended for the book trade, so information



Figure 6. TOC for *Take It From Me*, together with image of the book jacket

about such matters as print runs, availability and pricing, distribution rights, and distributors can be found in the data for each record. While some records may only contain an ISBN, title, and a projected release date (almost an equivalent of a CIP prepublication ONIX record), others are richly loaded with data, including jacket blurbs, reviews, links to the author's Web site, links to cover images, and more.

The ONIX-TOC project is just one of four BEAT-ONIX projects. It was the first, but BEAT has expanded its ONIX projects to take advantage of publisher descriptions (141,000 to date), sample texts either in HTML or PDF (twenty-four thousand), and contributor biographical information for authors, editors, illustrators, collaborators, and so on (fifty-seven thousand). In addition, there is a small test involving forty-four reading-group guides linked from the LC record to the publisher's Web site.

LC currently receives three versions of ONIX: versions 1.1, 2.0, and 2.1. New iterations tend to come out rather frequently, and publishers are not willing to reprogram for each new version, so there are many publishers still using version 1.1. A few publishers have moved up to version 2, but more waited for version 2.1. They are just now beginning to distribute data using that version, even though it has been available since June 2003. All versions through 2.1 are upwardly compatible.

---

EDiTEUR, the group responsible for the ONIX standard, will release version 3.0 in late 2006.<sup>9</sup> This latest version is essentially the same as version 2.1, but all deprecated tags have been removed. Thus, there is no compatibility with the older versions, so programmers do not have to take into account any deprecated tags. Changes to the ONIX standard seem to be moving more to changes in code lists associated with the standard rather than changing the standard itself. This allows the standard to remain stable longer and requires less programming when there is a change, such as for a new type of contributor to a work or a new language code to be added. The change can be handled more efficiently in a code listing.

Today, there are nearly sixty thousand of BEAT's ONIX-TOC records available on the Web. This is a steadily expanding figure because the pool of publishers making their ONIX data available to LC continues to grow. Since February 1998, counters have monitored access to BEAT's Web-based TOC files. Hits currently range from four hundred to five hundred per hour between 8:00 A.M. and 9:00 P.M. eastern time to around two hundred per hour overnight; however, the rate is increasing rapidly. By October 2005, more than 7.5 million hits had been recorded. In addition, as with most BEAT products, the records enriched to provide links to these records are redistributed by LC's Cataloging Distribution Service (CDS), making them available to users of OCLC and RLIN, as well as to other agencies that subscribe to CDS products.

## TOC Web survey

To determine whether BEAT's TOCs were, in fact, serving useful purposes, a simple Web survey was developed and a small selection of the HTML-TOC files modified to provide a link to it.<sup>10</sup> The survey was posted from August through October 2001 and elicited input from 360 Web users. When asked how they found the TOC file, 60 percent reported "from the bibliographic record in a library catalog," while 36 percent said "from an Internet search." Of those who responded to the question "Was this TOC information useful?" 84 percent replied in the affirmative. When asked "Did you go to the bibliographic record from the link on the TOC page?" 58 percent answered "yes." Of these, 57 percent indicated that they had "look[ed] over" the bibliographic record, and some had also clicked on the hot links within the bibliographic record to search for works by the same authors or on related subjects. Asked to describe themselves, 58 percent indicated that they were researchers or students, 23 percent were librarians, and 14 percent were casual users looking for information. Survey participants had an opportunity to

comment on their Web TOC experience. Their opinions confirmed and extended the various uses and overall serviceability of TOC information summarized by Pappas and Herendeen.<sup>11</sup> Librarians found BEAT's TOCs helpful in making acquisition decisions, especially in cases of expensive publications, and in downloading the TOC for addition to records in their OPACs. The survey was repeated about eighteen months later with almost identical results.

## 505 data

Although LC's D-TOCs and ONIX-TOCs records provide access to the online catalog, they do not provide entry to bibliographic records from within the catalog, because the bibliographic records contain only links in lieu of the TOCs themselves. To counter this drawback, a program was created to add full TOCs to the bibliographic records for the Web-based TOCs. Beginning February 2005, it proved possible to use this application to enrich bibliographic records by adding information that was previously only available through 856 links.

The 505 data are automatically generated from the TOC information in the files created for the D-TOCs and ONIX-TOC records. The program scans the TOC file and extracts out each line of the TOC data, treating each line as an element in the 505 field being constructed. For many TOCs, this works perfectly well to extract the chapter titles. In the case of multiline TOC titles, this approach causes a TOC title to become two or more elements in the 505 field, potentially causing confusion. Similarly, when multiple chapter titles are on one line, some muddling of the data will occur. Each application of the program will introduce the TOC with the legend: "Machine-generated contents note."<sup>12</sup> Because the scanned TOCs come in a wide variety of formats and structures, some errors are to be expected in the placement and configuration of the 505 textual strings. Space, hyphen, hyphen, space will be inserted after each line break within the TOCs. In many cases, chapter and page numbers will appear as captured from the scanned TOCs images. The 505 data will not undergo review for punctuation (see figure 7).

Approximately sixty thousand LC records with existing 856 links to TOC texts are being batch-processed, modified, and redistributed until all eligible records are enhanced. Initially, after consultation with LC's public service staff, TOCs that are four thousand or less bytes in extent have been declared eligible, but larger-sized records may become eligible for processing as described above. Later, eligible ONIX-TOC records may be similarly processed.



Figure 7. Sample bibliographic record with machine-generated TOC

## Conclusion

BEAT's TOC projects demonstrate how, in the electronic era, LC is taking traditional services and providing new and improved ways to capitalize on them in the digital age. These projects provide a model that might be of interest to others as they ponder issues and opportunities regarding bibliographic access and retrieval in today's growing electronic environment. By responding to expanding user needs through bibliographic enrichment initiatives such as TOCs, libraries will recognize that, whether in a traditional framework or in the digital environment, researchers can and do use the catalog the way an entire library is used—not only as a source of material and information, but also as a gateway to additional information. Through adding more keyword-rich information to the catalog, libraries can serve the extended information needs of the researcher as well as offer structured pathways to their own information resources. Offering such features as standardized subject

terminology and pervasive controlled headings, these catalogs are the result of more than one hundred years of intellectual effort and real capital. Considering the major investments made to create and maintain their catalogs, libraries everywhere should seek opportunities to build upon these investments to provide richer records in order to entice patrons to continue to include the online catalog as a rewarding access mechanism in their growing array of tools for information retrieval.

## References and notes

1. 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 and Classification Quarterly* 23, no. 4 (2000): 65–67.
2. R. Conrad Winkle, "An Analysis of Tables of Contents in Recent English-Language Books," *Library Resources and Technical Services* 43, no. 1 (1998): 14.
3. Pappas and Herendeen, "Enhancing Bibliographic Records," 63–64.
4. Holly Yu and Margo Young, "The Impact of Web Search Engines on Subject Searching in OPAC," *Information Technology and Libraries* (Dec. 2004): 168.
5. *Ibid.*, 169.
6. Ruth C. Morris, "Online Tables of Contents for Books: Effect on Usage," *Bulletin of the Medical Library Association* 89, no. 1 (Jan. 2001): 29.
7. More information regarding RichCat is available at [www.loc.gov/standards/catenrich](http://www.loc.gov/standards/catenrich) (accessed Dec. 21, 2005).
8. A streaming videocast recorded in January 2002, containing information relating to all of BEAT's TOC initiatives as of that date, may be viewed online at <http://lcweb.loc.gov/catdir/beat/eTOC/jan30-eTOC.html> (accessed Dec. 21, 2005).
9. For more information on ONIX, visit the EDItEUR home page, [www.editeur.org](http://www.editeur.org) (accessed Dec. 21, 2005). EDItEUR is the agency responsible for coordinating the various national ONIX groups and distributing the ONIX standard.
10. Survey results and comments are available for review at [www.loc.gov/catdir/tocsurveyresults.html](http://www.loc.gov/catdir/tocsurveyresults.html) (accessed Dec. 21, 2005).
11. Pappas and Herendeen, "Enhancing Bibliographic Records."
12. The 505 indicators for these machine-generated notes will be set to "8" (No display constant generated) and blank (Basic; single occurrence of subfield \$a).

## Index to Advertisers

ALCTS	32	EBSCO	48
Art & Tech	23	LAMA	cover 3
ASCLA	47	LITA	covers 2 and 4