

## BEAT Annual Report 2002

This year marked the 10<sup>th</sup> anniversary of the BEAT initiative, and the team celebrated with a half-day program on November 12, 2002 in the Library's Mumford Room. The schedule included opening remarks by John D. Byrum, Jr., Chief, Regional and Cooperative Cataloging Division, and BEAT chair, who has directed the team since its inception, and the program encompassed a half-day series of presentations on current and ongoing BEAT projects, as well as concluding remarks by Beacher Wiggins, Acting Associate Librarian for Library Services and by many of the Directors whose staff also support the BEAT Team. The closing activities provided team members and library staff in attendance an opportunity to gather informally.

In his remarks John Byrum reviewed early BEAT projects and highlighted just some of the principles that underlie the team's activities: working "Outside the Box," developing tools that aid in creating and in locating information, developing innovative work flows and policies, preparing pilot projects ("Proof of Concept"), and most recently, scaling up to production level some of the team's projects. A special point to note was that the celebration was also intended to recognize all the catalogers who have created the products BEAT is enriching and to appreciate public services and other staff for sharing their creativity and expertise in forming and implementing BEAT projects.

With a great deal that could have been said regarding past achievements, the anniversary program chose to highlight six of the current BEAT projects, and this report reviews and updates the current status those projects as well as describes the other noteworthy BEAT activities during the past year.

In short compass, and echoing some of Mr. Byrum's anniversary comments, the Bibliographic Enrichment Advisory Team (BEAT) is a Cataloging Directorate initiative aimed at developing tools to aid catalogers, reference specialists, and searchers in creating and locating information. Major components of the team's work are enriching the content of Library of Congress bibliographic records, improving access to the data the records contain, and conducting research and development in areas that can contribute to furthering these efforts. The team's membership represents a wide spectrum of the Library's functions reflecting the Cataloging Directorate's desire to provide benefits from its projects to as wide an audience as possible and to incorporate within its program objectives the needs and interests of various constituencies beyond those of technical services

### **BEAT Projects**

#### ***Tables Of Contents Projects***

##### *DTOC*

The Digital Tables of Contents project creates machine readable Table of Contents (TOC) data from TOC surrogates and these materials are subsequently HTML-encoded and placed on a server at the Library. The process cross-links the TOC to underlying catalog records. Both the catalog records themselves and the linked TOC data may be viewed through a Web browser by accessing the Library's online catalog access options. Over 11,000 TOCs have been created and linked, in this project.

During the year the team acquired new scanning equipment as well as updated its optical character recognition (OCR) software, all of which were brought on stream and into production during the first half of the year. The team revised its workflow and data handling so that much of the processing after initial digitization runs remotely directly from various cataloging workstations. The project also diffused its selection responsibilities with the goal of both increasing input (and consequently production) and of continuing to broaden the subject areas of the works represented in the data.

### *ECIP TOC*

One area of BEAT support has been for the Library's Electronic Cataloging-In-Publication (E-CIP) program, where the team has been instrumental in helping to implement a Tables of Contents component. The addition of TOC data to E-CIP records converges with BEAT's bibliographic enrichment mandate.

More than 3500 electronic CIP records were created and it is expected that the number of TOC records associated with these CIP records will continue to increase as the E-CIP program will continue to expand during the next few years. During Fiscal Year 2000, TOC data was included in the MARC records for approximately 21% of the titles processed as E-CIPs, and with the strong support of the Director for Cataloging and the use of automated editing techniques developed by team member David Williamson, and the most recent statistics show staff now adding TOC data for about 32% of the E-CIP titles. This activity is ongoing and separate from BEAT's own digital TOC initiative described elsewhere in this report.

E-CIP TOC operates in two modes of production. In the first, table of contents data is added to the body of the catalog record by catalogers and staff during the course of the regular E-CIP cataloging process.

In the second mode, still in the experimental stage, TOC data is created from the original source files, and using programming techniques similar to those outlined in the description of ONIX TOC (*below*), resulting HTML files are created and placed on the Web. Cross links between the catalog record and the TOC files are also made. The controlled vocabulary as used in subject assignment is added to the HTML keyword meta tags, providing additional search points to the material in the TOC record.

### *ONIX TOC*

ONIX (ONline Information eXchange) is a means of representing book industry product information and is being used by some publishers today to communicate that data electronically.

The Library receives this data directly, including coverage of retrospective material. Programming developed by a BEAT team member utilizes the data to create Table of Contents (TOC) records that the Library makes available on the Web. Hyper-links are made from this TOC data to the catalog record, and the reverse, thus allowing researchers to move from or to the Library's online catalog where they can make additional searches for related or other material. A catalog record with a related TOC file retrieved in an online search has the hot-link to the TOC displayed on screen. As in the Digital TOC project described above, Library of Congress Subject Headings from the catalog record are added to the HTML keyword meta tag for future use and also at the bottom of the

displayed TOC so that search engines can pick up and index these terms. To date the project has created about 31,000 ONIX TOC records.

More than one million hits were recorded on the TOC files section of the Cataloging Directorate Web pages, continuing to reflect the utility of the TOC files created by all of the BEAT TOC initiatives. In addition, TOC files are now being indexed by the Google search engine and it is anticipated that this will increase further the awareness of the enriched data provided by the TOC projects and will also offer increased opportunities for researchers to explore the riches of the Library's collection through the links that the TOC data provide to the LC Catalog and the catalog's search options.

A cybercast from January 2002 prepared as part of the LC staff Digital-Future-and-You series, containing information relating to all the TOC initiatives may be viewed online at <http://lcweb.loc.gov/catdir/beat/eTOC/jan30-eTOC.html>

All of the TOC projects benefited from the implementation of the batch cataloging utility this year. This capability greatly enhances the speed and efficiency of adding linking fields to catalog records and has helped streamline the processing of TOC materials.

### ***OTHER ONIX INITIATIVES***

#### *Publisher's Descriptions*

An outgrowth of the ONIX TOC initiative is the creation of records that contain Publisher's Descriptions of books. Based on ONIX encoded materials, file creation and linking is similar to that of the ONIX TOC initiative above. To date, the project has created approximately 27,500 such records, although links are currently made from the catalog record only in an ongoing fashion. It is anticipated that introduction of additional system capabilities within the near future will permit the timely creation of such links and in increasing volume.

By end of 2002, the Library was receiving ONIX data for TOC and/or descriptions from the following distributors, who represent many publishers and/or subsidiaries: Cambridge University Press, Eloquence Services, Houghton Mifflin, McGraw-Hill, John Wiley, and Princeton.

A fuller list of distributors and publishers who have supplied or who are contributing ONIX data was posted on the BEAT web page and is updated regularly.

#### *ONIX Sample texts*

This project makes links from LC catalog records to copies of sample texts from publications received from various participating publishers. This material might include such things as sample chapters, front-matter, illustrations, or other materials - - information that readers or prospective users of a book would find of interest. With the publisher's permission, these have been stored at the Library to insure long-term availability. By linking the catalog record to these items, the project significantly enhances the information about a book that the Library makes available to a researcher. The project added links to 2,278 titles from Cambridge University Press.

#### *ONIX Reading Group Guides*

The Reading Group Guide project, represents another outgrowth of BEAT's use of ONIX data and resulted in linking of forty-four catalog records for Houghton Mifflin publications to their associated reading group guides on the Web. This experimental BEAT project provides links to these discussion guidelines or topics for book groups for which the information is not on LC's Web server as it is with the Table of Contents and Publishers Description projects. Therefore, the bibliographic records in this project will be closely monitored by BEAT to determine the reliability of the URLs involved.

### ***BECites+***

BECITES+ (Bibliographies plus: Enhanced Citations with Indexes, Tables of contents, Electronic resources and Sources cited) enhances online bibliographies produced by LC staff, as well as the catalog records for the titles included in such guides, through the use of scanning and OCR technology. First and foremost, the project concentrates on providing links to scanned and OCR'd tables of contents, indexes, and back-of-book bibliographies of resources cited in the guides. A secondary focus begun in 2002 has been the scanning and OCR'ing of the full text of heavily used but out-of-print guides prepared by LC staff to Library collections, including those containing individual items not easily identified in the Library's online catalog.

Completed works within this project this year include guides on business history, African American business, Thomas Jefferson, and materials on Immigrant Arrivals to the United States. A number of additional works are in progress, covering additional business resources, guides to microfilm collections, prints and photographs resources, and manuscripts from several Middle Eastern monasteries, as well as a guide to Ladino publications in the Library of Congress.

Information concerning the titles in the project are available online, and further details about the project as well as a full list of completed bibliographies and other work in progress can be found at <http://www.loc.gov/rr/business/guide/>

### ***Web Access to Publications in Series***

One of the most noteworthy developments this year was in the dramatic growth and awakening awareness of the significance of this project. Many social science series of the "working paper/discussion paper" type are now available in electronic form. The focus of this project is to identify those that are available in electronic form and add the URLs to the LC records for these series. By linking to these electronic versions, LC provides more timely, comprehensive, and cost effective access to these series. A corollary to this project has been the identification of numerous high research value monographic series that have not been represented in LC's catalogs. These projects create added value by leading researchers via these links to significant information resources that offer full-text and/or abstracts for many of the titles in these series. These efforts provided access to the full, electronic texts of more than 15,000 individual monographs. At year-end more than 120 different serial publications, primarily in business and economics had been added to the project database, and these represent some of the most important publications from a wide range of institutions from throughout the world.

As a further enhancement, a pilot project was launched to create electronic resource records for the individual monographs of selected series. A number of links to examples and further description of this project can be found on the BEAT Web page.

### ***Web Access to Works in the Public Domain***

BEAT announced a new initiative to link LC bibliographic records to full text electronic copies of these same cataloged materials residing in collections of other institutions. Though these works, all in the public domain, have been digitized by various institutions, many of the original printed works are also in the collections of the Library of Congress. By linking LC catalog records to these electronic versions the Library expects to provide users with more unified and centralized access to materials of this nature as well as provide users of the LC collections or of LC catalog data rich and substantive information about the contents of these works as well as access to their texts.

In this project, the LC catalog record is modified to add an 007 field indicating a computer resource, to add a note for the existence of a digital reproduction (in a 530 field presenting information that the item is available in a different physical form), and to create an 856 linking field that points to a meta-level description record at the holding institution. In turn, that meta record provides links that allow the researcher to browse the full text itself. These operations are accomplished as a result of a program written by David Williamson, cataloging automation specialist. Some but not all of these modified LC catalog records are redistributed, depending on whether access points and other details are in conformance with AACR2.

The initial applications result from cooperative agreements with the University of Michigan (for materials digitized in its Making of America project, described at <http://moa.umdl.umich.edu/> and Indiana University (works comprising its Wright American Fiction, 1851-1875 project), described at <http://www.letts.indiana.edu/web/w/wright2/>. In order for records to qualify for enhancement in these projects, the electronic versions have to be an exact version of a print version represented in the Library's collection, as established by the presence of an Library of Congress Control Number (LCCN) for the electronic version that matches the LCCN for the print version. For the University of Michigan materials, it proved possible to enhance 1,267 LC bibliographic records, and for the Indiana University project 653 LC bibliographic records were linked to the online versions.

The project is actively seeking additional participants and hopes to expand next year.

### ***HLAS Reviews***

A new activity is the HLAS Reviews project. A separate database of records created for the *Handbook of Latin American Studies* (HLAS) is available for searching through the Library's integrated library system and includes reviews for most of the titles listed in this serial. These summaries are especially rich for natural language keyword searching. Except for a few in Spanish, the reviews are in English.

The purpose of the BEAT project is to extract reviews from the HLAS database for those monographs that can be matched to works in the LC collections, and insert them into the

corresponding records in the LC bibliographic database. This project began with reviews listed in the two most recent printed volumes of HLAS, v. 57 on social sciences and v. 58 on humanities. The initial application yielded enhancement of 5,995 LC bibliographic records. The Library of Congress Cataloging Distribution Service is redistributing these enriched records

### ***Adding Annotations and Resource Access***

A BEAT project to add descriptive annotations to the catalog record for many of the Library's Microform collections was undertaken at the end of the year. LC reference staff have created a web-based annotated bibliography, *A Guide to the Microform Collections in the Humanities and Social Sciences Division*, describing many of the Library's microform collections. The *Guide*, based on previous iterations of text-based bibliographies, adds many more collections received from 1996 through the present, as well as several collections that were previously undiscovered. It lists its entries alphabetically by collection, and each entry has an annotation describing the collection as well as information about individual guides to each of the collections themselves. However, the *bibliographic* records for these collections do not carry the same extremely descriptive data as is found in the online *Guide*.

Accordingly, this project will add the text of all the annotations to the underlying LC catalog record for those collections that have been assigned a Library of Congress Catalog Card Number (the LCCN). This will result in the record carrying a much fuller description of the collection identified in the catalog record, and should be very useful in helping researchers who find the entry assess and utilize that collection. BEAT is hopeful that this project will lead to similar initiatives aimed at further enriching the content of, or access to, other LC collections.

### ***DCC/LCC Correlations***

The creation or promotion of use of new or better tools for catalogers has always been a major focus of the BEAT Team. In keeping with that key objective, the team supported the expansion and the use of the capabilities of a program (MINARET) by which LC Classification is made available to LC staff catalogers online that extend to the provision of classification correlations in bibliographic records between Decimal (DDC) and LC Classification (LCC) numbers. Using the Library's own database of catalog records, the goal is to implement use of this MINARET capability for real-time search and retrieval of classification number correlations between DDC and LCC and the reverse for use in appropriate areas of LC catalog work.

Several demonstrations of the utility of this capability were made to staff.

### ***Other Areas of BEAT activity***

In addition to the projects noted above BEAT also participates in or supports a number of other initiatives. Among those are Area Studies Portals, support for the development and rollout of NewBooks, and support for implementation of BEAT-related recommendations from the Cataloging Bicentennial Conference. In addition, BEAT continues to receive inquiries regarding its projects, particularly those relating to TOC, from outside librarians. Project chairs or other staff provide direct replies and correspond with prospective partners, and team members maintain professional

contact with other institutions regarding projects of mutual interest.

### ***Outreach***

During the year BEAT provided the Library's Library Services Directors with a lively presentation of some of BEAT's projects, sponsored and participated in two demonstrations of the use of the correlations features of Minaret, and presented the demonstration and subsequent cybercast of the TOC initiatives in the LC "Digital Futures and You series," referenced under the TOC projects section above. The team also presented demonstrations of its projects at national and other conferences and its members are also active in various library-related professional organizations. The Team received feature coverage of the anniversary, and of the featured BEAT projects in the *LC Gazette*, and team members contributed four articles or updates to the *LCCN* publication.

The BEAT Web pages were also substantially expanded this year, with a number of new graphic and live links to resources added to sub-sections dealing with the team's many projects. Of particular note are the materials related to the various ONIX initiatives and for the Web Access to Works in the Public Domain and the Web access to Works in Series projects, all of which have links from the BEAT page at <http://www.loc.gov/catdir/beat/beat.html>

The team generally held biweekly meetings throughout the year, and the minutes of the meetings or news from BEAT was circulated widely throughout the Library. BEAT's current and retrospective reports, such as those for ALA, are available on the Web site, and the team continues to add documentation, links, and other materials that are illustrative of its projects.

### ***Staffing***

BEAT would like to acknowledge the long-term and sustained contributions and support of Peter Young and Winston Tabb, both of whom left the Library this year.