Increasing Access to Archival Records in Library Online Public Access Catalogs

MATTHEW BENJAMIN GILMORE

Now that archivists are becoming more familiar with using the US MARC (United States MAchine Readable Cataloging) formats for describing archival materials, they are in a position to capitalize on those formats' capabilities for data transfer by integrating machine-readable descriptions of their holdings into wider information systems. While Thomas Hickerson's article in this issue is concerned with the use by archivists of national and regional bibliographic utilities such as the Research Libraries Information Network (RLIN) and the Online Computer Library Center (OCLC), this paper will look at the use of institutional online public access catalogs (OPACs). It will review the utility of call-number searching and will then discuss some possible archival applications. In particular, the Wallace Archives, in the Honnold Library in the Claremont Colleges, is used as an example of where bibliographic descriptions of multiformat archival materials are available online through the library catalog, and where access has been increased to those materials using subject headings, date, and format in conjunction with call- or classification-number searching.

Subject Searching Using OPACs

The problem for any reference professional, be they archivist or librarian, is to turn a "fact" question into a "which source" question—that is, to take the patron's query, a request for a fact, and to determine...
which source out of those available will best supply that information. Librarians have constructed a variety of access points to materials and their intellectual contents to make this task easier and have disseminated these points through their manual and online catalogs. Cataloged books have four chief points of intellectual access: title, author, subject heading, and call number. Before automation of the library catalog and shelflist, a searcher needed to know at least one element of these (that is, first word of the title, or subject heading, or the author's last name) and its correct form according to established cataloging rules in order to locate a desired item.

Today, when searching an online public access catalog that has keyword searching and truncation capabilities, the user need not know the entire name, title, or subject heading to retrieve relevant references. The constraints that remain, however, are still those of a controlled search vocabulary—that is, an insistence on correct or standardized spelling of titles and subject headings and a knowledge of exact name forms (particularly those derived from Library of Congress Name Authorities or established according to the Anglo-American Cataloguing Rules, 2d ed.). Common and "stop" words (articles such as "an" and "the" or prepositions such as "with" and "to") must also be avoided. A search constructed using frequently occurring terms, which generates a large response, can defeat even the most sophisticated systems and cause great frustration to the patron.

The University of California, Los Angeles (UCLA) ORION online catalog, for example, will give nothing but an "overflow" message should the subject search "United States history" be attempted, no matter in what order the words are entered. This is because the concepts used to generate the retrieval sets are too large for the computer to manipulate gracefully, and unlike the card catalog, it is almost impossible on most OPACs in operation today to browse through just a few of the references retrieved.

This is a matter of concern to archivists who have been dubious, since the first discussion about integration with library systems, of the requirement of most bibliographic information systems that archival records coded in a MARC format contain at least one Library of Congress Subject Heading (LCSH). This requirement means that archivists frequently must use a very general heading rather than the specific local thesauri or otherwise controlled vocabularies which archivists have developed to describe their own materials. In a case where archival materials are entered directly into an OPAC, and not into a specially designed archival subsystem, the vague nature of many collection titles
and the general subject headings may result in those materials disappearing into a void; they could be lost in a computerized catalog that can only be searched by traditional access points.

More intellectual access points must be made available. As patrons and staff are gradually becoming aware, particularly through their contact with various types of automated retrieval systems, automation requires different and alternative strategies for subject searching from those used when performing a manual search. The most sophisticated OPAC users, such as reference librarians, will frequently use a rough "subject" search by using keywords in a search for a hypothetical title; their assumption is that the title will correspond with the subject. The full bibliographic records for titles retrieved this way (excluding the inevitable "false drops") will indicate the approved subject headings assigned by the cataloger. These headings can then be used in a real subject search option to locate more material. The rough search also reveals the classification numbers which predominate for the subject. Reference librarians will often send a patron to the stacks or the shelflist with the injunction that more books on the same subject can be found with similar call numbers on the shelves near the title(s) already located in the catalog. At UCLA librarians can demonstrate this feature to the patron by searching by call number in the ORION online catalog, although this option is not yet available on patron terminals.

Call Number Searching in OPACs

Call number access has led some proponents of online catalogs to conclude that call number searching would be useful to the patrons as well as to the librarian. Karen Markey, working with the Dewey Decimal Classification (DDC) scheme, emphasizes the implementation of automated classification number searching as "a major source of unique subject-rich terms that enhance subject access to bibliographic records." One of her major recommendations was to mount the entire DDC schedule online so that the patron's subject could be matched against the classification number used in the schedule. In summarizing her research, Markey found that subject searching was by far the most common use of the online catalog, even in those catalogs lacking formal capabilities to perform such searches. An alternate means of facilitating subject access is the use of the classification number. However, user ignorance of the subject significance of call numbers must be overcome. Most patrons simply see the call number (either library or archival) as a location indicator and have no idea that it might also pertain to the subject.
To rectify user ignorance, user aids should be created, probably in a thesaurus format which would clearly link a subject with its classification number. If the guides were available online, a classification number could be generated in response to the patron input of any subject term; Markey suggested this solution in 1983. For example, when the patron searches for “Islam-addresses, essays, lectures,” the OPAC would respond with the number of titles retrieved accompanied by the truncated call number DT4.5

Nancy Williamson and Lois Mai Chan have both studied the Library of Congress Classification (LCC) scheme as a tool for online searching. Williamson suggests that implementation of a thesaurus format is the direction that research should take in order to allow user exploration of hierarchical construction of the classification.6 Chan points out that the enumerative nature of LCC is an advantage in online searching since each number contains rather specific subject information.7 An advantage that LCC numbers have over DDC numbers is that they are short and, therefore, are usually applied similarly in a variety of libraries. This leads to little variation from library to library, as opposed to Dewey numbers which are often truncated.

Both Williamson and Chan favor mounting the entire classification scheme onto a subsidiary database, rather than relying simply on the call numbers of the records present in an existing online catalog. A similar idea is the development of a subsidiary database, dedicated to locally or nationally established authority files, which can be searched independently of the archival records (this was undertaken by Geac in the Smithsonian Institution archival subsystem).

An advantage to searching with LCC is that one has greater search precision—and therefore, fewer “overflow” responses—than by using the regular subject search option. This is particularly important when dealing, not with a localized archival database containing perhaps a few thousand records, but with a system-wide database of online information which might contain several million records. For example, searching records E178.A1 through E178.3.Z9 to retrieve references to materials on U.S. history is feasible whereas searching the subject term “United States history” is impossible. Of course, numerical searches would be even more useful if the retrieval set could be narrowed by specifying archival materials only or archival classification numbers only. Call number searches are actually more accurate than title “subject” searching because there should be no false drops unless an item has been misclassified by the cataloger.
Janet Swan Hill highlights several drawbacks to online classification searching, however. She notes changes in DDC over the years, local inconsistency in schedule application, or local multiplicity of systems (for example, a mixture of local, DDC, and LCC), and the failure of the call number and subject to correspond due to misclassification. The capability to browse the shelves from a computer terminal is a compensating advantage which would benefit reference archivists in those instances where shelving corresponds to classification number. (This is not always the case, however, and additionally, most archives do not permit patrons into their stack areas.) Like Chan, Hill also recognizes the rich subject content of the two classification systems, particularly that of LCC which has schedules occupying six linear feet.

Archival classification schemes are usually locally devised, based on provenance, and hierarchical in nature. Some sophisticated schemes specify four or more levels of detail—form division, section, and office, to media and format. These schemes would provide a useful source of information about archival holdings in an OPAC where some of the other MARC fields, which are directly searchable on local archival systems (such as media or format), would not be supported or available.

Online Classification and Reference Coding Retrieval

Despite the active research in the field of librarianship, little attention has been paid to applying classification- or reference-coding searching to automated records of archival materials. Many, if not most archives, of course, are not yet online with a local OPAC, and many do not use what they would consider to be classification numbers. At UCLA, for example, the existing classification scheme in the University Archives grew unwieldy through neglect and it is no longer used. Instead, a unique accession number is assigned to each series. However, integration and support of the library OPAC can be very desirable, particularly for small archival collections that fall under the part-time control of a librarian. Any mnemonic shelf location reference coding or classification system could be implemented online as an additional access point, providing records were in a MARC format acceptable to the OPAC and some method was used to control record length. In this, the Wallace Archives can serve as a model.

Background

In 1983, Irving Wallace donated his personal papers and collection of publications to the Claremont Colleges to be housed in Honnold...
Library Special Collections. Wallace, a best-selling novelist, and author of such works as *The Fabulous Originals, The Sins of Philip Fleming, The Chapman Report, The Man, The Plot, The Prize,* and *The R Document,* also coauthored with his wife Sylvia, son David Wallechinsky, and daughter Amy, *The Book of Lists* series and *People's Almanac* series, and until 1953 wrote short fiction and nonfiction articles which were published in national magazines. In that career, which began in 1931 with the precocious sale of his first article (Wallace was 17), he conducted many interviews with celebrities including Alexander Kerensky, Marlene Dietrich, and Pablo Picasso. Many of his books have been adapted to the screen or for television, most recently *The R Document.*

In his agreement with the Claremont Colleges, Wallace donated all drafts and notes of approximately twenty-five books written by him, all correspondence, contracts, business records, review files, corrected galleys (foreign and domestic), all original research materials, and copies of all the books in all the languages in which they were printed. Also included are research materials and manuscripts of some 500 short stories, magazine articles, interviews, and approximately seventy-five videotaped interviews with Wallace, some running to two hours in length. At present, the growing collection consists of approximately 3000 unit cataloging descriptions, ranging from single items to boxes of research material, envelopes of photographs, and groups of videocassettes.

To enable Wallace to have access to his own papers, the library cataloged them on a subsidiary database of the library's Total Library System (TLS) online catalog. Since the donation included a wide range of materials from pre-print to published works, including outlines, manuscript drafts, galleys, correspondence, private journals, contracts, review files, books, articles, screenplays, promotional items, and audio- and videotapes, all by Wallace and his family, and many with the same title, not only was a classification scheme necessary, but so, too, was a format designation.

**Classification Scheme**

Fifteen categories of material were devised by Honnold Library catalogers, each with a two-character mnemonic code (see fig. 1). This collection is organized around Wallace's books, the principle being collocation of all material relating to each published work, no matter what format. Original order is not respected, but provenance is maintained. Each book is given a code number, consisting of the first
Figure 1. Fifteen Categories of Material (Source: Honnold Library. “Irving Wallace Collection Classification Scheme.” Unpublished, 12 April 1983, p. 4).

two letters of the first significant word and a number representing the third letter. This code will allow for the insertion of new titles with the same first two letters of the first word (see fig. 2).

A character was assigned to each of the Wallace family authors: A for Amy Wallace, D for David Wallechinsky, I for Irving Wallace, and S for Sylvia Wallace. W was assigned to collective projects. Together with the dates and the W location code for the Wallace Archives, those three

AL5  The Almighty
CH3  The Chapman Report
FA2  The Fabulous Originals
FA5  The Fan Club
MA5  The Man
NY5  The Nympho and Other Maniacs
P14  The Pigeon Project
PL6  The Plot
PR5  The Prize
RD2  The R Document
SE2  The Second Lady
SE8  The Seven Minutes

Figure 2. Codes Assigned to Each Book (Source: Honnold Library. “Irving Wallace Collection Classification Scheme.” Unpublished, 12 April 1983, p. 4).
elements—format designation, title code, and author character—make up the call number equivalent. This is accepted by Claremont Colleges Total Library System as one long call number.

Each book that the TLS receives has either an LCC or DDC call number assigned using OCLC. As the Claremont Colleges system has several libraries and a multitude of locations in each, and as the computer does not allow duplicate numbers, a location designation is necessary to make distinctions between like or multiformat materials. This distinction takes the form of a five character prefix code (see fig. 3).

The initial character stands for the library, in this case Honnold. The Wallace Archives uses W instead of H. The second character stands for the location in the library, such as S for Special Collections. This is left blank for Wallace. The third is a format designation—"." for regular size, "-*" for folio, "-*" for reference folio, "-]" for double oversize, and "-I" for miniature. Wallace cataloging fills this and the next space with the two character format mnemonics described earlier. The fourth and fifth positions are used for a third level of specificity for collections in certain locations in each library, for example OX for the Oxford Collection in Honnold Library Special Collections.

The Wallace materials are cataloged at the item level, except for the background research materials for the books which are described at collection level. The MARC Books (monographic) format was used since the AMC format was not available when cataloging was started and the extra fields of the AMC format were not needed when the format was finally issued. Each item cataloged is given a unique bar code, which is a system requirement for the original circulation system. Figure 4 shows some sample records from the Wallace TLS Database.

**Retrieval**

The purpose-designed, segmented call numbers designed for the Wallace Archives allow for a variety of searches, from the broad to the

<table>
<thead>
<tr>
<th>Location Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H . DA 400.A5 B5</td>
<td>Honnold Library Open Stacks</td>
</tr>
<tr>
<td>HS.OXDA 400.A5 B5</td>
<td>Honnold Library Special Collections</td>
</tr>
<tr>
<td></td>
<td>Oxford Collection</td>
</tr>
<tr>
<td>HS-OXDA 400.A5 B5</td>
<td>Honnold Library Special Collections</td>
</tr>
<tr>
<td></td>
<td>Oxford Collection</td>
</tr>
<tr>
<td></td>
<td>folio</td>
</tr>
</tbody>
</table>

Figure 3. The Five Character Prefix Code (Source: Total Library System database)
Library Online Public Access Catalogs

precise. Particular books can be located in the various stages of production, beginning with research material all the way through to screen adaptation. Classification number searching is particularly helpful in the case of the Wallace Archives since, with almost all the material being authored by a member of the Wallace family, an author search would be too broad. It also provides a solution to the difficulty presented by the many manifestations of each work. Instead of having to search through the works of each Wallace or through every format of a title, specific aspects can be readily retrieved. Figure 5 shows how the searcher can retrieve on the format designation and on the book title code. Searching is possible in a shorter form, without the location/format designation, and such a search would retrieve everything with the Cutter number input, regardless of format (see fig. 6).

Implications

While virtually all the Wallace materials are cataloged item by item using the MARC Books format, the same principles can be extended to those described at the collection level using MARC AMC format, if a mnemonic or otherwise meaningful reference code is assigned to the collection. The searcher could begin, as do most patrons, by searching for known subjects or names in the OPAC. Retrieval would then reveal a reference code which could be used to search the collections; the process should emulate a patron reading the classification scheme for record groups or the summary record series description, or browsing the shelves. This capability would be enhanced if the collection were cataloged at series and box level (where each box contains discrete materials), so that the patron could browse box descriptions from the terminal.

Richard Lytle has shown in his work that traditional cataloging methods in archives (using either provenance or content indexing) have no significant affect on retrieval success. His conclusion “that both methods performed poorly was drawn by measuring overlap, the degree to which both methods found the same relevant folders in the collections; in other words, it is likely that many relevant folders were not retrieved. Another way of stating this is that neither method is very consistent or reliable, as measured by the overlap.” It is possible that implementing online access by classification scheme as an additional indexing method might help to improve this low rate of retrieval.

By taking advantage of the research done by scholars like Karen Markey, Nancy Williamson, Lois Chan, Janet Swan Hill, and Elaine Svenonius, archivists can add another significant, powerful, point of
Sample Records

W MSILE 8 1943A [W=Wallace, MS=Manuscripts, I=Irving, LE=Letter..., 1943A=date]
Wallace, Irving.
A letter from Australia [typescript] by Irving Wallace.
1943.
7 leaves; 28 cm.
Photocopy.
World War, 1939-1945
Football
1943
Foreign Service
barcode:10010827339

W MSICO 5 1945 [W=Wallace, MS=Manuscripts, I=Irving, CO=Columbus..., 1945=date]
Wallace, Irving.
Columbus was wrong [typescript] by Irving Wallace.
1945.
21 leaves; 28 cm.
Photocopy.
"I finished this article on Wilbur Glenn Voliva and Zion City on Oct. 11, 1945. It finally sold to True magazine where it appeared combined as 'When Zion blew its top'—April, 1946 issue. However, I used most of it with revisions in chapter I of my second published book, The Square Pegs, published in 1957 by Alfred A. Knopf."—I.W., Nov. 12, 1984.
Voliva, Wilbur Glenn.
Zion City, Illinois.
1945
True
Square Pegs
Alfred A. Knopf
barcode:10010827338

W FRIOD 4 1936 [W=Wallace, FR=film related, I=Irving, OD=Oddity..., 1936=date]
Wallace, Irving.
Oddity Hunter [film treatment] by Irving Wallace.
1936.
6 leaves; 28 cm.

Figure 4. Sample Records (Source: Total Library System database) (cont. on page 619)
Carbon copy.
"Only existing carbon copy of retyped final version of my first movie original story, finished in February, 1936."—I.W.
Curiosities and wonders
William Morris Agency.
barcode: 10010827337

W ZZIOB 4 1972 [W=Walace, ZZ=Miscellany, I=Irving, OB4=Obituaries, 1972=date]
[Obituaries] [Clippings]
19 items: ports.; 17-36 cm.
"Obituaries of friends and writing or publishing associates."—I.W., June 8, 1985.
Clippings of obituaries of IW's friends in publishing, writing, film, etc.
Obituaries.
barcode: 10010827341

Figure 4. Sample Records (Source: Total Library System database)

access to their online databases. Reference codes, based on an archive organized according to provenance and revealing the structure of records arrangement, would allow patrons to scan the related record series as if they were searching inventories. Format designations could be helpful in retrieving archival material, particularly if applied within a collection—for example to a box of photographs in a collection of personal papers. In fact, the formats designated for the Wallace Archives could be extended to include any type of format likely to occur, such as scrapbooks, photograph albums, and any kind of realia.

The AMC format has a genre/format field 655, as well as a field for classification numbers, and, if it proves feasible on a local system, archivists could use these as searchable fields in the holdings database, allowing for more precise searches and easier retrieval of materials. A survey of MARC AMC users indicated that all of those questioned already use the 655 field in conjunction with various thesauri, including LCSH. The survey responses also indicated that this helped with specialized retrievals, for example, locating materials for an exhibit. Given the facts that these data are already in existence for collections coded in MARC AMC, that many archives have well-defined classification systems which they use with a variety of in-house manual and
Sample Searches

The user inputs the TLS code for location searching, and the computer responds with the prompt:

LOCN.SS++++.++ C99 -------- AA.

The user then inputs the Wallace call number of whatever fraction would be useful.

• Searching using location code (just the five characters)
  LOCN.SS++++.++ C99 -------- AA (the computer prompt)
  W CRA
  retrieves all critical material (review) (CR) on Amy Wallace (A)

  LOCN.SS++++.++ C99 -------- AA
  W CSD
  retrieves all correspondence (CS) of David Wallechinsky (D)

  LOCN.SS++++.++ C99 -------- AA
  W PPI
  retrieves all preparatory prints (PP) for books by Irving Wallace (I)

  LOCN.SS++++.++ C99 -------- AA
  W BGI
  retrieves research (background) material (BG) used by Irving Wallace (I)

• Searching using location code and cutter (five characters and book code)
  LOCN.SS++++.++ C99 -------- AA
  W BGICH 3
  retrieves all background material (BG) on The Chapman Report (CH3)

  LOCN.SS++++.++ C99 -------- AA
  W MSSFO
  retrieves all manuscripts (MS) of Sylvia Wallace’s (S) The Fountains (FO)

  LOCN.SS++++.++ C99 -------- AA
  W PRIBL 2
  retrieves all promotional material (PR) on The Book of Lists 2 (BL2)

Figure 5. Sample Searches (Source: Total library system database)

The user inputs the TLS code for call number searching, CC, and the computer responds with the prompt:


Sample search

• Searching using the book code (without the location code)
  SS++++.++C9
  AL 6
  retrieves all material on Wallace’s The Almighty, whatever the format.

Source: TLS Wallace database.

Figure 6. Source: Total Library System database
automated systems, and that archivists are now looking increasingly to integrate information on their holdings into local online catalogs, the argument for implementing classification access is a persuasive one.

ACKNOWLEDGMENT

The author acknowledges the help of those UCLA Graduate School of Library and Information Science faculty who read and commented on this paper.

References


MATTHEW GILMORE


Additional References


