APPLYING DATABASE MANAGEMENT CONCEPTS TO BIBLIOGRAPHIC RECORDS

PETER SIMMONS
School of Library, Archival and Information Studies
University of British Columbia
Vancouver
Canada.

The concept of handling bibliographic records in the form of a coherent group of designated elements has never been applied completely. The creation of data elements known as record segments, with links between segments and between records, permits more flexibility in computer manipulation of bibliographic records.

INTRODUCTION

The appearance in 1966 of the first centrally distributed bibliographic records in machine-readable form introduced librarians to a number of new bibliographic concepts. The most important were these:

(a) the decomposition of bibliographic records into their most basic elements;

(b) the definition of new terms to describe the containers for these elements: fixed and variable length fields, subfields, indicators, and special codes; and

(c) a structure to organize all of these elements into a complete format.

A bibliographic format is the syntax of the language of bibliography. With human-readable formats, understanding the syntax is largely dependent upon correctly interpreting the significance of the order or arrangement of the elements. There are few connectors other than punctuation to help make the combination of elements intelligible, and the structure is a very condensed one. Nevertheless, professionals who handle bibliographic data have always been skilled at manipulating these elements, often without giving the matter much thought. Similarly, experienced users of catalogues and bibliographies are rarely conscious of any special syntax in a bibliographic citation because they have become familiar with the limited number of forms commonly found in the bibliographic works they consult.

BIBLIOGRAPHIC DATA ELEMENTS

The introduction of machine-readable bibliographic records required as a departure from the implicit identification of general areas of description by means of layout and punctuation to the explicit coding, or "tagging", of the basic elements, as well as a significant increase in the use of codes to represent data. Librarians and other information specialists quickly adapted themselves to using formatted bibliographic records in machine-readable forms, and a host of new formats were (and are still being) developed to suit various kinds of information systems.

Since that introduction of standardized computer-manipulable bibliographic records more than 20 years ago, librarians and other information specialists have exploited the potential of these records as the basis for many kinds of automated library applications. Yet the bibliographic records we use are not entirely well suited to their role as the foundation of bibliographic files.

Manual bibliographic files fall into several categories. The simplest are enumerative bibliographies, which probably come closest to a printed version of a computer file of bibliographic records. The most sophisticated and complex are library catalogues and certain kinds of printed indexes, their complexity derives from their ability to contain and display a number of relationships between bibliographic items. These relationships present special problems for information handlers: often they are not well supported, and sometimes not even permitted, by the standard formats commonly used for machine readable records.
FORMATTED BIBLIOGRAPHIC RECORDS

Two example will be used to illustrate this point. Figure 1 shows a typical formatted bibliographic record for what many would consider a single bibliographic item. The real record would consist of a series of fields containing subfields, many of which have been omitted from this illustration. For ease of comparison, the data elements are shown preceded by their numeric tags, the actual record would not necessarily be stored or displayed in this form. Examination of this record reveals that the sequence of the fields follows the numeric value of the field tags. The value of the field tags is designed to cause the fields to fall into the order in which they would appear in a displayed or printed version of this record in a library catalogue. This close connection between the machine record and an entry in a library catalogue is also evident in the use of punctuation. The marks of punctuation concluding the title statement (field 245, subfield @a) and the place of publication (field 260, subfield @a) are designed to provide the display format commonly used for catalogue entries, while the series statement (field 440, subfield @a) lacks final punctuation because the user’s software is supposed to provide enclosing parentheses when displaying the record.

If the format used for this example is formulated in accordance with ISO 2709 [1], the bibliographic level code of the record, which will appear as character 7 in the record label, will be “m” for monograph. However, field 440 shows a title which is not that of the monograph, but instead the title of a series to which the monograph belongs. In fact, it is a serial title appearing in a monograph record. Our example contains a hierarchy. While most of the fields describe the level indicated by the bibliographic level code, this field is at a higher level. Field 505, on the other hand, gives a list of component parts contained in the item, this field exists at a lower hierarchical level than a monograph.

We see that this record, in common with many bibliographic records, contains data elements describing several bibliographic entities which are related to one another. For the convenience of humans, the relationships can be expressed in a note addressed to the user. But for purposes of computer manipulation, the relationships can only be expressed implicitly by the tags assigned to the fields, and therefore the possible relationships are restricted to those for which appropriately tagged fields exist. The fields are closely aligned with a single set of cataloguing rules: AACR2 [2]. Thus although some bibliographic organizations might wish to show the names of the two persons who are editors of the series shown in field 440, the format does not have a field for these names because according to AACR2 these names should not appear in a record for an item which belongs to the series. The alliance of the format with specific cataloguing rules thus limits use of the format to those who agree to use the rules.

It is also evident that although authors and titles are given for the component parts listed in field 505, they have not been formatted with individual tags for subfield codes for computer use. Nor has any field in the format been designed specifically for such use, even though the cataloguing rules provide for the creation of analytic entries. The fundamental rule which applies to the description of the monograph — that the record is to be decomposed into basic elements for purposes of machine manipulation — cannot be applied equally to these additional authors and titles.

Treating these component parts in a less detailed manner is a result of common library practice. It is assumed that if the component parts are considered worth providing access to, they will be treated as separate bibliographic items. Unfortunately, economic considerations have precluded this treatment in all but a few institutions and certain special kinds of collections, even though a great many catalogue users would be well served by having such access. The increasingly powerful computers available today can be used to retrieve “hidden” entities such as these, but the simplicity of the format dictates that access is possible only for those with information retrieval software capable of full-text scanning or perhaps automatic keyword indexing, and even then only with great expenditures of storage space, processing time, or both.

The second example, seen in Figure 2, shows part of a tagged serial description. Once more we find a record that represents either a single bibliographic item or several items, depending on the treatment it receives. In this case the great majority of fields in the record (most of which are omitted from the example) describe the serial publication Canadian library.
HANDLING BIBLIOGRAPHICAL RECORDS

TAG SELECTED FIELDS AND SUBFIELDS OF DATA

245 @aPolymers in medicine /@cEditors: E. Chiellini and P. Giusti.

260 @aNew York :@bPlenum Press, @c1983.

440 @aSeries in polymer science and technology

505 @aH. Bauser and H. Chmiel. Improvement of biocompatibility of polymers through surface modification.--D.J. Lyman. The development of small diameter vascular prostheses.

700 Chiellini, Emo.

700 Giusti, Paolo.

Fig. 1. Records for a monograph formatted for computer use. Bibliographic level shown in directory: "m" for monograph.

TAG SELECTED FIELDS AND SUBFIELDS OF DATA

245 @aCanadian library.

362 @av. 16, no. 6-v. 25, no. 3; May 1960--Nov./Dec. 1968.

710 @aCanadian Library Association.

780 @aCanadian Library Association. @tBulletin. @x:0316--6058

785 @tCanadian library journal. @x:0008--4352

Fig. 2. Record for a serial formatted for computer use. Bibliographic level shown in directory: "s" for serial.
Field 780 informs us that this item had a previous life as the Canadian Library Association Bulletin. We may assume from examining field 362 that this former title existed prior to the appearance of number 6 of volume 16, which occurred in May 1960. But in order to find out anything more about that title, (such as how long it existed, who published it and where, and so on) we must leave this record and examine another whose control number is not revealed, though its standard number (ISSN) is. This makes examination of that record possible only after a search either by title or ISSN. We also see that the title Canadian library was dropped at the end of 1968 in favour of Canadian library journal, and once more we must search out the record for that title to discover what if anything replaced it, and so on. Nowhere in this record do we see any mention of the Canadian Library Council Bulletin, which is the same journal under a yet another title. To find any trace of it, or of the Canadian Library Association’s conference proceedings, which were absorbed into the journal in 1975, we must first retrieve other records.

In this case library cataloguing rules dictate that the separate titles will receive separate descriptions, so fields in the format are designed to accomplish that. Once more the format, having been designed to implement specific cataloguing rules, imposes limitations on how the data may be treated. A bibliographic agency might decide that its users would be better served by a single record showing the complete bibliographic history of this item in a single record, perhaps to be displayed in multiple locations in an alphabetical title listing. To accomplish this the agency must create that history from a number of successive records, which requires inventing fields to hold information about the place and publisher, volumes, issues and dates of coverage, and so on, for each of the variant titles, thus departing from the standard format which does not provide suitable fields for a record organized in this way.

The two formatted records shown so far are very simple examples of their types. Neither represents any kind of extreme or departure from the norm. Nor are the comments about the records meant to unfairly criticize existing formats, which are in widespread and profitable use in thousands of bibliographic agencies around the world. But these kinds of limitations, which most bibliographic agencies have been willing to accept and a few have circumvented with sophisticated software, still present difficulties for many libraries and other kinds of information centres.

**A NEW STANDARD BIBLIOGRAPHIC FORMAT**

Within the past few years a new format has been published which employs what are known as “record segments” to solve many of these problems. A record segment is a group of related fields within a record, with special data elements called “linking fields” to point at the relationships that exist and identify them by type. Segments, which have been adopted because of the kinds of limitations described above, were made possible by a change made to ISO 2709 in 1981. This Format for bibliographic information interchange on magnetic tape provides the rules according to which standard bibliographic formats are designed. Some countries have their own national version of this standard, such as the American 239.2-1979, American Standard Format for Bibliographic Information Interchange on Magnetic Tape [3], while others follow the ISO standard directly. The standard mandates a structure comprising a 24-character label, followed by a directory, and a variable number of variable-length fields. In its 1973 edition, the directory contained an entry for each field which consisted of the tag, the field length, and the address of the field within the record. The 1981 edition departed from that pattern in providing an additional “application specific” section of up to nine bytes.

The first format to take advantage of this change was CCF : The Common Communication Format [4], published by Unesco in 1984. The CCF, which has been described by Hopkinson [5, 6] and Simmons [7, 8], is different from other standard formats in two major ways. First, it neither includes its own cataloguing rules nor aligns itself with a specific set of published rules. Second, it takes advantage of this application-specific feature to create record segments. To provide the mechanisms which are needed to show relationships both between fields in a record and between records in a file, several linking fields are included in the CCF. They will be examined here and examples given of their use.
Figure 3 shows the same bibliographic item described in Figure 1, as it would appear in the CCF. Certain fields and subfields have been shown to emphasize the differences between the two formats, although in most other respects the records are treated similarly. The record shown here has been divided into a number of segments. Segment 0 is known as the “primary segment” since it contains the fields that describe the “target item” whose bibliographic level is coded in the record label. Two new characters now appear following each tag; these are the two application specific additions to the record directory. The first is the segment indicator, which permits the segments to be separated in the manner shown, even though in an exchange record the fields might be sequenced numerically by tag rather than by segment. The second character after the tag is a repetition counter, which serves to uniquely identify fields when they occur more than once in a record segment. Within segment 0 field 100 appears twice, the first with a repetition counter of 0, the second set at 1.

VERTICAL RELATIONSHIPS

The only other additions to the record are the fields tagged 015 and 080, which appear in all of the secondary segments. Used as links between segments, these fields are used in the CCF to show hierarchical (i.e., vertical) relationships. Field 015 shows the bibliographic level of the segment in which it appears. In this example segments 1 and 2 represent parts of the target monograph, so their fields 015 contain the code “a”, which is used to denote component parts. Field 015 in segment 3 contains the code “s” for the series named in it.

Field 080, General Vertical Relationship, has two subfields: subfield @b shows the segment to which the current segment (the segment in which this field appears) is related. In this record all of the secondary segments (segments 1, 2 and 3) are related directly to the primary segment (segment 0), so all of the subfield @b are set to 0. It is possible, however, to use this mechanism to show relationships between secondary segments. Subfield @a contains a code taken from a code list used exclusively by this field. This list contains only two codes: code 01 means that the current segment is lower in the hierarchy than the primary segment, code 02 means that the current segment is higher. Therefore we can see that in segments 1 and 2, which contains component parts, subfield @a of field 080 is set at 01, since the current segment (the component part) is lower in the hierarchy than the monograph described in the primary segment. In segment 3, which describes a series, the code must be 02, since a series is at a higher level than a monograph.

While the record shown in Figure 3 is only slightly more complex than the one shown in Figure 1, the separation of the several bibliographic items permits enrichment of the record. An agency may wish to include considerably more bibliographic information about any or all of the various items. An agency may wish to provide access in online system to each segment, since each segment records the description of a different bibliographic item. Or it may wish to split the record with its several segments into a number of independent records, while retaining the linking mechanism which points from one description to another and identifies the relationship. An example of the last will be shown below.

HORIZONTAL RELATIONSHIP

Figure 4 shows a CCF record for comparison with Figure 2. Here the relationships are horizontal rather than hierarchical, so CCF field 085, Horizontal Relationship, has been used. These segments are arranged in a chronological chain, starting with the primary segment representing the original title of the journal, which did not appear in Figure 2. Each successor title is contained in a segment that includes a field 085, which contains two subfields. In segment 1, field 085’s subfield @a contains code 21, taken from the list of codes used for this field. Code 21 indicates that the segment shown in subfield @b (the primary segment, segment 0) contains a former title. The same code is used in each of the other segments, since each contains a successor title. The subfields @b in each segment point to the previous segment, so that relationships within the chain are explicitly designated without relying on the preservation of the sequence of segments. The organization of segments shown here would be determined by local policy; some agencies might choose to reverse the sequence. Then the most recent title would be the primary segment, with each preceding title following in a chained sequence of secondary segments.
Fig. 3. Record for a monograph in CCF format. Bibliographic level shown in directory: "m" for monograph.
SELECTED FIELDS AND SUBFIELDS OF DATA

TAG  SEGMENT INDICATOR  REPETITION COUNTER  DATA

SEG-  201  0  0  @aCanadian Library Council. Bulletin.

SEG-  085  1  0  @a21@b0
MENT   015  1  0  @as
1  101  1  0  @a0316-6058
2  201  0  1  @aCanadian Library Association. Bulletin.

SEG-  085  2  0  @a21@b1
MENT   015  2  0  @as
2  201  0  2  @aCanadian library.
3  310  0  2  @aCanadian Library Association.
4  450  2  0  @av. 16, no. 6-v. 25, no. 3; May 1960-
         Nov./Dec. 1968.

SEG-  085  3  0  @a21@b2
MENT   015  3  0  @as
3  101  3  0  @a0008-4352
2  201  3  0  @tCanadian library journal.

Fig. 4. Record for a serial in CCF format. Bibliographic
level shown in directory: "s" for serial.
The CCF permits any kind of relationship to exist for which there is a code. The example shown has purposely been kept simple for comparison with the record shown in figure 2, in a real record it may be assumed that considerably more bibliographic data would be entered. And as with the monographic example in figure 3, it is possible to treat each of these segments as a separate record, while maintaining a linking field with a pointer and a code expressing the relationship.

The list of codes for this field permits the expression of links between earlier, later, and variant editions of a work; supplements and the items they are issued with, translations and the originals from which they were translated, and an item and one or more reviews of the item which have been published elsewhere. In each case the links may be established in either direction, or in both directions, and there is no restriction on the number of such relationships that may be expressed, since field 085 is repeatable. The experienced bibliographer will see from the list in this paragraph that the relationships permitted by the format are sufficient to permit the user agency great latitude in determining its own choice of rules for describing materials and for displaying the descriptions.

FIELD TO FIELD LINKS
In addition to links between record segments, the CCF has a special field to permit the expression of links between individual fields. In most machine formats, when it is necessary to express a link between a publisher and a place, or an ISBN and the publisher with which it is associated, these pairs of elements are linked implicitly by being included in the same field. Another kind of link that may be expressed in many of rules for description is between an author and the organization with which he or she is affiliated, or the author's address. Figure 5 shows the primary segment of the record in figure 3 as it would appear with the authors' affiliations recorded. Field 330, taken from the CCF, has been used for affiliation. Since there are two authors and two affiliations, and because the name of the author and the name of the affiliation organization are in separate fields, it is necessary to link each author with his affiliation.

CCF field 086, Field to Field Linkage, has been used for this purpose. Each field 086 has three subfields. Subfield @a shows the field linked from, complete with its three-digit tag, segment indicator and repetition counter, so that the entire string occupies five characters. Subfield @c shows the field linked to, similarly identified with a five-digit string. Subfield @b is used to contain the code which identifies the type of relationship existing between these two fields. In the example shown, code AA is used to denote a link between author and affiliation. Other codes in the CCF are used to link a publisher and a place, an ISBN and a publisher, a field and a transliterated version of it or part of it, and a field and a version of it or part of it in another script.

The linkage shown in the example in figure 5 is so straightforward as to be virtually unnecessary. It might be argued that sequence is sufficient to show linkage, the first affiliation would implicitly be linked with the first author, and the second affiliation with the second author. But many cases are far less obvious. It is common to find three authors, of whom have the same affiliation, or four authors with two affiliations, and so on. The use of field 086 makes these inter-field links explicit, removing the opportunity for errors caused by ambiguity.

LINKING BIBLIOGRAPHIC RECORDS
The use of record segments permits the description of as many as thirty-six related bibliographic entities in a single record, each designated by a single alphanumeric identification character. However many agencies creating bibliographic records will choose to eliminate hierarchical or horizontally linked records, preferring to house separate bibliographic entities in separate records.

To express relationships between records, the CCF uses field 010 to contain the identifier of another record. Figure 6 shows two segments of our monographic example. The primary segment again describes the monograph which is the “target item” of the record. In segment 1 we see a field 080 which points to a series (coded s in field 015) that is hierarchically higher (shown by code 02 in subfield @a of field 080) than the primary segment (shown by code 0 in subfield @b of field 080). The remainder of the segment consists only of field 010 which shows an address: this is the identifi-

8
HANDLING BIBLIOGRAPHICAL RECORDS

SELECTED FIELDS AND SUBFIELDS OF DATA

Fig. 5. Selected field from a CCF-formatted record showing field-to-field links.

Fig. 6. Two segments of a CCF-formatted record showing a link to another record.
cation number of another record which contains the description of the series. Field 010, Record Identifier for Secondary Segments, points to another record; it thus permits the expression of relationships between independent records rather than between segments of a single record.

CONCLUSION

In the past two decades the development and increasingly widespread use of international standard rules for bibliographic description and standard data formats have accustomed information professionals to handling bibliographic records in the form of a number of elements within a formal structure. As the catalogues, bibliographies, and computer files dependent upon these standards grow in size and value, there is increasing pressure to resist all but trivial changes in the standards. Yet parallel to the acceptance of standardized, machine-readable records, development in such fields as relational database theory and practice are resulting in the rapid growth of more sophisticated data consistently reduced to its most elemental level. The application of this concept to bibliographic records, facilitated by the use of record segments, permits each bibliographic agency to maintain maximum control over the organization of its database. Thus records containing hierarchies, chronologically successive titles, and other kinds of related entities may be created, explicitly designated, and manipulated in ways best suited to the purpose of the organization and to the abilities of the database management software available.

REFERENCES


