

Library Resources & Technical Services

ISSN 0024-2527

July 2002

Volume 46, No. 3

ARTICLES

- Panning for Gold** 79
Utility of the World Wide Web for Metadata and
Authority Control in Special Collections
Nadine P. Ellero
- “Garbage” In, “Refuse and Refuse Disposal” Out** 92
Making the Most of the Subject Authority File in OPAC
Marguerite E. Horn
- Cooperative Cataloging, Vendor Records, and
European Language Monographs** 105
Charlene Kellsey

American Library
Association
Association for Library
Collections & Technical
Services
Visit *LRTS* online at
www.ala.org/alcts/lrts.
For current news and
reports on ALCTS activi-
ties, see the ALCTS
Newsletter Online at
[www.ala.org/
alcts/alcts_news](http://www.ala.org/alcts/alcts_news).

EDITORIAL BOARD

Editor and Chair

John Budd

University of Missouri

Members

Kathleen R. Brown, *North Carolina State University*

Lynn S. Connaway, *NetLibrary Incorporated*

Jeanne M. Drewes, *Michigan State University*

Martin M. Kurth, *Cornell University*

Jack Montgomery, *Western Kentucky University*

Linda L. Phillips, *University of Tennessee*

Richard E. Sapon-White, *Oregon State University*

Stanley J. Wilder, *University of Rochester*

Barbara A. Winters, *University of Georgia*

Ex-Officio Members

Charles Wilt

Executive Director, ALCTS

Miriam Palm, *Stanford University, emerita*

Editor, ALCTS Newsletter Online

Douglas Raber, *University of Tennessee*

Book Review Editor, *LRTS*

Library Resources & Technical Services (ISSN 0024-2527) is published quarterly by the American Library Association, 50 E. Huron St., Chicago, IL 60611. It is the official publication of the Association for Library Collections & Technical Services, a division of the American Library Association. Subscription Price: to members of the Association for Library Collections & Technical Services, \$27.50 per year, included in the membership dues; to nonmembers, \$55 per year in U.S., Canada, and Mexico, and \$65 per year in other foreign countries. Single copies, \$15. Periodical postage paid at Chicago, IL, and at additional mailing offices. **POSTMASTER:** Send address changes to *Library Resources & Technical Services*, 50 E. Huron St., Chicago, IL 60611. Business Manager: Charles Wilt, Executive Director, Association for Library Collections & Technical Services, a division of the American Library Association. Send manuscripts to the Editorial Office: John Budd, Editor, *Library Resources & Technical Services*, University of Missouri, 221M Townsend Hall, Columbia, MO 65211; (573) 882-3258; fax: (573) 884-4944; e-mail: buddj@missouri.edu. Advertising: Bill Coffee, c/o Benson, Coffee & Associates, 1411 Peterson Ave., Park Ridge, IL 60068; (847) 692-4695; fax: (847) 692-3877. ALA Production Services: Troy D. Linker, Kevin Heubusch; Ellie Barta-Moran, Angela Hanshaw, Kristen McKulski, and Karen Sheets. Members: Address changes and inquiries should be sent to Membership Department—*Library Resources & Technical Services*, 50 E. Huron St., Chicago, IL 60611. Nonmember subscribers: Subscriptions, orders, changes of address, and inquiries should be sent to *Library Resources & Technical Services*, Subscription Department, American Library Association, 50 E. Huron St., Chicago, IL 60611; 800-545-2433; fax: (312) 944-2641; subscription@ala.org.

Library Resources & Technical Services is indexed in *Library Literature*, *Library & Information Science Abstracts*, *Current Index to Journals in Education*, *Science Citation Index*, and *Information Science Abstracts*. Contents are listed in CALL (Current American—Library Literature). Its reviews are included in *Book Review Digest*, *Book Review Index*, and *Review of Reviews*.

Instructions for authors appear on p. 220–21 of the October 2001 issue and on the LRTS Web page at www.ala.org/alcts/lrts. Copies of books for review should be addressed to Douglas Raber, Book Review Editor, *Library Resources & Technical Services*, School of Information Science, University of Tennessee, 804 Volunteer Blvd., Knoxville, TN 37996-4330 (raber@utk.edu).

© American Library Association 2002

All materials in this journal subject to copyright by the American Library Association may be photocopied for the noncommercial purpose of scientific or educational advancement granted by Sections 107 and 108 of the Copyright Revision Act of 1976. For other reprinting, photocopying, or translating, address requests to the ALA Office of Rights and Permissions, 50 E. Huron St., Chicago, IL 60611.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992. ∞

Publication in *Library Resources & Technical Services* does not imply official endorsement by the Association for Library Collections & Technical Services nor by ALA, and the assumption of editorial responsibility is not to be construed as endorsement of the opinions expressed by the editor or individual contributors.

Panning for Gold

Utility of the World Wide Web for Metadata and Authority Control in Special Collections

Nadine P. Ellero

This article describes the use of the World Wide Web as a valuable name authority resource and tool for special collections analytic-level cataloging and the specific goal of “fully discovering” the names of people who lived in the past as well as those from the present. Current tools and initiatives such as the Name Authority Component of the Program for Cooperative Cataloging (NACO) and the Library of Congress Name Authority File have a specific mission and are partially helpful. Web resources encompassing special collections are often intricate and require global and enhanced resources to continue what have been the guiding principles, tradition, and value of cataloging: to discover works via many points of entry; to find works by or about the same person, topic, or title; and to continue the great cataloging legacies of standards and cooperation.

From August 2000 through January 2002, the Historical Collections and Services Department of The Claude Moore Health Sciences Library (TCMHSL), along with the guidance of the Head of Intellectual Access at TCMHSL, used the Web as a tool for name authority work with the digitization of The Philip S. Hench Walter Reed Yellow Fever Collection. Specifically, authority data for names of persons were created and supplemented with information obtained via resources found on the Web.

Russell and Spillane (2001) presented a synthesis of the function of authority control in general cataloging practice and the utility of the Web in obtaining information on authors. They focused on the benefits of the NACO initiative and presented a case for using the Web for contact information for authors. This Web information would supplement the authority record in the 670 field. Author and company Web sites and catalogs of the world's national libraries are two examples of sources. In addition, they discussed the evaluation of this contact information mined from the Web. Web sites created by the author or institution are considered as authoritative as corresponding print reference sources, whereas Web pages or sites created by “fans” need to “be treated with a healthy degree of skepticism” (78).

While Russell and Spillane concentrated on contemporary authors, this article will show how the use of the Web, in a very focused and specialized way, can be extremely valuable for special collections cataloging and metadata creation for fully discovering the names of people who lived in the past as well as those from the present.

While the NACO initiative and the Library of Congress Name Authority File (commonly known as the NAF) are unsurpassed, they were designed to include

Nadine P. Ellero (npe6f@virginia.edu) is Head of Intellectual Access, The Claude Moore Health Sciences Library, University of Virginia Health System in Charlottesville.

Manuscript received April 1, 2002; accepted May 17, 2002.

names based on items of “literary warrant” as they are added to library collections and therefore serve a very specific purpose. Our project, while involving both famous and little known people derived from analyzing a manuscript collection, fell outside the mission of the NAF. As more and more institutions (i.e., libraries, archives, and museums) in the United States and around the world process special collections of unpublished materials on an analytic level and make these resources available on the Web, an enhanced and global system for authority records will become essential. It is interesting to note that in almost ten years the NAF has increased an astonishing 667%. In 1992, the NAF contained 500,000 name authority records (*Library of Congress Information Bulletin* 1992); as of April 13, 2002, the total was 3,835,384 (Sturtevant 2002). Additionally, in 1994 the Library of Congress Name Authority File became the Anglo-American Authority File (*Library of Congress Information Bulletin* 1992) and in 2001 19.43% of the new Name Authority Records (NARs) were contributed by international sites.¹ We were surprised to find that even with this amazing expansion of the NAF, many names of prominent persons were not found, including United States Congressional Gold Medal honoree Aristides Agramonte, Surgeon General of the United States Army Raymond W. Bliss, President of the National Academy of Medicine in Columbia Roberto Franco, and U.S. Secretary of War James W. Good. An enhancement to the NAF for facilitating faster personal name identification would be to include a qualifier field indicating the person’s profession (e.g., actor, author, historian, lawyer, physician, etc.). While there is a 678 biographical or historical data field in the MARC option of data fields for authority records, it was often not present from our experience. In a sampling of 264 NAF records, 5 contained a 678 field, 141 did not contain a 678 field or other field to indicate profession, and 108 contained a 670 field that indicated profession. The balance of NAF records that contained fields indicating profession were 8 with 510 fields and 2 with 400 fields. In addition, projects such as this one could report to NACO on data obtained for death dates of persons. Some examples from this project include: Edgar Erskine Hume, Foster Kennedy, and William Dosite Postell.² The Web implies an international scope; therefore international and enhanced resources will be required to continue what have been the guiding principles, tradition, and value of cataloging: to discover “works” via many points of entry; to find works by or about the same person, topic, or title (such as all the versions of the Bible); and to continue the great cataloging legacies of standards and cooperation.

Background

The Philip S. Hench Walter Reed Yellow Fever Collection, held by Historical Collections and Services of TCMHSL, is

an archive of largely primary source documents focusing on the discovery of the transmission of yellow fever. Walter Reed and his assistants James Carroll, Aristides Agramonte, and Jesse Lazear proved that the *Aedes aegypti* mosquito was the transmission vehicle for the yellow fever virus. Philip S. Hench’s passion for this subject led him to collect books, articles, correspondence, photographs, and artifacts from the Yellow Fever Commission of 1900. The archive also includes military artifacts and photographs of American troops during the Spanish-American War and the First American Occupation of Cuba (1898–1902), and is therefore international in scope. (For more information, see the Web exhibit *This Most Dreadful Pest of Humanity: Yellow Fever and the Reed Commission, 1898–1901* at www.med.virginia.edu/hs-library/historical/yelfev/tabcon.html and the Philip S. Hench Walter Reed Yellow Fever Collection Web site at: <http://yellowfever.lib.virginia.edu/>.)

TCMHSL’s curator, Joan Echtenkamp Klein, realizing the value and importance of digitizing and preserving this collection, assembled several library staff to complete an application for the Institute of Museum and Library Services (IMLS) National Leadership Grant in December 1998. On September 24, 1999, the library received notice of an award of \$250,041. The goal of this project was to digitize a large selection of the primary resources (i.e., the written correspondence, photographs, artifacts, and maps) both as images and searchable text to create a Web site for displaying, searching, and learning about the Yellow Fever Commission and discovery of the transmission of yellow fever. The grant writing and work on the project itself was a successful experience of cross-departmental cooperation requiring expertise from all areas of library operations, especially cataloging for the metadata aspects. The alignment of technical services and special collections in this creation of digital objects and Web sites will undoubtedly be a continuing trend (Crosby 2000; Bradshaw and Wagner 2000). The work at TCMHSL on building the Yellow Fever Web site is a testimonial to the richness of cataloger and curator collaboration and communication.

Primary Documents and “Metadata”

The project team along with David Seaman, director of the Electronic Text Center at the University of Virginia, decided to encode the text using the Text Encoding Initiative (TEI) Guidelines, which are a subset of the Extensible Mark-up Language (XML). Robust flexibility was desired for searching as well as the ability to create pre-determined searches by subject, personal name, etc. XML was also a promising archival choice and one that would allow for future applications as yet unknown. Seaman designed a template for the insertion of tagged text and the

project team designed a data grid to capture the essential descriptive metadata for senders (i.e., a sender of a letter), recipients (i.e., a receiver of a letter), broad subjects, and significant geographic names and places.

The popular definition of metadata as “data about data” is pervasive and poor. Hearn (1999, 7) described metadata structures as “the development of metadata [that] can be thought of as a loosely consolidated effort to create a standard structure for differing communities to use for the description and retrieval of records.” Glogoff and Forger (2001, 9) used metadata “as the indexing that is applied to electronic information.” In other words, there is the structure (i.e., XML language or MARC) and there are the standards in the form of content (i.e., the Library of Congress Name Authority File and the prescribed form of entry and standard conventions set forth in the Anglo-American Cataloguing Rules (AACR2) for the creation of entries. The goal of metadata, as stated by Milstead and Feldman (1999, 25), is to “improve matching by standardizing the structure and content of indexing or cataloging information.” Vellucci makes several important observations that we aimed for in our metadata approach and design:

. . . the successful use of multiple metadata schemes in the library environment will depend on authority control . . . (2000, 33)

Metadata are data that describe the attributes of a resource; characterize its relationships; support its discovery, management, and effective use; and exist in an electronic environment. (34)

For purposes of this article, the phrase “descriptive metadata” will be used. Descriptive metadata include types of data that, by their nature, are of an intellectual content and include such elements as author (in the Yellow Fever Collection, mostly known as senders of letters), recipient (the receiver of a letter), personal name as subject, topical subjects, geographic locations, etc. These descriptive metadata (as the entry in a 100 or 600 MARC field, for example) require consistency and completeness. The consistency aspect serves to control input of the data in a tag or field and better guarantees resource discovery. Completeness of name-entry—that is, having both a surname and forename whenever possible—also facilitates resource and information discovery and retrieval. In the language of catalogers, this is authority control. In addition, anyone familiar with working with primary source documents knows that names of people and places are often incomplete, have variations in spelling, and frequently contain nicknames. These unique materials, especially the letters, require the creation of titles for XML headers, and dates often have to be supplied or

approximated. One of the most challenging aspects of these materials is the high degree of personal name variation, which demands authority control for insuring accurate transcription and effective discovery of the descriptive or intellectual content contained in these primary resources.

At the time of this project, a full authority system—that is, one allowing for automatic mapping of variant names to a complete or controlled name—was not developed for this Web site. A collection level MARC record was added to the library’s online catalog, but not the 5,562 individual items that comprise the Web site collection. Likewise, the authority controlled personal names from the Web site were not added to the library’s online catalog authority file. The personal names list (or the Who’s Who on the Web site) contains *see* references and lists alternate forms of names following the controlled name and functions as an interim measure for collocating name variations as well as providing brief biographical data. A future enhancement would include an active authority system applied to the Web site as is available in many automated library catalogs. DiLauro et al. (2001) describe an automated name authority control system for the creation of controlled metadata name entries, employing an indexing scheme able to locate and “learn” specific patterns and facilitate collocation of related names and/or concepts. This automated name authority system still requires an added controlled name element or field. Other possibilities that utilize an authority number may allow faster computer indexing and minimize the size of records that in current models and practice require adding controlled name entries in addition to names as they appear in text.

Authority Control and the Web as Authority Source Information

The head of Intellectual Access (a member of the project team), who functioned as the metadata consultant, created a personal names list. For completeness and consistency, the personal names list was used by production staff to enter the correct (i.e., controlled or authoritative form) and complete forms of personal names. Later, the personal names list was renamed “Who’s Who” (<http://yellowfever.lib.virginia.edu/reed/whoswho.html>). A list of geographic names was also created and became the “Places” component (<http://yellowfever.lib.virginia.edu/reed/places.html>). Both the Who’s Who and Places lists reside on the “Collection” side of the Philip S. Hench Walter Reed Yellow Fever Collection Web site.

Personal names appeared in many forms: first and last name; last name only; first name only; nicknames, etc. Often these names were of significant personages, such as presidents of countries, surgeon generals, or key players in

the yellow fever story. The example of George Sternberg illustrates the common problem with the collection material. Sternberg was surgeon general of the United States Army from May 30, 1893, to June 8, 1902, and a member of the Yellow Fever Commission. In the primary documents, his name can be found in the following forms:

Dr. Sternberg
G. M. Sternberg
General Sternberg
Geo. Sternberg
Sternberg
Uncle George Miller

The descriptive metadata (located in the header of the XML record) used the inverted form of “Sternberg, George Miller” and the header and title sections of the XML record used the direct form of “George Miller Sternberg.” Sternberg is an example of an easy or regular case, and his name was quickly and easily found in the NAF. Other challenging examples are described in the next section.

The Process

Name-authority procedures implemented in common cataloging workflow are often done with the work in hand. For this project, work-in-hand procedures were not an option. The required speed of production needed to meet the grant deadlines, combined with the fact that the digitized documents in the beginning were not loaded on a server and easily viewed or consulted, prevented work-in-hand processing for name authority. Accuracy and complete identification of people and places remained important to this project and efforts were made to maintain goals of effective information discovery and retrieval. In the early stages of processing, names were established without having full authority control applied and a different form of personal name was often chosen by the metadata analysts. Again, due to grant deadlines, the usual routine of consulting the NAF with the work in hand was not possible. In later processing, when a personal name authority was discovered for a locally created name (e.g., Carlos Finlay), it was noted with the phrase “Partial LC Authority [10/3/2000] Full LC is: Finlay, Carlos Juan, 1833–1915” and in that way flagged for a later processing cleanup routine. Whenever a name was added to the personal names list, the creation date was entered in brackets and corresponded to the date marked on worksheets that the metadata analysts gave to the head of Intellectual Access. The date served as a reference for backtracking to the original worksheets. A decision was made to not include birth and death dates with personal name entries.

As a result of the constraints discussed above, a hybrid workflow was developed. The workflow is referred to as

hybrid because it was an adaptation to facilitate meeting the grant requirements and to allow the employment of several part-time staff working simultaneously. The workflow was hybrid and not an ideal. An ideal workflow would have included the metadata analysts searching the NAF as documents were examined and data grids completed, with problem names set aside for later in-depth investigation. The hybrid workflow that was instituted required the compilation of lists of names by the metadata analysts. These lists of names or worksheets included bits of information that would enable the discovery and determination of proper or complete forms of personal names, names that were later worked on by the head of Intellectual Access as she created the authoritative personal names list and handled problem names. The salient bits of information included years, countries, activities, and so on—essentially anything that would assist in identifying a particular person. In addition, “Input Guidelines” for creating entries for incomplete names were devised. These guidelines were placed on an Intranet page available to all project staff (see figure 1). The personal names list was a simple html tagged file with a refresh command, which enabled the list to be built as project staff were completing the data grids and ensured that project staff would see the most current additions and corrections to the list (see figure 2).

The example of Hagedorn illustrates the hybrid process. On one analyst’s list appeared:

Hagedorn, [s.n.] author of a book on Leonard Wood

(Note: [s.n.] was a convention used when the first name was unknown.)

When “Hagedorn” was first searched in the NAF (via the Online Computer Library Center, Inc. [OCLC]) with the first name unknown, the search produced 93 entries. The examination of 93 entries was quickly deemed an ineffective use of time. Following a hunch, the head of Intellectual Access decided to use the Web and the Google search engine to find Hagedorn’s first name. The simple search phrase of “Hagedorn Leonard Wood” was used and almost immediately the full name of “Hermann Hagedorn” was discovered. These Web searches were often not complicated, as many names were unique and, when coordinated with the bits of information supplied by the metadata analysts, resulted in short hit lists. Since this collection was processed in a health sciences library and not a general academic library, there were few immediately accessible print world biographies, etc. The NAF was partially helpful, but not very useful when only the last name, such as “Hagedorn” or a common name such as “Henderson,” was supplied. Due to the nature of the collection material, many governmental Web sites were fre-

Transcribe only top level names in the personal name metadata field. Top level names include all persons associated with the yellow fever experiments, all members of the Reed family, and close relatives and friends. Many names appear in the letters that are nicknames or terms of endearment that will not be transcribed for subject access of personal names. **When names appear incomplete*, the following conventions will apply:**

- **Lawrence, Mrs.** (when only the title is known and not the first name).
- **Fort Thomas [Arizona Territory?]** (when the name of fort is given and you think you might know the location), see also agreed upon geographic and place terms.
- **[Beech...?], [s.n.]** (when transcribing what you think the last name is, and no first name is given)
- **Methren, [s.n.]** (when the first name was not given and we can not surmise the first name)
- **Rawley Springs [s.l.]** (when you know the name but not the location)
- **[s.n.], Jacob** (when the surname was not given and we can not surmise the surname)
- **[s.n.], Ellen** (same as above and we were given the information "Sister Ellen", see also agreed upon personal name terms)

* *The above conventions were adopted from the Anglo American Cataloging Rules, 2d edition, 1998 revision, section 22.20 "Undifferentiated Names" (418–19)*

Do not transcribe names for which only the words "brother/brother-in-law" or "sister/sister-in-law," etc. appear and the specific person can not be identified.

Figure 1. Input Guidelines

quently used and bookmarked. The Web became a gold mine of information (see tables 1 and 2). The instability of URLs, however, continues to be problematic. During this project and the final months of writing this article, the information sources consulted in table 1 continued to manifest URL changes. For purposes of documentation in this project, paper copies were made of all Web pages that supported the discovery and establishment of persons' names.

Seaman encouraged the work and development of the personal names list for purposes not only of standardization in the entry of descriptive metadata but also as a self-help resource for visitors of the Yellow Fever Web site and as a reference aid for e-mail inquiries. With this encouragement, the decision was made to enhance the list with as much information about each person as possible. A minor but interesting person on the Yellow Fever Web site was Mabel L. Conat, to whom Hench wrote in 1940 at the Public Library of Detroit. Conat's signature was not clear on her letter and while her name was typed on the responding letter from Hench, it was often the case that the spelling of a name from a respondent letter was misspelled. In this case, it turned out that the spelling was correct and a Web search provided verification. Conat was not in the NAF, but was mentioned in the ACRL's Guide to Policies and Procedures (www.ala.org/acrl/policy/polyindx.html). Her name was verified as Mabel L. Conat in Chapter 15. It was interesting to discover that she was ACRL president in 1942–43.

The Yellow Fever Web site features many Hispanic names, as the site involved relations with Cuba. One worksheet highlighted the name Estanislao Pardo Figueroa along with the notation of "President of the Academy of Medicine, Lima, Peru." Due to the hybrid workflow procedures, much of the personal name work was done in isola-

tion from the primary documents and name entry and identification was given scrupulous attention. As expected, Figueroa was not in the NAF. A string search of "Estanislao Pardo Figueroa" on the Web brought up the document "Imágenes Históricas de la Medicina Peruana" (http://200.48.26.79/bibvirtual/libros/Medicina/Ima_Histo_Med_Per/cap_27.htm). Chapter 27 noted that Estanislao Pardo Figueroa was president of the Academy of Medicine from 1919 to 1921.

The Web also proved to be helpful in the case of Francisco Dominguez Roldan. The metadata analysts struggled with this name, as it appeared in every possible form. A Web search led to the page entitled HISA (www.fiocruz.br/coc/hisa/itiC.htm), a sublink of the Web site of Casa de Oswaldo Cruz, a foundation that provides research and special projects and is part of the Brazilian medical institution known as Fundação Oswaldo Cruz. The HISA page contained the entry "Centenario del Nacimiento del Dr. Francisco Dominguez Roldan: 1864–1942," which led to a short biographical entry (www.fiocruz.br/coc/hisa/CTIC.HTM#1696). This Web search provided both name verification and the additional facts of Dominguez Roldan's birth and death dates.

Names that did not bear identification clues on the worksheets, such as Harold W. Jones, became prime candidates for employing the Web after NAF searches failed to produce sought-after names. Interestingly, Jones appeared on the National Library of Medicine's Web site at www.nlm.nih.gov/exhibition/tour/portraits3.html.

In our final analysis of May 2002, the personal names list (Who's Who on the Yellow Fever Web site) contained 1,692 unique name entries (not including the *see* references for other forms of names) and 265 places. Of the 1,692 names, 272 (16%) were verified and constructed using the NAF, leaving 1,421 (84%) names established by

```

<HTML>
<HEAD>
<TITLE>
PERSONAL NAMES MASTER LIST:
"WHO'S WHO" ON THE PHILIP S. HENCH WALTER REED
YELLOW FEVER COLLECTION WEBSITE
</TITLE>
</HEAD>
<BODY>
<meta http-equiv="Refresh" content="300;url=http://knowledgeweb.lib.virginia.edu/docs/cmas/intellect/wrnamaster.html">
<H1> MASTER LIST OF PERSONAL NAMES: "WHO'S WHO" ON THE PHILIP S. HENCH WALTER REED YELLOW FEVER COLLECTION
WEBSITE
</H1>
&nbsp;

<P><B>List Created and Maintained by Nadine P. Ellero *** LAST UPDATED 11/19/2001 AT 11:42 AM *** </B>

<P><B>Each category of information is delimited by a pipe symbol (|) according to the following format:</B>

<P><B>Last Name, First Name Middle Name | Additional information; Other names; etc. | Authority Source </B>

<P><B>NOTE: NAMES THAT HAVE AN ASTERISK (*) BEFORE THEM AND/OR ARE BOLDED CONTAIN INSTRUCTIONS TO BE
REPLACED WITH A FULLER FORM OF THE NAME. AS AN INTERVENING STAGE,AND FOR THE PUBLIC, THESE NAMES WILL
BECOME SEE REFERENCES TO THE PROPER FORM.</B>
<P>&nbsp;

<P><B>KEY</B>
<P><B>(date) = Reference Date(s) to place the person and their position/function in time</B>
<P><B>Local Source = Form of name created from original document(s)</B>
<P><B>LC = Library of Congress </B>
<P><B> [date] = Date the name was added to the list </B>
&nbsp;

<P>Abbot, William Richardson |Opened and taught a Charlottesville Institute attended by Reed; Was born 1839 -; not sure if the same as Prof. Abbott |
Local Source [8/21/2001]

<P>Abbott, Gen.|General Abbott (some military person) | Local Source [9/12/2000]

<P>Abbott, Prof. | Professor at the University of Virginia | Local Source [10/23/2000]

<P>Abbott, [s.n.] | See document 00916034; not sure if this is any of the other listed Abbotts | Local Source [9/11/2001]

<P>Abel, Dr. | Doctor in Louisville (1939) | Local Source [10/13/2000]

<P>Abercrombie, John William | (1866-1940) Alabama State Senator 1896-1898; University of Alabama President 1902-1911; United States Congressman
1913-1917; Superintendant of Education 1920-1926 | Local Source [4/10/2001]

<P><B>*Abercrombie, [s.n.] | REPLACE WITH ABERCROMBIE, JOHN WILLIAM |

<P>Acheson, Dean | Secretary of State, United States (January 21, 1949-January 20, 1953); see: www.spartacus.schoolnet.co.uk/USAacheson.htm | Partial
LC Authority [8/21/2001] Full LC is: Acheson, Dean, 1893-1971.

<P>Acker, Mollie Flint |South Bend, Indiana (1924); markup has Mollie "Heink" Acker which is not right |Local Source [11/16/2001]

<P>Acton, H. W. |Author; Physician? |Local Source [11/14/2001]

<P>Adams, E.S. | Major General, Adjutant General, The Adjutant General's Office, War Department, United States; | Local Source [8/29/2001]

```

Figure 2. Sample of the Personal Names List

**This page
intentionally
blank**

**This page
intentionally
blank**

<P>Adamson, Estelle |Student, Tulsa, Oklahoma (1927) | Local Source [10/22/2001]

<P>Adamson, William C. |Congressman from Georgia, United States; served from March 4, 1897, until December 18, 1917; see: <http://bioguide.congress.gov/scripts/biodisplay.pl?index=A000051> |Partial LC Authority [11/6/2001] Full LC is: Adamson, William Charles, 1854-1929.

<P>Agostini, Dr. | Physician, Cuban sanitation, 1908 | [3/20/2001]

<P>Agramonte, Aristides |A doctor; born June 3, 1866 and died August 17, 1931; congressional gold medal recipient | Local Source [10/2/2000]

<P>Agramonte, Eduardo | Father of Aristides Agramonte, died 1872? | [3/20/2001]

<P>Agramonte, Frances F. | Wife of Aristides Agramonte | Local Source [4/26/2001]

<P>*Agramonte, Mrs. |REPLACE WITH AGRAMONTE, FRANCES F. Wife of Aristides Agramonte | Local Source [10/17/2000]

<P>Ahrendts, J. L. |Representative, John Wyeth & Brother Inc., Philadelphia, Pennsylvania (1942) |Local Source [9/3/2001]

<P>Ainsworth, Frederick Crayton | United States Major General |Partial LC Authority [10/23/2000] Full LC is: Ainsworth, Frederick Crayton, 1852-1934.

<P>*Ainsworth, [s.n.] | MIGHT BE THE SAME AS FREDERICK CRAYTON AINSWORTH |1905 U.S. Army Medical Officer | 3/20/2001

<P>Alberti, Steward | Worked with Reed at Fort Apache | Local Source [8/20/2001]

<P>Albertini, A. Diaz | Cuban Physician; Director of Finlay Institute in Havana; full name is Antonio Diaz Albertini, however he preferred signing his name as "A. Diaz Albertini [8/17/2001] | Local Source [4/26/2001]

<P>*Albertini, Antonio Diaz | REPLACE WITH ALBERTINI, A. DIAZ |

<P>*Albertini, Dr. | REPLACE WITH ALBERTINI, A. DIAZ | Local Source [10/2/2000]

<P>Alderman, Edwin Anderson | Former University of Virginia President |Partial LC Authority [10/5/2000] Full LC is: Alderman, Edwin Anderson, 1861-1931.

<P>Alexander, Martha | Worked for the Journal of the History Medicine and Allied Sciences (1951) | Local Source [9/6/2001]

<P>*Allery, P. | WRONGLY TRANSCRIBED. SHOULD BE TILLERY, P. A. |

<P>*Allery, P. | WRONGLY TRANSCRIBED. SHOULD BE TILLERY, P. A. |

<P>Allison, William B. | United States Representative to Congress 1863-1871 and Senator 1873-1908 from Iowa |Partial LC Authority [10/23/2000] Full LC is: Allison, William B. (William Boyd), 1829-1908.

<P>Alspaugh, Edna |Student, Tulsa, Oklahoma (1927) | Local Source [10/22/2001]

<P>*Alvare, Dr. | REPLACE WITH ALVARE, IGNACIO |

<P>Alvare, Ignacio | Cuban physician, Havana; helped Hench with research on Camp Lazear, 1940 | Local Source [4/27/2001]

<P>Alvarez, Jacinto Mendez | Volunteer in the Yellow Fever Experiments, 1900, 1901 | [3/20/2001]

<P>Alvarez, Joaquin Maria | Cuban Physician; identified site of cemetery | Local Source [5/10/2001]

Figure 2 (cont'd). Sample of the Personal Names List

project staff. Of the 1,692 names, 107 were found to have authoritative biographies located on the Web and the corresponding URLs were added to the Who's Who list. In May of 2002, we rechecked the names not found originally

in the NAF and discovered that two (William Abbot and Domingos Freire) had been added to the NAF. The two names represent a percentage of less than 1% of the unfound personal names added in the intervening time

Table 1. Web Sources Frequently Used for People

Web Source: People	Web Address	Comments
Biographies of the U.S. Chiefs of the Army Corps of Engineers	www.hq.usace.army.mil/history/coe.htm	The title of this Web site as reviewed July 25, 2002, is "Commanders: Portraits and Profiles."
Famous West Point Graduates	www.dmi.usma.edu/Milresources/Generals/famgrads.htm	The title of "Famous West Point Graduates" changed to "Notable USMA Graduates," as viewed July 25, 2002.
Political Graveyard	http://politicalgraveyard.com/index.html	
Principle Officers of the Department of State	www.state.gov/www/about_state/history/officers.html	
Surgeon Generals of the Public Health Service	www.surgeongeneral.gov/library/history/sylist.htm	
Surgeon Generals of the United States Army	www.armymedicine.army.mil/history/tsgs/default.htm	
United States Congressmen	http://bioguide.congress.gov/biosearch/biosearch.asp	
United States State Senators	www.senate.gov/search/index.html	
United States Department of Interior's Secretary of the Interior	www.doi.gov/anniversary/secretaries.html	The United States Department of Interior's Web page had the following notice: "Access to the DOI website has been restricted in compliance with a court order. Select DOI Web pages will be made available to the public through a private internet service provider." (As viewed February 20, 2002.)
United States Party Leaders in Congress 1789–1997: Vital Statistics	www.house.gov/rules/97-136.htm	
United States Presidents	www.americanpresidents.org/ www.whitehouse.gov/history/presidents/index.html http://lcweb2.loc.gov/ammem/ndlpedu/orientation/preslist.html	
United States Secretaries of State	www.state.gov/www/about_state/history/sectravels2.html#tenure	
United States Secretaries of the Treasury	www.treas.gov/Architext/AT-allquery.html	
[United States] Secretaries of War and Secretaries of the Army: Portraits and Biographical Sketches by William Gardner Bell	www.army.mil/cmh-pg/books/sw-sa/SWSA-Fm.htm	
Yellow Fever Experimentations Congressional Gold Medal Awardees	http://dallaslibrary.org/CGI/goldmedals/yellowfever.html http://clerkweb.house.gov/histHigh/Congressional_History/goldMedal.php	

Table 2. Web Sources Frequently Used for Events

Web Source: Events	Web Address	Comments
<i>The Army Medical Department 1865–1917</i> by Mary C. Gillett	www.armymedicine.army.mil/history/booksdocs/spanam/gillett3/	
Public Health in Cuba 1865–1917	www.armymedicine.army.mil/history/booksdocs/spanam/gillett3/ch9.htm	This is Chapter 9 of a larger work: <i>The Army Medical Department 1865–1917</i> by Mary C. Gillett.
Records of the Military Government of Cuba	www.archives.gov/research_rom/federal_records_guide/military_government_of_cuba_rg14.html	Contains significant collection of Major General Ludlow's papers.

(i.e., time between the name was created for the Web site and the time the name appeared in the NAF) and in these instances, less than one month and eight months, respectively. Refinement work continues on the Web site and the Who's Who list is expected to grow over the next year.

Impact/Conclusion

The Web has the potential for tremendous impact on the ability of archivists, catalogers, and historians to flesh out the description of collections, bibliographic records, and the stories behind and surrounding primary materials. The

Web saved time and allowed us to work more effectively and meet project deadlines. We locally “cooperated” along departmental lines at TCMHSL to share expertise and created a Web resource that satisfied preservation goals, digitized primary materials, and enriched access to the yellow fever materials with the Who’s Who name list and biographical links. This local and internal cooperation became vital when grant deadlines necessitated a quick processing speed and the relinquishment of an ideal authority control procedure. Many personal names were mined from the Web and used for authority establishment, enhancement, and control, due to the size and continuing growth of the Web and associated metadata. In particular, government and educational sites are getting better and becoming more complete with information, easier to search and browse, and more prevalent. Schreiner and Somers (2002) have also recognized the Web as a gold mine of sources and compiled a listing of biographical Web resources.

Tillet (2001, 169–70) has proposed the construction and application of an International Name Authority File System as part of IFLA’s (International Federation of Library Associations and Institutions) effort of Universal Bibliographic Control (see: www.ifla.org/VI/3/ubcim.htm):

We could shift our attention from a single authorized form that everyone in the world had to accept and could instead share parallel or complementary records through the Internet—moving more into what I’ve called for years “access control” . . . Rather than exchanging authority records with the overhead of locally maintaining such a file, we would instead create a virtual database on the Internet that allowed simultaneous searching of multiple national authority files.

Tillet (2001, 3) describes a situation where not only is there an international authority file, but also the ability for searchers to customize their view of names or for library catalogs or search engines to customize a view for a certain audience(s):

We want to have the authorized form preferred by a library as the default offered to most users, but we can also envision offering user-selected preferences, through client software, or “cookies” that let the user specify once what their preferred language, script, or cultural preference is—for example for spelling preferences when cultures have variations, like American English and spelling preferences in the United Kingdom: labor and labour.

This kind of customization can be extremely useful for Web sites that involve international collaboration or cross international boundaries contentwise, such as the Yellow Fever Web site. A by-product of the Yellow Fever Web site was the Who’s Who names list that could become a source, albeit small, of names to aid in cooperatively adding to an international personal name depository. As OCLC has led the way for sharing bibliographic records, another system, whether it be one gigantic personal names depository (Barrueco Cruz et al. 2002) or a Z39.50 link is needed to facilitate order and increase searching precision and recall on the Web depositories (Tillett 2000). Tillett (2001, 5) states:

Authority control will help users of the Web to benefit from collocation and search precision that authority control enables and do it in ways that are meaningful to users in their preferred language and script.

It has been recognized that before such an international name authority system can be facilitated, professionals need to discuss and address the following questions as outlined by Françoise Bourdon (2001, 8–9):

[I]n May 2001, FRANAR decided to move its work toward the definition of functional requirements for authority records. It seems more pertinent first to know more about nature and functions of authority data we want to manage before defining an international numbering system supposed to identify them . . . When the concerned entities, the elements of data which constitute authority records, and the real or possible users of these records have been defined, how should we organize them? What are the characteristics of each entity, of each element of data, of each user? How are these components linked together to finally create an information system?

Work involving a standard authority number such as the International Standard Authority Number (ISAN), with author numbers as the authority or controlled form, has been proposed and is in early stages of design and implementation (Snyman and van Rensburg 2000).

The Philip S. Hench Walter Reed Yellow Fever Collection Web site is but one Web site that has used the NAF and the Web to flesh out its story and provide complete names of persons. It would have been easier and better if there were an international and enhanced name authority file, files, or system from which the many Hispanic names, such as Francisco Dominguez Roldan, could have been searched and verified. Would that an auto-

matic name authority control system existed for creation and searching of Web site materials. Work needs to be done to create new tools and depositories for the future. The Web as a communication tool, data exchange tool, and information warehouse or gateway offers us (i.e., catalogers, descriptive metadata creators, et al.) the opportunity to fly forward into the digital age. Vellucci (2001, 42) strongly urges catalogers to “expand their concepts of authority control, for although the underlying goals will remain the same, the authority control process will change. . . . And most critically, catalogers must actively participate in the development of system architectures and data registries.” Setting standards, working cooperatively, describing documents, and enhancing searching are the key characteristics in the cataloger’s hall of fame. OCLC, Cooperative Online Resource Catalog (COCR), RLIN, and NACO stand in this hall of fame as well as all the local initiatives not written about but seen in library catalogs, Web pages, and Web sites that have been enhanced with descriptive metadata. We need to seize these opportunities to create new systems and sources of information as well as continue to share new ways of working that get our work done faster and more efficiently (as Russell and Spillane [2001] concluded) and cooperatively create and contribute to products (such as an international Web name authority system) that can be consulted and shared by all. While panning for personal names on the Web was a highly useful and fun application for creating controlled name entries, the process can be made easier and more effective with an international and enhanced Web name authority system.

Looking even further into the future, Berners-Lee, Hendler, and Lassila (2001, 41-42) describe how the application of XML and RDF (Resource Description Framework) can encode meaning and relationships between concepts to enable the following scenario:

Suppose you wish to find the Ms. Cook that you met at a trade conference last year. You don’t remember her first name, but you remember that she worked for one of your clients and that her son was a student at your alma mater. An intelligent search program can sift through all the pages of people whose name is “Cook” (sidestepping all the pages relating to cooks, cooking, the Cook Islands and so forth), find the ones that mention working for a company that’s on your list of clients and follow links to Web pages of their children to track down if any are in school at the right place.

Here is another type of “name-hunt,” not too unlike the hunting experienced in establishing full names of persons on the Yellow Fever Web site and the dreams that exist in creating a global name depository or international

name authority file system. Information retrieval and precision (i.e., resource discovery) can be enhanced by harnessing the power of an international name authority system. We can employ techniques such as XML and RDF to the infrastructure of a new and improved Internet. A fulfilled dream would be to employ an automatic name authority system, a system that does not mandate a particular “right form” of name, yet recognizes related forms. Behind the scenes there may be an ISAN, ISADN, etc., pulling these variants together powered by some yet to be determined software and hardware. Cataloging principles and especially those of authority control will light and lead the way for a future built on standards, cooperation, and more effective resource description and discovery.

Notes

1. See www.loc.gov/catdir/pcc/intco/graphs01.html. Accessed Aug. 5, 2002.
2. Edgar Erskine Hume, death date of 1952. Accessed Aug. 5, 2002, www.arlingtoncemetery.com/eehume.htm; Foster Kennedy, death date of 1952. Accessed Aug. 5, 2002, www.whonamedit.com/doctor.cfm/1181.html; William Dosit Postell, death date of 1982. Accessed Aug. 5, 2002, www.tulane.edu/~matus/postell.html.

Works Cited

- Barrueco Cruz, José Manuel, et al. 2000. Personal data in a large digital library. Paper presented at Fourth European Conference on Research and Advanced Technology for Digital Libraries, September 18–20, 2000. Accessed March 11, 2002, <http://openlib.org/home/krichel/phoenix.html>.
- Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. The semantic web. *Scientific American* 284, no. 5: 35–43.
- Bourdon, Françoise. 2001. Functional requirements and numbering of authority record (FRANAR): To what extent authority control can be supported by technical means? Paper presented at sixty-seventh IFLA Council and General Conference, August 16–25, 2001. Accessed March 12, 2002, www.ifla.org/IV/ifla67/papers/096-152ae.pdf.
- Bradshaw, Elaine Beckley, and Stephen C. Wagner. 2000. A common ground: Communication and alliance between cataloger and curator for improved access to rare books and special collections. *College and Research Libraries* 61, no. 6: 525–34.
- Crosby, Ellen. 2000. Technical services in twenty-first century special collections. *Cataloging and Classification Quarterly* 30, no. 2/3: 167–76.
- DiLauro, Tim, et al. 2001. Automated name authority control and enhanced searching in the Levy Collection. *D-Lib Magazine* 7, no. 4: 1–8. Accessed January 11, 2002, www.dlib.org/dlib/april01/dilauro/04dilauro.html.
- Glogoff, Stuart J., and Garry J. Forger. 2001. Metadata protocols and standards: Bringing order to our digital objects. *Internet Reference Services Quarterly* 5, no. 4: 5–14.

- Hearn, Stephen S. 1999. Metadata structures and authority control. *Technicalities* 19, no. 6: 7–9.
- LC achieves cataloging milestones. *Library of Congress Information Bulletin* 51 (June 29, 1992): 290.
- LC name authority file expanded: Becomes Anglo-American authority file. *Library of Congress Information Bulletin* 53 (July 11, 1994): 293.
- Library of Congress. 2002. E-mail to the author.
- Milstead, Jessica, and Susan Feldman. 1999. Metadata: Cataloging by any other name. *Online* 23, no. 1: 25–31.
- Russell, Beth M., and Jodi Lynn Spillane. 2001. Using the Web for name authority work. *Library Resources and Technical Services* 45, no. 2: 73–79.
- Schreiner, Susan A., and Michael A. Somers. 2002. Biography resources: Finding information on the famous, infamous, and obscure. *College and Research Libraries News* 63, no. 1: 32–35, 39.
- Snyman, M. M. M., and M. Jansen van Rensburg. 2000. NACO versus ISAN: Prospects for name authority control. *The Electronic Library* 18, no. 1: 63–68.
- . 1999. Reengineering name authority control. *The Electronic Library* 17, no. 5: 313–22.
- Sturtevant, Carolyn (NACO Coordinator, Library of Congress). 2002. Further comments from Sturtevant, May 29.
- Tillett, Barbara B. 2000. Authority control at the international level. *Library Resources and Technical Services* 44, no. 3: 168–72.
- . January 2001. “Authority control on the Web.” Paper presented at bicentennial Conference on Bibliographic Control for the New Millennium, Final Version, January 23, 2001. Accessed March 11, 2002, http://lcweb.loc.gov/catdir/bibcontrol/tillett_paper.html.
- . August 2001. “A virtual international authority file.” Paper presented at sixty-seventh IFLA Council and General Conference, August 16–25, 2001. Accessed March 11, 2002, www.ifla.org/IV/ifla67/papers/094-152ae.pdf.
- Topper, Joby, and Historical Collections and Services. The Claude Moore Health Sciences Library. University of Virginia. 1997. “This most dreadful pest of humanity”: Yellow fever and the Reed Commission, 1898–1901. Accessed July 18, 2001, www.med.virginia.edu/hslibrary/historical/yelfev/tabcon.htm.
- Velluci, Sherry L. 2000. Metadata and authority control. *Library Resources and Technical Services* 44 (1): 33–43.

“Garbage” In, “Refuse and Refuse Disposal” Out

Making the Most of the Subject Authority File in the OPAC

Marguerite E. Horn

Subject access in the OPAC, as discussed in this article, is predicated on two different kinds of searching: subject (authority, alphabetic, or controlled vocabulary searching) or keyword (uncontrolled, free text, natural language vocabulary). The literature has focused on demonstrating that both approaches are needed, but very few authors address the need to integrate keyword into authority searching. The article discusses this difference and compares, with a query on the term “garbage,” search results in two online catalogs, one that performs keyword searches through the authority file and one where only bibliographic records are included in keyword searches.

Early catalog use studies indicated that most searching in a catalog was for known items (Cochrane 1985; Bodoff and Kambil 1998). With the advent of computerized catalogs, subject searching came to be the predominant target for users (Drabenstott and Vizine-Goetz 1994; Hildreth 1997; Matthews 1997; Bodoff and Kambil 1998). Early OPACs provided for subject searching only by the subject heading of the bibliographic record. However, keyword searching came into use almost immediately, with most OPACs allowing for word searches in subjects, titles, and notes. A decade ago, the big question was whether keyword searching alone would suffice for subject access. The conclusion was a resounding “no!”: controlled vocabulary (authorized terms) was absolutely necessary—but only if the relevant cross references were also supplied (Frost 1989; Jamieson, Dolan, and Declerck 1986; Marner 1993; Micco 1991; Smith 1991; Tillotson 1995). Users could not be expected to know the authorized subject term in order to perform subject searches. Markey (1988) suggested loading the entire Library of Congress Subject Headings (LCSH) into the OPAC to overcome this deficiency. Most libraries today, however, make do with authority records and cross references for headings actually used in the bibliographic records in their own catalogs.

Subject searching in OPACs continues to be problematic (Hildreth 1997; Matthews 1997; Yee and Layne 1998). For average users, a subject is just anything they wish to know “about.” The searcher has little or no understanding of the distinction in a catalog between “keyword” searching and “subject” searching. Most catalogs use the term “keyword” to mean “free text” and “subject” to mean “controlled vocabulary” searching. Nor does the user understand that subject searches

Marguerite E. Horn (HORNMA@sysadm.suny.edu), former Head, Cataloging Dept., University Libraries, University at Albany, State University of New York, is now New York Library Systems Implementation Specialist, Office of Library and Information Services, System Administration, State University of New York.

Manuscript received April 9, 2002; accepted May 16, 2002.

**This page
intentionally
blank**

**This page
intentionally
blank**

**This page
intentionally
blank**

**This page
intentionally
blank**

are based on left-anchored string searching, while keyword searches are generally based on words within a subject, title, or elsewhere in the bibliographic record (Yee and Layne 1998). Moreover, the average user does not understand that subject searches are based on *controlled* vocabulary used in the bibliographic record (for instance, LCSH) and represented in an authority file (Markey 1988; Cherry 1992; Drabentstott 1998; Smith 1991). Greenberg (1997) notes the failure of most OPACs even to refer to LCSH as the source of subject headings. Matthews (1997) identifies that even a keyword search of LCSH authorized headings (excluding cross references) will retrieve records only about 50% of the time.

In the typical online catalog, the distinction between keyword and controlled vocabulary subject searching, although present, is almost completely opaque to the user. Whether OPAC users actually choose "keyword" or "subject" as their search mode, the plain fact is that both end up being natural language searches in the absence of any guidance concerning the subject heading structure. If the term entered in a subject search happens to be the first word of the authorized form, then the user will likely find relevant citations. If the term entered in a subject search also happens to be the first word of a cross reference in a catalog that displays cross references, then the user will also be correctly directed. However, if the term entered in a subject search is a word *within* a subject heading or cross reference, then the user misses the authority control structure of the catalog.

How can we improve users' success in subject searching? One plan of attack would be to enrich the MARC bibliographic record, which is intellectually impoverished at best: it contains a very limited amount of conceptual and terminological variation upon which a search engine can operate (Drabentstott and Vizine-Goetz 1994). One method for enrichment would be to add more subject headings or improve subject analysis (Smith 1991; Drabentstott and Vizine-Goetz 1994), but this is not common nor often perceived as important—and would certainly be very time consuming. Nor does this approach solve the problem of user failure to understand controlled vocabulary.

Another method of enrichment would be adding tables of contents to the bibliographic record—in effect adding more keywords (Bodoff and Kambil 1998). But this method does not direct the user toward controlled vocabulary. One way around this lack would be to add keywords as cross references in the authority record (Micco 1991; Rada et al. 1988). However, since most catalogs do not use the authority record in keyword searches, this strategy would be of help only in a subject phrase search. Greenberg (1997, 112) notes, "Despite the popularity of [keyword searching], there has been little effort to link keyword searching to OPAC reference structures."

Users are frequently taught in online instruction classes to try a word search first; locate a relevant citation; examine the subject headings associated with the item; and then either do a new search with the relevant subject or, if the catalog allows it, request a further search by related items (Marner 1993; Greenberg 1997; Aanonsen 1987). Many librarians also use this approach. However, the average user of keyword searching rarely has the patience to wade through the retrieval set to actually perform the second, relevant search. Hildreth notes that users do not understand why keyword searches fail; we either need to train users better or improve our retrieval systems. The latter approach is preferable because "there will always be fewer systems to improve than users to instruct" (1997, 61). Borgman (1996, 501) concludes, "Most end users of online catalogs are perpetual novices who lack the requisite conceptual knowledge for searching. They need assistance in the translation process, whether provided by the system itself, by instruction in using the system, or by a search intermediary."

Hildreth's second approach to improving users' success would be to enhance systems rather than enrich records. Drabentstott and Weller (1995; 1996) have suggested a solution: use search trees to improve retrieval by subject. One of the suggestions includes "requiring the system to check whether keyword searches on user-entered queries that match cross references retrieve additional titles and enable/disable the 'expand search' option based on this system check" (1996, 535). Although this method uses the cross references in a keyword search, it is unclear if the intent is to search cross references as free text or as phrase searches. Micco suggests a system that "takes uncontrolled terms from wherever possible in the MARC record . . . and links these terms to the controlled vocabulary of the primary LCSH heading assigned to that work" (Basista, Micco, Rambler 1991, 89). It appears that this approach would use the established heading but not make use of cross references.

Libraries have long recognized the necessity of an online authority file, containing not only the authorized term but cross references directing the user to the correct term. Libraries are committing substantial staff time and dollars to maintaining and improving authority files. Meanwhile, these very authority files and the benefits they provide are often lost to the users of keyword searching. Because, for the most part, keyword searches are based on terms found in the bibliographic record only, these searches completely miss the cross references built into the authority file. It is worth repeating that the two methods of subject searching—subject term and keyword—are only joined together through the bibliographic record. The authority file is missed in the keyword search completely and the subject phrase search is only useful if the term used is the first word of the cross reference or the authorized form. What if keyword searches searched

cross references (4xx fields) in the authority record first and returned the related bibliographic records?

Demonstration of Keyword Searching through the Authority File

Method

For the purposes of demonstration, I sought a cross-reference term in LCSH that did not contain any of the words that were part of the authorized term (in order to avoid search results based on the occurrence of the word in the authorized subject heading). The term “garbage,” which is a cross reference to the term “refuse and refuse disposal” (and also “organic wastes”), turned out to be an excellent example.

I performed the test searches in two large Geac ADVANCE catalog systems: University at Albany, State University of New York (UAlbany) and New York University (NYU). At the time of this investigation, the two libraries had chosen different options in the Geac ADVANCE indexing structure. At UAlbany, keyword searches were automatically submitted to the authority file, returning not only words within a bibliographic record but also words within a cross reference. NYU had chosen the most common keyword indexing option, namely one that does not send keyword queries to the cross-reference structure.

Figure 1 shows how the Library of Congress Subject Authority record for “refuse and refuse disposal,” is displayed in the UAlbany catalog. The term searched, “garbage,” which is a MARC tag 450 or cross reference in the authority record, is bold-faced.

The UAlbany system permits separate keyword searches by title, author, subject, series, notes, and words in all fields. For this study, I searched by both subject word and keyword (all fields). A subject-word query searches the authority file, not only finding the term as the first element of an authorized form or cross reference, but also as a term within an authorized form or a cross reference. A keyword (all fields) query not only searches the authority file in the same manner, but also searches each bibliographic record by title, author, series, and notes.

Three different searches on “garbage” made up this study: subject, subject word, and keyword. The subject query (*s=garbage*) is a left-anchored phrase search, accessing the bibliographic records through the authority file. The subject-word query (*sw=garbage*) searches the authority file for the term within an authorized heading or a cross reference. The keyword query (*w=garbage*) searches the authority file in the same manner as the subject word query and in addition searches bibliographic records for the term within other fields, including title and

008			990830j anannbab	b ana
010	BB	a	sh 85112316	
040	BB	a	DLC	
		c	DLC	
		d	DLC	
053	B0	a	HD4482	
		b	HD4485	
		c	Economics	
053	B0	a	TD785	
		b	TD812.5	
		c	Engineering	
150	BB	a	Refuse and refuse disposal	
360	BB	i	subdivision	
		a	Waste disposal	
		i	under types of industries, industrial processes, facilities and institutions, e.g.	
		a	Construction industry—Waste disposal; Metals—Finishing—Waste disposal; Universities and colleges—Waste disposal	
450	BB	a	Discarded materials	
450	BB	a	Disposal of refuse	
450	BB	a	Garbage	
450	BB	a	Household waste	
450	BB	a	Household wastes	
450	BB	a	Rubbish	
450	BB	a	Solid waste management	
450	BB	a	Trash	
450	BB	a	Waste disposal	
450	BB	a	Waste management	
450	BB	a	Wastes, Household	
550	BB	a	Sanitation	
550	BB	a	Factory and trade waste	
550	BB	a	Pollution	
550	BB	a	Pollution control industry	
550	BB	a	Salvage (Waste, etc.)	
550	BB	a	Street cleaning	
550	BB	a	Waste products	
670	BB	a	LC database, May 7, 1999	
		b	(household waste; household wastes)	

[Note: The Geac system uses “B” to represent a blank, does not display delimiters, and places subfields on separate line]

Figure 1. Authority Record for Refuse and Refuse Disposal

notes. In most OPACs, these last two queries would only search the bibliographic record; in UAlbany’s OPAC, the word search is sent to the authority file for words in headings or cross references.

Results

A subject search in Geac (*s=garbage*) is a *phrase* search. The results are presented to the user as a subject index screen, alphabetically, from the authority file with bibliographic records attached. (Geac ADVANCE requires an authority record for each bibliographic heading.) Figures 2A and 2B present similar results in the public OPAC view

for the same subject search (s=garbage) at UAlbany and NYU. The NYU search presents additional information to the user by presenting all LC authorized headings and cross references alphabetically near "garbage," even if there are no bibliographic records in the NYU catalog (for instance, "Garbage [see Organic wastes] LCSH [0]). In both OPACs, the user who enters "garbage" as a subject search will be directed to use "organic wastes" or "refuse and refuse disposal" (as well as other terms starting with "refuse").

Of 51 records for "refuse and refuse disposal" at UAlbany (figure 2A), only one title, *Garbage as you like it*, had the word "garbage" in the bibliographic record. Of the 17 records for "refuse as fuel," one record for "organic wastes," and two records for "refuse collection," none had the word "garbage" in the bibliographic record. This means that in the most common keyword search (that is, one that does not send the keyword to the cross-reference structure), only one of these 71 records would be returned.

Of 41 records for "refuse and refuse disposal" in the NYU catalog (figure 2B), only two had "garbage" in the bibliographic record. Of eight records for "refuse as fuel," two records for "refuse collection," and one record for "rag pickers," none had the word "garbage" in the bibliographic record. This means that in an ordinary keyword search, only two of these 52 records would be returned.

I next searched "garbage" as a subject word in both catalogs. In most catalogs, including NYU, this search will only look for the word within a subject heading used in a bibliographic record. But at UAlbany, this search also looks for the word *within* an authorized heading *or* cross reference.

Figure 3A (UAlbany), shows an authority index screen from the public OPAC, similar to that produced by the subject search, but including only entries with "garbage" as a word (i.e., it does not present the authority file in the alphabetical neighborhood of "garbage" as the first word). This search returns the same records as the first for those headings and cross references beginning with "garbage," but additionally returns the cross reference "medical garbage *see* medical wastes," because "garbage" is a term within this cross reference, and also returns the cross reference for "University at Arizona Garbage Project." The "medical wastes" bibliographic records did not contain the term "garbage" anywhere. If this query had searched the bibliographic record only, then only the "garbage collection (Computer science)" and "Garbage Project (University of Arizona)" records would have been returned, because these are the only records with "garbage" in a subject heading.

Figure 3B (NYU) represents a traditional subject-word query, which searches for a keyword within the subject headings of the bibliographic record only (i.e., *not* incorporating the authority file or its cross references). The user is pre-

Browsing Subjects: S=garbage

Subject Heading	No. of Titles
1. Garaudy, R. (Roger)	
2. [See] Garaudy, Roger	(Subject) 5
3. Garaudy, Roger	(Subject) 5
4. Garay, Eugenio Alejandrino, 1874–1937	(Subject) 1
5. Garay, Juan de, 1528?–1583	(Subject) 1
>>>	
6. Garbage	
7. [See] Organic wastes	(Subject) 1
8. Garbage	
9. [See] Refuse and refuse disposal	(Subject) 51
10. Garbage as fee—Law and legislation—United States	(Subject) 1
11. Garbage as fuel	

Subject Heading	No. of Titles
Garbage as fuel	(continued)
1. [See] Refuse as fuel	(Subject) 17
2. Garbage collection	
3. [See] Refuse collection	(Subject) 2
4. Garbage collection (Computer science)	(Subject) 1
5. Garbage Project (University of Arizona)	(Subject) 1
6. Garbarz, Moshe	(Subject) 1
7. Garbett, Cyril Forster, Abp. of York, 1875–1955	(Subject) 1

Figure 2A. (UAlbany) Online Catalog—Heading Browse

sented with a browse screen of titles, all of which have "garbage" in a subject heading. Of these six, three are for "garbage can models of decision making," one is for "Garbage Project (University of Arizona)," one is for "garbage collection (Computer science)," and one is for "Memphis (Tenn.)—Garbage strike 1968." In addition to retrieving far fewer records than the UAlbany "sw=garbage" search, this query yields what might fairly be called low precision as well.

Finally, I searched the term "garbage" as a keyword, resulting in 113 records at UAlbany and 74 records at NYU. The browse screen for both results is an undifferentiated list of bibliographic records with no indication of where the term appeared. The word search at UAlbany implicitly searches the authority file with its cross references and also the bibliographic file; the same search at NYU searches only bibliographic records.

In figure 4A (UAlbany), all of the records in the first two searches (subject and subject word) are returned in this search, as well as all records with "garbage" somewhere in the bibliographic record *beyond the subject fields*. Because there is no indication that cross references are being searched, users may be confused as to why they actually retrieved some of the records in response to the search. As noted above, in discussion of the subject search and subject-word search, 71 of the 113 records do not have the term "garbage" anywhere in the record. Of the remaining

Browsing Subjects: S=garbage

Subject Heading	No. of Titles
1. Garawi	
2. [See] Sudan grass	(LCSH) 0
3. Garay, Juan de, 1528?-1583	(LCSH) 1
4. Garay, Martin de, 1760-1825	(LCSH) 1
5. Garay, Sindo	(LCSH) 1
6. Garba, Joseph Nanven, 1943-	(LCSH) 1
>>>	
7. Garbage	
8. [See] Organic wastes	(LCSH) 0
9. [See] Refuse and refuse disposal	(LCSH) 41
10. Garbage Analysis Programme	(LCSH) 0
11. Garbage as feed	(LCSH) 0
12. Garbage as fuel	
13. [See] Refuse as fuel	(LCSH) 8

Browsing Subjects: S=garbage

Subject Heading	No. of Titles
1. Garbage can models of decision making	(LCSH) 2
2. Garbage can models of decision making— Congresses	(LCSH) 1
3. Garbage collection	
4. [See] Refuse collection	(LCSH) 2
5. Garbage collection (Computer science)	(LCSH) 1
6. Garbage collectors	
7. [See] Sanitation workers	(LCSH) 0
8. Garbage pickers	
9. [See] Ragpickers	(LCSH) 1
10. Garbage Project (University of Arizona)	(LCSH) 1
11. Garbage trucks	
12. [See] Refuse collection vehicles	(LCSH) 0
13. Garbagemen	
14. [See] Refuse collectors	(LCSH) 0

Figure 2B. (NYU) Online Catalog—Heading Browse

42 records returned, 14 have “refuse and refuse disposal—<subdivision>” as a subject heading and seven have what could be considered related headings of “marine waste” or “environmental engineering.” Therefore, if this search had operated as a bibliographic keyword search, only 42 titles would have been returned, 21 having nothing to do with waste management at all. Literary titles, song titles, and descriptive notes make up the remainder of the results of this search.

In figure 4B (NYU), there is a return that appears similar to the UAlbany search, but does not include titles that would have been returned from cross references. Of 74 titles returned, only 19 had “refuse and refuse disposal—<subdivision>” as a subject heading; seven had a related environmental heading; six were the same as in the subject word search. This left 42 titles (more than 50%) that were totally unrelated to waste products. NYU has a much larger collection of popular song recordings than UAlbany, resulting in a higher number of unrelated titles. Certainly a user presented with these results would be hard put to find an

Your Search: SW=Garbage

Subject Heading	No. of Titles
1. Garbage	
2. [See] Organic wastes	(Subject) 1
3. Garbage	
4. [See] Refuse and refuse disposal	(Subject) 51
5. Garbage as feed—Law and legislation—United States	(Subject) 1
6. Garbage as fuel	
7. [See] Refuse as fuel	(Subject) 17
8. Garbage collection	
9. [See] Refuse collection	(Subject) 2
10. Garbage collection (Computer science)	(Subject) 1
11. Garbage Project (University of Arizona)	(Subject) 1
12. Medical garbage	
13. [See] Medical wastes	(Subject) 3
14. University of Arizona—Garbage Project	
15. [See] Garbage Project (University of Arizona)	(Subject) 1

Figure 3A. (UAlbany)**Online Catalogue—Title Summary**
Your Search: SW=GARBAGE

Author/Title	Year
1. Ambiguity and command: organizational perspectives on military decision making	1986
2. Lentz, Richard Sixty-five days in Memphis : a study of culture, symbols, and the press	1986
3. Rathje, William L. Rubbish! : the archaeology of garbage	1992
4. The logic of organizational disorder	1996
5. Jones, Richard, 1954- Garbage collection : algorithms for automatic dynamic memory management	1996
6. Organizing political institutions : essays for Johan P. Olsen	1999
6 titles in list	

Figure 3B. (NYU)

appropriate record, find the correct subject heading, and then resume the search.

Summary and Conclusions

Subject searching in most OPACs remains problematic because users rarely know the difference between “key-word” and “subject” searching. They have little conception of controlled vocabulary except when stumbling over a cross reference in a phrase search. Hence, most OPAC queries turn out to be no better than keyword searches. Unlike the UAlbany catalog, most OPACs do not even take advantage of the authority file in keyword searches. That is,

Your Search: W=Garbage

Author/Title/Volume	Year
1. Lee, James A., 1922– The gold and the garbage in management theories and prescriptions	1980
2. Savas, E. S. The organization and efficiency of solid waste collection	1977
3. Young, Dennis R., 1943– How shall we collect the garbage? A study in economic organization.	1972
4. Fairfield, Roy P. Humanizing the workplace	1974
5. Rist, Ray C. The pornography controversy : changing moral standards in American life	1974
6. Hsu, Vivian Ling Born of the same roots : stories of modern Chinese women	1981
7. Dos Passos, John, 1896–1970 The garbage man : a parade with shouting	1929
8. Herbert, Brian Sidney's comet : being an account of the remarkable events which occurred during the approach of the Great Garbage...	1983
9. Platt, Charles Garbage world	1967
10. Darlington, Arnold Ecology of refuse tips	1969
11. Leckie, James O., 1939– Other homes and garbage : designs for self-sufficient living	1975

[An additional 102 records are presented to the user in a similar manner]

Figure 4A. (UAlbany)

they do not return bibliographic records having the search term in any of the fields, nor do they return cross references having the search term in any portion of the cross reference.

This investigation has revealed at least some strategies libraries can adopt to help solve this problem. For example:

1. A keyword query should be sent to the authority file first, returning authorized headings and cross references that inform the searcher of the authorized/controlled vocabulary headings. The sample search "sw=garbage" in the UAlbany OPAC returned the authority index screen, suggesting the appropriate subject headings through cross references and also finding the term within a cross reference.
2. If keyword searches are sent to the authority file, then the user should be presented with the authorized headings first (i.e., index screen with cross references), with an option to continue the search to bibliographic records only. Presenting users with an undifferentiated list of records is not helpful (as in the returns for "w=garbage"). Greenberg (1997, 112) notes that "Perhaps intelligent access to reference structures could even help to resolve a number of the

Your Search: W=Garbage

Author/Title/Volume	Year
1. McQuade, Walter, comp. Cities fit to live in and how we can make them happen; recent articles on the urban environment.	1971
2. Other homes and garbage : designs for self-sufficient living	1975
3. Fanning, Buckner. Throw away the garbage	1976
4. Lee, James A., 1922– The gold and the garbage in management theories and prescriptions	1980
5. Melosi, Martin V., 1947– Garbage in the cities : refuse, reform, and the environment : 1880–1980	1981
6. Xavier, Ismail Norberto. Allegories of underdevelopment [microform] : from the "aesthetics of hunger" to the "aesthetics of garbage"	1982
7. Young, Dennis, 1943– How shall we collect the garbage? A study in economic organization.	1972
8. Perls, Frederick S. In and out the garbage pail	1969
9. Erganian, George K. Effects of community-wide installation of household garbage-grinders on environmental sanitation	1973
10. Kelly, Katie. Garbage; the history and future of garbage in America.	1981
11. Born of the same roots : stories of modern Chinese women	1981
12. Neal, Homer A. Solid waste management and the environment : the mounting garbage and trash crisis	1987
13. Ambiguity and command: organizational perspectives on military decision making	1986
14. Lentz, Richard. Sixty-five days in Memphis : a study of culture, symbols, and the press	1986
15. Hershkowitz, Allen. Garbage burning : lessons from Europe: consensus and controversy in four European states	1986
16. Fassbinder, Rainer Werner, 1946– Plays. English. Selections	1985
17. Hershkowitz, Allen. Garbage management in Japan : leading the way	1987
18. Kirshner, Dan. To burn or not to burn : the economic advantages of recycling over garbage incineration for New York City.	1985
19. Dixon, Stephen, 1936– Garbage : a novel	1988
20. Rush to burn : solving America's garbage crisis?	1989

[An additional 54 records are presented to the user in a similar manner]

Figure 4B. (NYU)

retrieval overload problems associated with keyword searching." A searcher may have no idea why records for "refuse and refuse disposal" are retrieved, for example, when the word searched ("garbage") does not appear in the bibliographic record.

3. If a keyword search were sent to the authority file first, then adding common terms to the authority file as cross references would increase chances of returning more relevant records. Following a suggestion by Micco (1991), we might use a work's table of contents as a guide to terms that might usefully be added to the authority record as cross references. If, for instance, the keyword "rubbish" were in the table of contents, but not a cross reference on the authority record for the corresponding subject heading, adding it as a cross reference would improve retrieval.

These results raise the question of how much preemptive control OPAC designers should exercise over users' choices when they select a particular search type. For instance, in most OPACs, the default condition for a keyword search is "keyword anywhere." However, the default condition for a subject search is most often a left-anchored phrase search. Even a subject keyword search typically will not access the authority file. Hence, one kind of strategy consistent with the findings of this paper would be to redefine a keyword search as a *subject* keyword search including access to the authority file. The user does not need to know this; this approach, in most cases, will improve both precision and recall.

In an era of patron empowerment, this may not be a popular move, but at least for the naive user, it may initially provide the most useful results. Experienced users can always opt for more advanced techniques.

Increasingly sophisticated search and retrieval software, together with complex bibliographic record structures, offer the possibility of significant improvement in the performance of subject searches in online library catalogs. But this will not happen unless we take an innovative approach to exploiting the controlled indexing and searching capabilities of the next generation of integrated library systems. We already know what some of these strategies might look like. We may not be able to reduce the incidence of garbage in, but we can certainly reduce the incidence of garbage out.

Works Cited

- Aanonson, John. 1987. A comparison of keyword subject searching on six British university OPACs. *Online Review* 11, no. 5: 303-13.
- Basista, Thomas, H. Mary Micco, and Linda Rambler. 1991. Designing the OPAC user interface to improve access and retrieval. *Microcomputers for Information Management* 8, no. 2: 87-103.
- Bodoff, David, and Ajit Kambil. 1998. Partial coordination I: The best of pre-coordination and post-coordination. *Journal of the American Society for Information Science* 49, no. 14: 1,254-69.
- Borgman, Christine L. 1996. Why are online catalogs *still* hard to use? *Journal of the American Society for Information Science* 47, no. 7: 493-503.
- Cherry, Joan M. 1992. Improving subject access in OPACs: An exploratory study of conversion of users' queries. *Journal of Academic Librarianship* 18, no. 2: 95-99.
- Cochrane, Pauline A. 1985. *Redesign of catalogs and indexes for improved online subject access: Selected papers of Pauline A. Cochrane*. Phoenix, Ariz.: Oryx.
- Drabenstott, Karen M. 1998. *Understanding subject headings in library catalogs*. Ann Arbor: School of Information, University of Michigan.
- Drabenstott, Karen M., and Diane Vizine-Goetz. 1994. *Using subject headings for online retrieval: Theory, practice, and potential*. San Diego: Academic.
- Drabenstott, Karen M., and Marjorie S. Weller. 1996. Failure analysis of subject searches in a test of a new design for subject access to online catalogs. *Journal of the American Society for Information Science* 47, no. 7: 519-37.
- . 1995. *Testing a new design for subject access to online catalogs*. Ann Arbor: School of Information and Library Studies, University of Michigan.
- Frost, Carolyn O. 1989. Title words as entry vocabulary to LCSH: Correlation between assigned LCSH terms and derived terms from titles in bibliographic records with implications for subject access in online catalogs. *Cataloging and Classification Quarterly* 10, no. 1/2: 165-79.
- Greenberg, Jane. 1997. Reference structures: Stagnation, progress, and future challenges. *Information Technology and Libraries* 16, no. 3: 108-19.
- Hildreth, Charles R. 1997. The use and understanding of keyword searching in a university online catalog. *Information Technology and Libraries* 16, no. 2: 52-62.
- Jamieson, Alexis J., Elizabeth Dolan, and Luc Declerck. 1986. Keyword searching vs. authority control in an online catalog. *Journal of Academic Librarianship* 12, no. 5: 277-83.
- Markey, Karen. 1988. Integrating the machine-readable LCSH into online catalogs. *Information Technology and Libraries* 7, no. 3: 299-312.
- Mamer, Jonathan C. 1993. Measuring the success of keyword search strategy in an online catalog. *Technical Services Quarterly* 11, no. 2: 1-11.
- Matthews, Joseph R. 1997. Time for new OPAC initiatives: An overview of landmarks in the literature and introduction to WordFocus. *Library Hi Tech* 15, no. 1/2: 111-22.
- Micco, Mary. 1991. The next generation of online public access catalogs: A new look at subject access using hypermedia. *Cataloging and Classification Quarterly* 13, no. 3/4: 103-32.
- Rada, Roy, et al. 1988. Creating and evaluating entry terms. *Journal of Documentation* 44, no. 1: 19-41.
- Smith, Elizabeth H. 1991. Enhancing subject accessibility to the online catalog. *Library Resources and Technical Services* 35, no. 1: 109-13.
- Tillotson, Joy. 1995. Is keyword searching the answer? *College and Research Libraries* 56, no. 3: 199-206.
- Yee, Martha M., and Sara Shatford Layne. 1998. *Improving online public access catalogs*. Chicago: ALA.

**This page
intentionally
blank**

**This page
intentionally
blank**

Cooperative Cataloging, Vendor Records, and European Language Monographs

Charlene Kellsey

The appearance in OCLC and RLIN of minimal level catalog records from European book vendors for European language monographs and their effect on cataloging department workflows and cooperative cataloging efforts have been matters of concern expressed recently at ALA meetings and in the library literature. A study of 8,778 catalog records was undertaken to discover how many current European language monographs were being cataloged by the Library of Congress, by member libraries, and by vendors. It was found that vendor records accounted for 16.7% of Spanish books, 18% of French books, 33.6% of German books, and 52.5% of those in Italian. The number of libraries enhancing vendor records in OCLC was found to be only approximately one-third the number of libraries contributing original records for European language books. Ongoing increases in European book publishing and the increasing globalization of cataloging databases mean that the results of this study have implications not only for local cataloging practice but for cooperative cataloging as a whole.

Cooperative cataloging, or the sharing of the work of creating catalog records for books being added to libraries' collections, has been important to librarianship for a long time. It began with the distribution of catalog card sets by the Library of Congress (LC) in the early twentieth century and accelerated in the 1970s with the development of online databases, such as OCLC and RLIN, to which members could contribute original records. The ideal of cooperative cataloging has been to create a catalog record for any given book only once and then share the record with other libraries that need it, thus eliminating duplication of effort and diminishing the amount of original cataloging that any single library would have to do. The development of the Program for Cooperative Cataloging (PCC) in the 1990s was a successful effort to guarantee a standard level of quality in the records contributed by participants at the same time encouraging the contribution of larger numbers of high quality records.

A new development, however, seems to be undermining some of the progress in cooperative cataloging that has benefited libraries to date. Beginning in 1996, OCLC and RLIN began loading minimal level catalog records from several European book vendors into their databases. A number of articles in the library literature recently have raised questions about the value of these minimal level vendor catalog records for European language monographs and their effect on catalog department workflows and national cooperative cataloging efforts (Beall 2000; Shedenhelm and Burk 2001). OCLC maintains that its database, WorldCat, is not just a cataloging database anymore. Vendor

Charlene Kellsey (kellseycz@glinda.colorado.edu) is Assistant Professor, Monographic Cataloger, and Classics Bibliographer at the University of Colorado at Boulder Libraries.

Manuscript received February 18, 2002; accepted April 22, 2002.

records are valuable for the acquisitions process and useful to reference librarians trying to identify the existence of a title for their patrons (OCLC 2002). Vendor records are basically brief acquisition records that do not contain classification numbers or subject headings, and although catalogers at the Casalini Libri firm have recently been trained by a Library of Congress representative in the LC classification and subject heading systems, their enhanced catalog records will only be available to customers for an extra charge; they will not be available through the utilities (Casalini 2002). In defense of the vendors, it is really not their responsibility to provide full cataloging records for the books they sell; hiring catalogers to do this incurs costs that vendors, as businesses, need to recover. In fact, Casalini Libri estimates that they will charge two Euros per enhanced record (approximately \$1.80). While this may seem like a bargain to many libraries, it must be remembered that these records still will not meet the standards for full level U.S. records. The larger concern about delegating a heretofore public cooperative activity to private companies also needs to be addressed. Thus the effect on cooperative cataloging of vendor records in the databases needs to be documented and understood so that the library cataloging community can effectively respond to the changes these records have generated in cataloging work processes. The present study is intended to begin that documentation in order to inform the discussion of problems and possible solutions.

The effect of vendor records on cataloging department processes has several aspects. Since the vendor records do not contain classification numbers or subject headings, and name and series headings often do not match the form of heading found in the U.S. national authority file, the records require almost as much work by catalogers as creating an original record. Books for which a vendor record is found in OCLC or RLIN, however, often go to the copy cataloging unit, where time and effort may be spent to determine that the item needs the attention of an original cataloger. The time and effort required of original catalogers to upgrade a minimal level vendor record may be similar to that required to create a new record but the effect on costs to their institution may be different, and many libraries that are allowed to add new records to a national database such as OCLC may not be authorized to enhance existing records. The result of this situation would seem to be that more libraries would download vendor records and upgrade them in-house before adding them to their catalogs, which anecdotal evidence suggests is happening. Fewer libraries would then be contributing full level catalog records to the national databases and more libraries would be duplicating the effort of upgrading the records in-house, thus undermining the cooperative cataloging that benefits all libraries as well as the companies to which many libraries outsource their cataloging.

A recent retrospective study of Italian monographs documents the trend of an increasing number of vendor records for Italian monographs in OCLC at the expense of member-contributed records (Kellsey 2001). That study found that, in 2000, 60% of records for books in Italian were contributed to OCLC by a vendor while 30% were contributed by LC and only 10% by member libraries. This was a significant change from 1996 when 24% percent of records for Italian books were contributed by LC and 76% by member libraries. The current study expands that investigation to include French, German, and Spanish language monographs, in addition to Italian, and covers a variety of subject areas in order to determine whether the results of that earlier study remain valid when larger numbers of records are examined. If vendor records represent an increasing percentage of catalog records for European language monographs, then it would also be important to know how many libraries contribute original records and how many upgrade vendor records in OCLC, so the study also collected that information.

Method

The online catalog of the University of Colorado at Boulder Libraries was used to gather the data for this study. The UCB Libraries is a member of the Association of Research Libraries, with holdings of more than three million volumes (not including documents, microforms and special collections) and annual acquisitions of approximately 30,000 new monographic volumes. Around 22,000 undergraduate students and 4,000 graduate students are registered. Ph.D. programs exist in all of the subject areas represented in this study except German, which offers an MA, and Italian, which offers a BA. The libraries receive approval slips from Otto Harassowitz, Casalini Libri, Blackwell's, and Aux Amateurs du Livre, so active ordering of western European language monographs is a regular part of the acquisitions program. The Cataloging Department participates in OCLC's Enhance program and the Program for Cooperative Cataloging's Name Authority Cooperative program (NACO), Subject Authority Cooperative program (SACO), and Bibliographic Cooperative program (BIBCO). It is Cataloging Department policy to enhance minimal level records in OCLC before exporting them to the local catalog.

To gather data for this study, the "Create Lists" function of the Innovative Interfaces online system in the UCB Libraries was used to collect records of books cataloged in 1999 and 2000 in a number of different call number areas. They included: B (philosophy), D (general European history), DC (history of France), DD (history of Germany), DF (history of Greece), DG (history of Italy), DP (history

of Spain), DT (history of Africa), F 1201-3799 (history of Latin America), PA (classics), PQ (French, Italian, and Spanish literature), and PT (German literature). The lists generated were then sorted by language and by the library codes found in the MARC tag 040 (cataloging source) and then manually tabulated to discover the number of records originally contributed to OCLC by the Library of Congress, by member libraries, and by European book vendors for English, French, German, Italian, and Spanish language monographs. The vendors were Casalini Libri, Iberbook International, Otto Harrassowitz, Puvill Libros, and Jean Touzot.

After the initial tabulation of the number of records input by each category (LC, member, and vendor), a further tabulation was done of how many member libraries were contributing new records and how many were enhancing vendor records for European language monographs. Since concern has been expressed in the literature and at ALA meetings that the appearance of large numbers of vendor records in the national databases may be having an effect on the cooperative cataloging efforts of member libraries, documenting the current state of cooperative cataloging seemed crucial. Recommendations for change, whether locally or nationally, need to be based on accurate knowledge of current practice.

The tabulation of member library contributions and enhancements was done by creating charts of all the library code symbols in subfield \$a of the 040 MARC tag, which represents the library that initially created the record. The number of occurrences of that code were then counted for each language. In order to determine how many libraries were upgrading vendor records, the first code appearing in subfield \$d was counted for records that had a vendor's symbol in the subfield \$a. Although often more than one code appears in subfield \$d, representing other libraries that have modified the record in some way, in OCLC it is impossible to tell what a particular library has done to the record. Based on cataloging experience, it seems that the first library to modify the record usually adds a call number, verifies the name and series headings, and adds one or more subject headings. Although other libraries may add a call number from a different scheme or additional subjects, most of the critical work has been done by the first library, so it was decided only those would be counted.

Records already in the UCB catalog were used for this investigation, rather than trying to capture information as books passed through the catalog workflow, for several reasons. One was to avoid interruption of the workflow in a large and busy cataloging department. More significant, however, was the importance of gathering a large set of data in order to improve the reliability of conclusions drawn from it. By selecting all the records in several call number areas for the two most recently completed cataloging years,

it was possible to analyze information from 8,778 records in a matter of months, rather than having to wait for enough new items to come in, at irregular intervals, to gather a large amount of data. The call number areas chosen represent European literature, history, and area studies as well as the humanities fields of classics and philosophy. These fields were selected as being the most likely to have significant numbers of European language monographs on which to base the study.

Results

Table 1 represents the number of records for monographs in English versus those in European languages (French, German, Italian, and Spanish) for the subject areas examined. Although the percentages in each language vary greatly by subject area, and the amount of Spanish literature received by UCB may be unusually large, it can be seen from the totals that more than half of the monographs purchased by UCB in these areas are not in English. Since the data in table 2 show that the Library of Congress catalogs only 23–38% of European language monographs, versus almost 75% of English language monographs in these areas, it is clear that European language monographs represent a significant cataloging workload for those libraries collecting them.

In fact, one of the motivating factors for the development of the National Coordinated Cataloging Program (precursor to the Program for Cooperative Cataloging) was the Library of Congress's need for help in getting cataloging done for European language materials. The initial participants were assigned subject areas for which they would contribute records, almost all of which were in area studies, literature, and humanities in Spanish, German, French, and Italian (Rosenblatt 1993). This need appears not to have diminished.

The number of vendor records varies quite a bit by language, from a low of 16.7% for Spanish books to a high of 52.5% for those in Italian, reflecting the contribution of records by the different vendors (table 2). Because a much larger number of books in Spanish are received, however, the actual number of vendor records for Spanish books is larger than for the other three languages.

Tables 3 and 4 show the number of libraries contributing original records for European language monographs versus the number upgrading vendor records in OCLC. Totals for each column were not included since many libraries contributed or enhanced records in more than one language. As can be seen when comparing the tables, many fewer libraries upgrade records than contribute original records. One of the surprising findings of this part of the study was the large number of libraries that contribute five

Table 1. Number of Monographs in English and European Languages, Cataloged at University of Colorado, Boulder, 1999–2000

Classification	English Language		4 European Languages		Total Monographs
		%		%	
B	840	83.9	161	16.0	1,001
D	769	91.2	74	8.8	843
DC	152	77.0	45	22.8	197
DD	110	60.0	73	40.0	183
DF	79	70.5	33	29.5	112
DG	199	49.6	202	50.4	401
DP	67	36.0	119	64.0	186
DT	324	95.3	16	4.7	340
F	357	52.2	327	47.8	684
PA	270	46.2	315	53.8	585
PQ (Fre)	377	32.3	790	67.7	1,167
PQ (Ital)	101	52.3	92	47.7	193
PQ (Spa)	209	10.3	1826	89.7	2,035
PT	197	23.1	654	76.9	851
Totals	4051	46.1	4727	53.9	8,778

Table 2. Source of Cataloging for English and European-Language Monographs, 1999–2000

	LC		Member		Vendor		Total Records
		%		%		%	
English	3,026	74.7	996	24.6	29	0.7	4,051
French	229	25.3	514	56.7	163	18.0	906
German	237	22.8	453	43.6	350	33.6	1,040
Italian	122	25.7	103	21.7	249	52.5	474
Spanish	869	37.7	1,052	45.6	386	16.7	2,307
Total foreign	1,457	30.8	2,122	44.9	1,148	24.3	4,727
Total	4,483	51.0	3,118	35.5	1,177	13.4	8,778

or fewer original records in these four languages. This would seem to indicate that many libraries, while not contributing large numbers of records for foreign language books, do contribute a few records for items they receive that do not yet have a record in OCLC, in the spirit of cooperative cataloging. The fact that only roughly a third as many libraries enhance vendor records in OCLC, as shown in table 4, certainly indicates a loss of some of the benefits of cooperative cataloging. Further study of the libraries that contribute original records but do not enhance records would be useful in order to identify barriers to upgrading records and possible incentives that would encourage more libraries to upgrade records for the benefit of all.

Discussion

With the exception of the preliminary study noted previously, there have not been any previous quantitative studies of vendor records in OCLC (Kellsey 2001). There have, however, been a few previous studies of the avail-

ability of LC and member records in OCLC. Metz and Espley (1980) studied 396 monographs received at [Virginia Polytechnic Institute and State University] and Struble and Kohlberger (1987) studied 7,062 items at the University of Pittsburgh. Both studies were concerned with the availability of cataloging copy in OCLC on receipt of the books and after different periods of time in order to optimize the cataloging workflow and minimize multiple searching for the same item. Although the goals of those earlier studies differed from the current study, it is possible to extrapolate some comparable data from their tables to provide a view of the availability of LC and member records over a period of twenty-one years.

Table 5 illustrates the source of catalog records found in the three studies. Several caveats should be kept in mind when interpreting this table. Metz and Espley selected their books into groups of American imprints, British imprints and other foreign imprints by a ratio of 4:2:1 and since their total number of books was small (396), they looked at only 58 foreign imprints. Since the current study

did not separate British imprints from U.S. imprints, but included them in the English language category, and Metz and Espley only gave percentages, not numbers of items, in their tables, there was no way to calculate their inclusion, so they have been omitted from table 5. Metz and Espley also did not specify countries included in foreign imprints, so this category may have included languages in addition to the four in the current study. The percentages for LC copy were taken from table 3 of Metz and Espley's study, and the percentage of member copy was inferred from this (which would include any original cataloging their library may have had to do).

The study of Struble and Kohlberger notes that they excluded Slavic materials, and their tables include a breakdown by the four languages used in this study (although they included Portuguese with Spanish). The author of the current study calculated the percentages in table 5 from the numbers of items listed by Struble and Kohlberger in their charts. In calculating the percentages, items with no copy at the end of the study were included with the member copy since presumably the study library would have had to then catalog them.

It is interesting to note the close similarity in percentages of LC records for English language books between the 1980 and 2002 studies. The increase in LC records for foreign books from 22.4% to 30.8% was also noted from 1996 to 2000 in the study of Italian monographs (Kellsey 2001). The higher percentages of LC records for both English and foreign books in the 1987 study may have been due to several causes. The 1987 study included all imprint dates received during the study period, and while most of the books had recent imprints, a not insignificant number had imprint dates four or more years old, allowing more time for LC copy to appear. The 1980 study included only the previous two years' imprints. The current study did not examine imprints, but since the cataloging backlog at UCB is negligible and acquisitions emphasize current imprints, most of the items likely had imprints in the last several years. It should also be noted that while the 1987 study and current study had a comparable sample size, two-thirds of the books in the 1987 study were in English and only one-third were in foreign languages. In the current study, 46% of the books were in English and 54% were in foreign languages, so while the differences found in the two studies could be due to a genuine increase in LC cataloging in the late 1980s, which has since declined, they could be due simply to differences in collecting with concomitant variations in copy availability. Further studies using additional libraries and done at periodic intervals would be needed to settle this question.

Implications

The implications of this study intersect with several developments in the international arena. Worldwide publishing output has been increasing for many years. From 1980 to 1990 it increased 18% (Reed-Scott 1996); in the 1990s increases in book publishing continued in most European countries (UNESCO 1999). OCLC reports that for the period 1988–1994, 59.3% of foreign titles cataloged by OCLC libraries were from western Europe (Reed-Scott 1996). Obviously, the need for cataloging western European monographs is not about to disappear in the near future. In fact, the goal of the Global Resources Project, jointly sponsored by the Association of Research Libraries (ARL) and the Association of American Universities (AAU), is to increase the acquisition of unique materials from targeted areas by American university libraries. Although several of the projects deal with Asian publications, the Latin Americanist Research Resources Project targets publications in Spanish from Argentina and Mexico and the German Demonstration Project targets German language materials (Reed-Scott

Table 3. Number of Member Libraries Contributing Original Records in OCLC for European-Language Monographs, 1999–2000

No. of Records	1–5	6–19	20 +	Total No. of Libraries
French	92	16	4	112
German	100	10	6	116
Italian	46	2	0	48
Spanish	193	29	8	230

Table 4. Number of Member Libraries Upgrading Vendor Records in OCLC for European-Language Monographs, 1999–2000

No. of Records	1–5	6–19	20 +	Total No. of Libraries
French	31	6	2	39
German	33	12	3	48
Italian	35	2	4	41
Spanish	42	10	3	55

Table 5. Source of Catalog Records for English and Foreign Monographs, 1980–2002 (%)

	English		Foreign			
	LC	Member Vendor	LC	Member Vendor		
Metz and Espley (1980) N=396	77.6	[22.4]	22.4	[77.6]		
Struble and Kohlberger (1987) N=7062	84.4	15.6	47.7	52.3		
Kellsey (2002) N=8778	74.7	24.6	0.7	30.8	44.9	24.3

1996; see www.arl.org/collect/grp/grp.html for updates on these projects).

Another development involves the encouragement of international participation in the large cataloging databases. OCLC has been actively recruiting international members and contribution of records since the mid-1980s with the specific purpose of reducing duplicate cataloging and encouraging resource sharing (Brown 1992). By 2000, OCLC had participating libraries from 64 countries. With the technical advances in electronic communication in the last few years, it has become easier for libraries around the world to access and contribute to OCLC, and national libraries of several countries have joined this effort, some of them through the PCC program of the Library of Congress (Byrum 2000). These developments represent progress toward one of IFLA's stated goals, that each country should have responsibility for cataloging its own imprints (Holley 1996). At the same time, the introduction of catalog records

from non-English-speaking countries, from both libraries and vendors, has spotlighted the problems of differing standards in cataloging rules, notes in the language of the country creating the record, and the very thorny problem of lack of a universal authority file for names, corporate bodies, and uniform titles and series. Subject headings also present a challenge, since they need to be in the language of the catalog users, yet strict comparability of terms is often not possible between languages.

Although progress is being made toward developing international solutions to the problems described above, U.S. libraries still need to deal with the current reality of having to modify records for foreign language monographs before incorporating them into their local catalogs. Names and series have to be checked and modified in accordance with the U.S. authority file; notes have to be translated into English; LC subject headings have to be added; and a classification number usually needs to be added, especially if a library uses the LC classification system, since few libraries outside the United States use that system.

In the context of these developments in the larger world of publishing and cataloging, the implications of the results of the current study are troubling. Already an average of 24% of records for monographs in the four major western European languages are being entered into OCLC by vendors, with higher percentages in German and Italian. This proportion can only increase as more foreign libraries begin to add records also. Although their records may be fuller than the vendor records, they will still need modifications as noted above. At the same time, the number of U.S. libraries that upgrade records for western European language books is only about a third of the number that contribute original records for these books. The result is that more libraries are having to modify the same records locally, rather than one library upgrading a record that others can use, which is the antithesis of the goal of cooperative cataloging.

Identifying this trend is only the first step. Further studies of the practices of libraries that need to catalog western European language books would be helpful as would identifying perceived barriers or lack of incentives to upgrading records in the national databases. Exciting as the progress in international cataloging cooperation and convergence of standards is, we also need to have discussions at

all levels on the impact of this globalization of cataloging on local library cataloging practices and workload.

Works Cited

- Beall, Jeffrey. 2000. The impact of vendor records on cataloging and access in academic libraries. *Library Collections, Acquisitions, and Technical Services* 24: 229–37.
- Brown, Rowland C. W. 1992. OCLC involvement in international programs and the impact on academic libraries. In *The role of the American academic library in international programs*, edited by Bruce D. Bonta and James G. Neal. Greenwich, Conn.: Jai.
- Byrum, John D. 2000. The emerging global bibliographic network: The era of international standardization in the development of cataloging policy. In *What in the world . . . cataloging on an international scale: Papers from the ALCTS Preconference, June 26, 1998*. *Library Resources & Technical Services* 44, no. 3: 114–21.
- Casalini, Barbara. 2002. Communication at the WESS Research and Planning Committee Meeting, ALA Midwinter Meeting, Jan. 19.
- Holley, Robert P. 1996. IFLA and international standards in the area of bibliographic control. In *Cataloging and classification standards and rules*. New York: Haworth.
- Kellsey, Charlene. 2001. Trends in source of catalog records for European monographs 1996–2000. *Library Resources & Technical Services* 45, no. 3: 123–26.
- Metz, Paul, and John Espley. 1980. The availability of cataloging copy in the OCLC data base. *College & Research Libraries* 41: 430–36.
- OCLC. 2002. Communication at the OCLC Enhance Sharing Session, ALA Midwinter Meeting, Jan 19.
- Reed-Scott, Jutta. 1996. *Scholarship, research libraries, and global publishing*. Washington, D.C.: Association of Research Libraries.
- Rosenblatt, Susan. 1993. The National Coordinated Cataloging Program from the participant's perspective. *Cataloging & Classification Quarterly* 17, no. 3/4: 189–99.
- Sheddenhelm, Laura D., and Bartley A. Burk. 2001. Book vendor records in the OCLC database: Boon or bane? *Library Resources & Technical Services* 45, no. 1: 10–19.
- Struble, Carol A., and Paul B. Kohberger, Jr. 1987. Statistical survey to determine availability of cataloging copy on OCLC. *Cataloging & Classification Quarterly* 7, no. 3: 13–22.
- UNESCO. 1999. *Statistical yearbook*. Paris: UNESCO; Lanham, Md.: Berman.