Case Study

SUSAN E. DAVIS, editor

American Medical Association's Historical Health Fraud and Alternative Medicine Collection: An Integrated Approach to Automated Collection Description

JAMES G. CARSON

Abstract: From 1913 to 1975, the American Medical Association's Department of Investigation assembled more than 300 cubic feet of files on health fraud, quackery, "patent" medicines, and alternative medicine. In 1988, the AMA obtained a grant from the National Library of Medicine to process and catalog these materials, now known as the Historical Health Fraud and Alternative Medicine Collection. Using Minaret software (a stand-alone USMARC AMC cataloging system) in combination with WordPerfect word-processing software, the project staff developed procedures that allowed it to generate textual and index entries for the printed guide to the collection as well as upload USMARC AMC records directly to the OCLC (Online Computer Library Center) union catalog.

Background of the Project

THE AMERICAN MEDICAL ASSOCIA-TION'S Historical Health Fraud and Alternative Medicine Collection (hereafter referred to simply as the Historical Health Fraud Collection) consists of more than 300 cubic feet of files on health fraud, quackery, "patent" medicines, and alternative medicine. The collection originated as the office files of the AMA's Department of Investigation, which existed from 1913 to 1975 and was charged with answering inquiries about fraud, quackery, and alternative medicine.

A combination of factors led to the abolition of the Department of Investigation in 1975. By that time, government agencies such as the Food and Drug Administration and the Federal Trade Commission were largely duplicating the department's investigative functions. The AMA Library accepted responsibility for both the records and the information-dispensing function. The library gradually phased out active gathering of new information on health fraud and questionable therapies as private organizations such as the National Council Against Health Fraud moved into this arena.[1]

The Department of Investigation's files, though no longer active, contained an unparalleled wealth of original source material on thousands of fraudulent or alternative health practitioners, products, and practices that the department investigated during its sixty-two years of existence. Recognizing the unique historical value of these files, in 1988 the AMA Division of Library and Information Management applied for and received a two-year, $165,000 grant from the National Library of Medicine to process the collection, describe it

in MARC Archival and Manuscripts Control (AMC) format, and produce a collection guide.[2]

The grant funds permitted the AMA library to develop automated procedures that integrated what are typically related-but-separate operations. In order to provide wide access to information about the collection through a nationwide bibliographic utility, USMARC AMC records were to be added to OCLC (Online Computer Library Center), the national network with which the AMA Library is affiliated. This objective could be combined with that of efficiently producing a conventional guide to the collection, complete with indexes, and a local searchable database containing more detailed data than could appropriately be entered in OCLC. The procedures described here could be adapted for use by other archives, whether or not the same software and bibliographic utility described here are involved.

Creating USMARC AMC Records

In archival terms, the AMA's Historical Health Fraud Collection is an alphabetical subject file that constitutes a single large record series. Within this series, holdings range from single folders, in the case of many minor topics, to several cubic feet on topics of great interest, such as claimed cures for alcoholism, cancer, and obesity. This variation in depth of coverage within the collection leads to corresponding variation in the descriptive approach. The project staff created a master collection-level catalog record, supplementing it with separate AMC records for each major subseries, i.e., holdings on a single subject. Subseries of sufficient size and complexity were additionally described by a folder list (which is not a part of an AMC record). Because the

---

[1]The National Council Against Health Fraud, Inc., P.O. Box 1276, Loma Linda, CA 92354. Its resource center is located at Trinity Lutheran Hospital, 3030 Baltimore, Kansas City, MO 64108.

[2]Department of Health and Human Services, Public Health Service, National Institutes of Health, Resource Grant G08 LM04637, Arthur W. Hafner, Principal Investigator.

collection contained more than 3,000 sub-series, as many as four alphabetically adjacent minor subseries were combined into a single record. This effort reduced the number of MARC records to approximately 950.

Original plans for the Historical Health Fraud Collection project called for AMC records to be created directly in OCLC.[3] This approach was ruled out early in the project for several reasons. At the time, OCLC lacked subject-searching capability, and searching the OCLC database for in-house reference purposes would have involved cumulatively expensive connect-time charges.[4] The most serious problem, however, related to authority control.

Bibliographic networks such as OCLC naturally and legitimately require prior verification of personal and corporate names and other headings used as access points in catalog records; this verification is carried out using appropriate national authority databases such as the Library of Congress name authority file. However, in archival contexts this process can become a black hole into which mountains of work-time disappear to small discernable purpose. Very few of the names encountered in a typical archival collection are those of published authors or other similarly prominent entities. Hence they are unlikely to be found in the relevant authority files.[5] However, the use of many such names, and of other headings such as names of medicines and similar products, is highly desirable to provide access points for local searching of the collection.

Fortunately, this dilemma arose just as the first personal-computer-based AMC systems, Minaret and MicroMARC:amc, were

becoming generally available. The AMA Library eventually decided to install Cactus Software's Minaret system on the project's OCLC workstation, an AT-class personal computer.[6] This permitted project staff to do the original cataloging in Minaret and then transfer the records to OCLC. This decision allowed the development of a customized database configuration that includes both standard MARC and local versions of each of the USMARC AMC subject added entry (6xx) fields, as well as a special local field for product names. The OCLC version of a catalog record includes only fields with standard MARC tag numbers and omits the corresponding local fields. Cataloging staff verifies terms used in the standard 6xx fields in the relevant authority files, namely LC name authority and National Library of Medicine subject headings. Terms used in the local 6xx fields are subject only to a much more streamlined local authority-control system built into Minaret. In general, the local fields are preferred except for persons or other entities that appear to be of sufficient prominence to justify their inclusion in a national database.[7]

In conjunction with word-processing and other auxiliary software, Minaret has become the heart of an integrated automation approach that uses only two inputting procedures to generate five different products (see Figure 1). The processing staff enters AMC catalog records in Minaret and cre-

---

[3]The Historical Health Fraud Collection is the only portion of the AMA Archives to be cataloged in OCLC.

[4]OCLC's recently introduced EPIC service has filled this gap.

[5]On a previous project in which the author participated, the hit rate for authority searching in a similar context was about 2 percent.

[6]Cactus Software, Inc., 15 Kary Way, Morristown, NJ 07960-5604. Among the factors favoring Minaret were its built-in authority control routine, variety of inputting-form options, flexibility in formatting output, and automatic index updating.
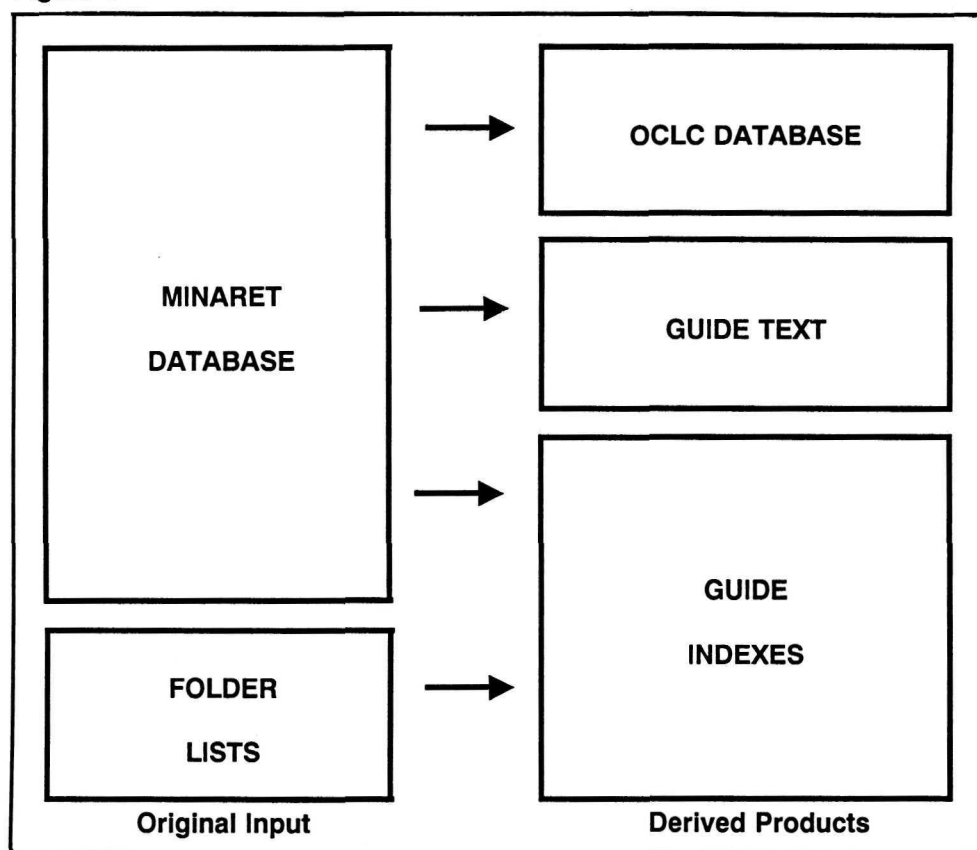
[7]This strategy can complicate local searching on the Minaret database, because a searcher may not know whether a given search term appears in the standard MARC version or the local version of a given field. But this is not a serious difficulty, for Minaret's free-form search editor allows searches with Boolean operators involving multiple fields. Formulating a free-form search can present problems for a computer novice, but we have streamlined the process by using SuperKey, a RAM-resident utility, to create search macros that take care of all the necessary keystrokes except for the search term itself.

Figure 1

ates folder listings with WordPerfect as the collection is processed. The Minaret records are then manipulated to produce OCLC catalog records. With the help of the search-and-replace and other editing conveniences of WordPerfect, they also yield textual and index entries for the collection guide. Index entries are also drawn from the folder lists.

Producing a Guide from Minaret Records

Text conversion. The process for producing guide text entries takes advantage of Minaret's form-editor feature. A Minaret form is, in effect, a template through which catalog records are viewed. The guide-entry form includes only those US-MARC AMC data elements that appear in the collection's printed guide: record title, dates, extent, call number, and note fields.[8] Once this form is invoked in Minaret, the operator creates an export file and then transfers it to WordPerfect. In Word-Perfect, search-and-replace macros (stored instructions that simplify the repetitive re-

---

[8] "Call numbers" for the collection are simply inclusive box/folder numbers. For example, 0106-07/0107-03 means that the materials described are to be found beginning in folder 7 of box 106 and ending with folder 3 in box 107.

placement of one text string or formatting code with another) perform a number of editing functions, the most notable of which exploits WordPerfect's paragraph numbering feature to assign serial numbers to the entries. (See Figures 2a and 2b.)

Figure 2a

```
035      ‡aAMC89-000140 ‡bAMA
040      ‡aAMA ‡eappm ‡cAMA
099   9  ‡a0113-03/0116-01
049      ‡aAMAF
110  2   ‡aAmerican Medical Association,
‡bDept. of Investigation.
245 00   ‡aRecords. ‡pCathartics, ‡f1904-
1973.
300      ‡a1.0 cubic ft. (3 boxes).
520      ‡aCorrespondence, reports, adver-
tisements, articles and clippings, press re-
leases, and promotional and supplementary
materials concerning cathartics. ‡bThere are
six folders of material on cathartics in gen-
eral. The rest concern individual cathartics,
mostly patent medicines, but also quack de-
vices such as the "Sphincter Muscle Expan-
der." Among the more prominent cathartics
are "Cereal Meal," Phillip's Milk of Mag-
nesia, and Zo-Ro-Lo.
555 0    ‡aA folder list is available for this
material.
```

Portion of a USMARC AMC catalog record as it appears in Minaret.

Figure 2b

```
125.     Cathartics, 1904-1973.
1.0 cubic ft. (3 boxes).
Call number: 0113-03/0116-01
SUMMARY: Correspondence, reports, ad-
vertisements, articles and clippings, press
releases, and promotional and supplemen-
tary materials concerning cathartics. There
are six folders of material on cathartics in
general. The rest concern individual cathar-
tics, mostly patent medicines, but also quack
devices such as the "Sphincter Muscle Ex-
pander." Among the more prominent ca-
thartics are "Cereal Meal," Phillip's Milk
of Magnesia, and Zo-Ro-Lo.
A folder list is available for this material.
```

Collection guide entry, derived from record in Fig. 2a.

Building an index. The procedure for deriving index entries from Minaret catalog records also involves the Minaret form editor. For this purpose, project staff have defined a form that contains only the call number and the subject added entry (6xx) fields. In this form the tag numbers are replaced by two-letter mnemonic codes, e.g., "PN" for a personal name. Again, an export file created with this form is transferred to WordPerfect.[9] There the serial number of the corresponding guide entry replaces the call number, and a macro appends this serial number to each index entry. Next, another set of macros appends each entry to one of seven index files, depending on the index code that precedes it.[10] In the final step, WordPerfect sorts the index files alphabetically to move the new entries into proper sequence.

The system also derives index entries from folder lists originally entered in Word-Perfect.[11] In addition to columns for folder title, dates, and box/folder number, the folder-list format includes a column for index codes. A processor will enter the appropriate two-letter code in this column whenever a folder title is suitable for inclusion in one of the seven indexes—for example, when it comprises the name of a person, corporate body, or product. Another series of macros then strips the folder list down to include only folder titles and call numbers, thus corresponding to the

---

[9] Unlike the guide-text conversion routine described in the previous section, this procedure must be performed separately for each catalog record. To streamline it as much as possible, all the keystrokes needed to generate the export file are stored as a SuperKey macro.

[10] There are indexes of personal, corporate, conference/meeting, geographic, and product names; titles; and topical subjects.

[11] This procedure would be unnecessary if every appropriate heading appearing in a folder list were also incorporated as an added entry in the corresponding catalog record. The decision not to follow this practice was largely a concession to time constraints and may be reconsidered in the future.

output from the Minaret index form.[12] From this point on, the procedure is exactly the same as for the 6xx AMC fields. (See Figures 3a, 3b, and 3c.)

The result of these steps (which, once the procedures and macros are established, are employed far more routinely than their description here may convey) is a guide to the collection that provides a clear description of the collection (including indexes) in an effective, recognizable format. The same data entry also produces a locally searchable Minaret database which can be accessed in a wide variety of ways—even by minimally trained personnel or by researchers themselves, using the search macros described in note 7. As described in the next section, it also produces records to be added to a national database.

Uploading Minaret records to OCLC

The purchase of the Minaret system for the project did involve one major uncertainty. While Minaret produced records that conformed to the OCLC implementation of the AMC format and were hence OCLC-compatible, it did not originally have the capacity to upload records directly to OCLC. At this stage, it would have been necessary to export records onto tape and then send the tape to OCLC—a cumbersome procedure that would have added significantly to the project's expenses. The solution was to develop a direct upload protocol using a modem, which would obviate the necessity for tape uploading.

Eventually, following a few false starts and several discussions with OCLC personnel, the author and Cactus Software president Geoffrey Mottram developed a procedure based on one previously devised by Richard Aroksaar and Ellen Traxel of the Pacific Northwest Regional Library, National Park Service.[13] The initial version of the OCLC upload routine worked as follows:

The first step was to strip the local fields out of the records to be uploaded; this was accomplished by exporting them to a separate database within Minaret. Staff then transferred this record set to WordPerfect, where search-and-replace macros rectified some minor format differences between

---

[12]This process is actually performed on a copy of the folder list; the original folder list is naturally retained (on disk as well as in hard copy).

[13]The original procedure is described by Aroksaar and Traxel in *OCLC Micro 5* (June 1989): 9-11. The American Medical Association's adaptation was described briefly by Marion Matters in the *SAA Newsletter*, March 1990, 11.

Figure 3a

692	‡aCeremel.
692	‡aCitrolax.
692	‡aCream of Magnesia.

Product-name added entries, derived from US-MARC AMC catalog record shown in Fig. 2a.

Figure 3b

Index	Folder Title	Date(s)	Folder No.
CN	Cerag Company	1916	0114-04
PR	Colonaid	1957-1960	0114-05
PR	Correctol	1958-1959	0114-06
PR	Cryst-L-Dex	1936-1939	0114-07
PR	Dorsey's Mixture	n.d.	0114-08
CN	Druggists Cooperative Association	1913-1917	0114-09
PR	Dunbar's System Tonic	1913-1937	0114-10

Portion of corresponding folder list showing product-name entries (index code "PR").

Figure 3c

Ceremel. 125
Chamberlain's Colic Remedy. 129
Chase's Kidney Pills. 130
Citrolax. 125
Citrophan. 148
Clarke's Blood Mixture. 224
Collum Dropsy Remedy. 158
Colonaid. 125
Connelley Liquor Cure. 14
Correctol. 125
Cosmic Wave Vitalizer. 173
Cream of Magnesia. 125
Crotalin. 181
Cryst-L-Dex. 125
Cystex. 184

Portion of product-name index, showing inter-mixed entries from AMC record and folder list.

Minaret and OCLC. For example, Minaret requires subfield delimiters at the beginnings of all subfields; in OCLC, the ‡a delimiter is omitted where subfield *a* is the first subfield in a field. Hence, one of the search-and-replace macros stripped out ‡a delimiters that occurred at beginnings of fields.

After these modifications, the resulting file went through a routine that transformed it into a script that could be read by ProComm communications software and transmitted via modem to OCLC. Aroksaar originally developed the "transcat" utility file which performs this transformation; the utility is available via the Fedlink bulletin board ALIX.[14] Since this utility was originally designed for use in a book-oriented environment, its output required some modification for archival purposes, notably by replacing the books-format workform command with the appropriate workform command for the AMC format.[15] Thus, in-

stead of going directly into ProComm, the "transcat" output file was first loaded into WordPerfect again, where another series of search-and-replace macros replaced the workform commands and made other necessary changes. Staff then manually inserted appropriate passwords and identification numbers, and transferred the file to ProComm, which transmitted it to OCLC.

Recently, drawing on the Health Fraud project's experience, Cactus has added an upgraded version of the OCLC upload utility for Minaret that completely eliminates the need for auxiliary massaging in a word processor. This latest upload utility includes a special upload form and a DOS utility, "mkscript," that incorporates the "transcat" routine. These features accomplish the stripping of local fields, elimination of superfluous subfield delimiters, substitution of AMC workform commands, and all other necessary changes. User passwords and identification numbers need only be inserted once in an auxiliary text file; the utility then includes them automatically in each output file produced by the "mkscript" routine. This output file is then loaded directly into ProComm and transmitted. What OCLC "sees" during this process is cataloging text being entered at "home" position on the workstation screen, one line at a time. The process takes from 60 to 90 seconds for an average record containing between twenty and thirty fields.

The version of the upload utility used on the Health Fraud project places catalog records in the OCLC "save" file, from which project staff then retrieve and "produce" them in a separate manual operation. This is the final step that actually places a record in the OCLC online union catalog and assigns it a serial number. The ProComm script

[14]ALIX can be dialed at 202-707-9656; the "transcat" routine is in files area #3, files section.

[15]The process of entering an OCLC record always begins by calling up the appropriate workform for the MARC format desired. The workform includes prompts for required fields and others that are commonly used.

Although data can be entered at the "home" position on the screen, rather than on the workform, the latter must still be present.

could incorporate this step; however, problems such as undiscovered typographical errors and communication difficulties during the uploading session may result in the necessity for last-minute changes. It is easier to make these changes in the "save" file than in a record which has already been "produced."

Comment

It is difficult to quantify the impact of these automated procedures on the AMA's Historical Health Fraud Collection project.

However, a reasonable estimate is that it would have taken the project staff at least twice as long to create OCLC catalog records, guide text, and guide indexes manually. More likely, of course, these helpful additional finding aids would never have been developed. Thus, the Historical Health Fraud Collection project's automated procedures have saved roughly four person-years of work, and can serve as a useful model for other repositories in using automation to improve collection access with minimal descriptive effort.