Notes on Operations

Evaluation of Three Record Types for Component Works in Analytic Online Catalogs

Herbert H. Hoffman

Works contained in collections and anthologies are a significant body of information stored in libraries. For the retrieval of such works, online catalogs today rely mostly on contents notes and added entry fields. Four criteria for analytic catalogs are suggested: a search for a specific work should retrieve all units of that work; it should retrieve only that work, without false drops; the search should require only one pass; and the resulting display should clearly collocate all retrieved works. It is suggested that "In" analytics as described in AACR2 rule 13.5A promise better results than analytic entries based on contents notes and added entries.

Bibliographic databases that make up library online catalogs contain bibliographic records that represent the books, discs, cassettes, and other items a library has on its shelves. When such an item contains one sole work, the item and the work it contains are perceived to be one and the same thing. It is rarely difficult to retrieve such works. But when one item contains two or more works, it may be difficult to link the works to the item. That may be part of the reason why online catalogs are still hard to use, as Borgman recently reiterated (Borgman 1996).

Much groundwork has been done to distinguish works from items. Lubetzky held that the "work"—the "literary unit"—was the basis of bibliographic description (Lubetzky 1963). Hoffman (1976) has attempted to show that all publications, regardless of medium or format, are first of all "works" (i.e., essays, poems, novels, plays, symphonies, etc.) contained in "books" (i.e., monographic publications, collections, periodicals, cassettes, etc.) that come in "sets" of one or several volumes, and that there are no exceptions. The ERIC thesaurus of 1978, in its basic list of publication type categories, supported this view by defining "book" as "pure form or 'empty container,'" contrasted with "creative works" such as "poetry, literary works, essays, novels, short stories . . ." (Educational Resources Information Center 1978, 178D). These definitions lend weight to Lubetzky's proposition that it is the work—the intellectual creation—and not the package or the container on which the catalog should focus.

Herbert H. Hoffman (hoffm@earthlink.net) is Catalog Librarian, Santa Ana College, Santa Ana, California. Manuscript received January 1, 1998; accepted for publication April 4, 1998.
The bibliographic description of “analyzed parts” versus “containing items” was treated in detail by McCallum nearly two decades ago (McCallum 1980). Still, as Howarth (1997, 9) described the situation, today’s online catalogs seldom manage to represent the “work independently of the physical format.” As Weintraub and Shimoguchi phrased it (1993, 178), a reliable analytic catalog would not only remain a “vehicle for transmitting bibliographic information about whole books” but would “reveal information couched within these books” as well. In such a catalog, the bibliographic records of all of an author’s works, large or small, those that stand alone and those that are “component parts,” would be indexed so that any of them can be searched for and retrieved with one simple keystroke or click.

While a true analytic online database on a national scale does not yet exist, a case could be made for it. Hoffman and Magner (1985) found that for every item on the shelf that is listed in the catalog, there might be five works embedded in collections and anthologies, works that are not listed—not in the catalog nor in any other index, finding list, or bibliographic aid. Thus, leaving the periodical literature aside, a library of 100,000 items might contain half a million works that are not easily accessible to patrons searching the online catalog. Poulsen (1990), in studies at six other libraries, had similar findings. Such data justify the conclusion that our libraries contain enough uncataloged or insufficiently cataloged works to make improved analytic catalogs worthwhile. Hagler (1997, 13) seconds this objective when he suggests that future editions of AACR should require that “an agency provide access to every work . . . appearing within each catalogued document.”

The sheer numbers of works, however, call for a cooperative effort. Given the advances being made in computer technology and telecommunications today, it is conceivable that in addition to a MARC database of books qua containers, a comparable utility for component works will become feasible, a utility that would enable readers to access the works contained in library collections and anthologies, much as they now use INFOTRAC and similar indexes to access the works contained in periodicals. The organizational and administrative aspects of such an undertaking, however, are not within the scope of this paper. We restrict ourselves to a discussion of how different bibliographic record structures affect the search for, retrieval, and display of component works in analytic or partially analytic online catalogs.

Let us consider what an analytic catalog ought to achieve to facilitate the retrieval of specific works. It seems there are four major goals:

1. Once a work has been identified, the catalog should retrieve all versions of that work the library owns, not just some. At present, few online catalogs provide complete and reliable access to works embedded with others in collections and anthologies. Most online catalogs are at least partially blind to component works because the contents of collections and anthologies are not listed in the bibliographic records that make up the database. A sample of 44 anthologies of drama retrieved from the catalog of one large American university revealed that only 15 of them were represented by bibliographic records that contained tables of contents. There was no clue to the contents of the other 29 anthologies. In that library, a student looking for a given work will find some manifestations of that work but will not find all manifestations for that work that the library owns. And because most libraries today download their records from the same pool of MARC records, few other online catalogs are likely to give better service.

2. The catalog should retrieve only only versions of that work and no others. There should be no irrelevant or unrelated titles retrieved, that is, no false drops. In many online catalogs, a search for author “Beethoven” and title keyword “Octet” will retrieve many examples that do not contain any of Beethoven’s octets. For example, this search will retrieve a
record that contains Beethoven’s Septet op. 20 and Mendelssohn’s Octet op. 20. Both search terms are contained in the record for the item, but they do not describe one work; they are not linked specifically to a particular work.

3. It should be possible to achieve the desired results in one pass. Readers should not have to try several approaches before they are confident that they have exhausted all possibilities. A search for author “Schnitzler” and title “Game of love” in many catalogs will draw a blank. Libraries that own Corrigan’s *Masterpieces of the Modern Central European Theatre*, however, do have a copy of the play. The MARC record has a searchable contents field for the title. But there is no searchable author field for “Schnitzler.” To find Schnitzler’s play, then, the reader must do a second title search, which, paradoxically, will only work if the author’s name is left out.

4. The retrieved records should be located in an uncluttered, unambiguous screen display. In today’s online catalogs, as Carlyle (1997) points out, the second objective of the Paris Principles tends to get short shrift. The second objective requires that the catalog collocate all of an author’s works that the library owns (Verona 1971). To display a work in the online catalog means to show its title and, if applicable, its author or authors, in a prominent position on the screen. A user searching for Shaw’s play *Arms and the Man*, for example, expects to see a screen that displays both the author name and the title. However, in most systems, a brief display will include data from MARC field 100 (for author) and MARC field 245 (for title), with results similar to those seen in figure 1. This display is not as clear as might be expected. From the information given, it is not clear that only the last two items contain the play sought. Moreover, neither “Shaw” nor “Arms and the man” are displayed on screen. The reader must search for that information in the full display of each record on subsequent screens.

To construct analytic catalogs that achieve all four goals, librarians must determine which record format will produce the desired results. In the *Anglo-American Cataloguing Rules*, 2d ed., 1988 revision (AACR2R), three record structures for analytics are mentioned. They are contents notes, added entries, and “In” analytics.

The first method—contents notes—is described in rule 13.4A. The rule states (p. 300) that it is “the simplest means of analytics” and that it is “usually limited to a citation of title or name and title.” Blackwell’s “Blackwell Table of Contents” project focuses on this method, making title-searchable contents lists for selected collections and anthologies available for downloading.

The second method—added entries—is described in rule 13.2A, which reads in part (p. 300): “this method is appropriate when direct access to the part is wanted without creating an additional bibliographic record for the part.” It is clear from the wording that direct access to component works has always been a desirable feature of library catalogs. It must be remembered that the rules date back to a time when the standard was a card catalog. An added entry then was an extra card. When filed in proper order, readers had direct access to a given title and found all manifestations of it neatly collocated in the drawer. In the online world, however, an added entry is no longer a separately

<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>TITLE</th>
<th>PUBLICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hille, L. Rust</td>
<td>Lust, violence, sin, magic</td>
<td>c1993</td>
</tr>
<tr>
<td>Grey, M Cameron</td>
<td>Angels and awakenings</td>
<td>1994</td>
</tr>
<tr>
<td>Kernan, Alvin B.</td>
<td>Classics of the modern theater ([1965])</td>
<td></td>
</tr>
<tr>
<td>Barnet, Sylvan</td>
<td>Eight great comedies</td>
<td>c1958</td>
</tr>
</tbody>
</table>

*Figure 1. Display of Titles Retrieved for Shaw's Play Arms and the Man.*
Fields in MARC records do not automatically produce collocated displays. No matter what software a library may have bought, access to embedded works is not as direct as it once was.

The third method—"In" analytics—is described in great detail in rule 13.5A and B. The rule states that an "In" analytic entry be made (p. 300) "if more bibliographic description is needed than can be obtained by displaying it in the note area."

**Contents Notes**

Many librarians will say that a contents note in the bibliographic record for a book or other item containing different works will suffice. There are two ways of entering data into the contents field of a MARC record (field 505). The basic method uses subfield a; similar results can be achieved by placing the titles of component works into the subtitle, i.e. MARC field 245 subfield b. The second, enhanced method uses subfields t and r in MARC 505.

**Basic 505 Field for Collections**

*(All Works by the Same Author)*

Let us first look at the basic contents note method for collections. Librarians raised on AACR2R, Akers’ *Simple Library Cataloging*, and similar classics might not be in the habit of distinguishing clearly between collections that contain works by the same author and collections that contain works by different authors. The two types of items, however, are cataloged by different rules. That is why we prefer a distinction and use the term “collection” here to designate the former type of item. We shall use the term “anthology” for the latter. In a collection, then, the author’s name would be in field 100 of the bibliographic record. The titles of component works can be added in field 505 subfield a or, occasionally, in 245 subfield b. It might seem that this simple method of analytics makes good sense, but there is a problem. Because all strings in 505 subfield a and in 245 subfield b are in one subfield, only a title keyword search is possible. There is no way to instruct a computer to search for the exact title phrase.

An example would be the collection *Eight Plays* by Tennessee Williams (see figure 2). The author’s name appears in the searchable 100 field. But a search for the exact title “Summer and Smoke” would fail. Only a search for author’s name plus one or more title keywords will retrieve this collection. When the retrieved item is displayed on the screen, of course, it will show the title “Eight plays,” not “Summer and smoke,” thus hiding the col-

---

<table>
<thead>
<tr>
<th></th>
<th>Williams, Tennessee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Collected plays ............................................... 1</td>
</tr>
<tr>
<td>2</td>
<td>Eight plays ...................................................... 1</td>
</tr>
<tr>
<td>3</td>
<td>Summer and smoke ............................................... 1</td>
</tr>
<tr>
<td>4</td>
<td>Sweet bird of youth, and two othe................................ 1</td>
</tr>
</tbody>
</table>

---

**Figure 3. Display of Results for a Search for Williams’ “Summer and Smoke.”**
location feature. If the library owns several versions of *Summer and Smoke*, the reader might see a summary display like that seen in figure 3. Such a display confuses the searcher because it is not clear which lines actually point to the play sought.

The basic table of contents method is unable to respond to exact title searches. This is a problem for the reader who begins with an exact title search, then finds out that the retrieved item is not available. The reader has two choices: give up or do a second search, just to be sure that there really are no other copies of the play in the library. And even if the library does not own any other collections or anthologies that contain the work, this second pass is still necessary to be sure all avenues were exhausted.

There is another problem. Keyword searches are vulnerable to false drops. Suppose a reader searched the catalog for author's name “Shaw” and title keyword “Man” (as in *Man and Superman*). It is easy to see that there might be a hit, but that the play the reader retrieved was *Arms and the Man* instead. A keyword search for Palestrina’s “Ave Maria” might well retrieve Hyperion CDA 66850, an item that does not contain the work but does contain one titled “Ave verum corpus, natum de Maria virgine.” Unless the library’s catalog takes word order and proximity into account, a reader looking for the poem *Lost World* might well retrieve the story “World Well Lost” instead.

**Figure 4. Bibliographic Record for an Anthology Using MARC Field 505 Subfield a.**

**BASIC 505 FIELD FOR ANTHOLOGIES (WORKS BY DIFFERENT AUTHORS)**

If an author’s works are contained in an anthology together with other authors’ works and field 505 subfield a is the only analytic field in the host item record, then author searches or author and title searches will not work.

An example is the anthology *Nineteenth-Century British Drama* (see figure 4). Although Oscar Wilde’s *The Importance of Being Earnest* is contained in this anthology, a perfectly logical author/exact title (or even an author/title keyword) approach will fail because authors’ names, in this case, are not contained in any author-indexed field. Pragmatic souls will say that all one has to do to achieve the desired retrieval is to treat the author’s name as a title keyword and include it in the search argument. But this is a stopgap approach, inconsistent with the concept of authorship, and therefore not recommended as a permanent catalog design feature. Library users should not be required to do mental rearrangements like this in order to succeed at the catalog.

**ENHANCED 505 FIELD FOR COLLECTIONS (ALL WORKS BY THE SAME AUTHOR)**

The enhanced 505 field has repeatable subfields t for titles. An example would be *Three Plays* by August Wilson (see figure 5).

**Figure 5. Bibliographic Record for a Collection Using MARC Field 505 Subfield t.**
Unlike the basic 505, the enhanced 505 with subfield t can, if the library’s software allows it, be fully title indexed and can then be searched by exact title phrase as well as by keyword. The author’s name is in field 100, the title in 505 subfield t. Thus, exact title, keyword, and author/title searches will succeed. The possibility of a false drop is remote. But readers would find it hard to notice the collocation feature. For even if one has searched for and found a work title in a browsing display, the machine is likely to generate a final item display that shows the title contained in field 245, not the work title contained in field 505. An unsuspecting, maybe less sophisticated user looking for the title Fences might, on seeing the summary display in figure 6, pick line 2 and take home the wrong book.

**ADDED ENTRIES**

Some believe that the best way to provide analytic access to component works is to use added entries. This is the method advocated in such rules as 13.2A and 21.30M1 of AACR2R. It requires the addition of a suitable combination of 7XX fields to the host item’s bibliographic record, either alone or together with a 505 field.

**7XX Fields for Collections (All Works by the Same Author)**

For a collection it would be easy to add a number of 740 fields to the item record, one for each component work title. Because in such a situation the name of the author of all the works is in the 100 field, an author/title search would be successful. An example is *Three by Tennessee* (see figure 8). As in the situations described.
above, however, when the host item is displayed on the screen, the title shown will be that contained in the 245 field, thus hiding the identity of the retrieved work behind the item or document title.

7XX Fields for Anthologies
(Works by Different Authors)

For an anthology, one or more 700 fields would be indicated, where subfields a and t contain the authors and titles, respectively, of component works. It might look as if exact title, title keyword, as well as author/title searches are provided for in this case. There is a problem, though, when author names in 700 subfield a and titles in 700 subfield t are not linked. For example, a search for author “Palestrina” and title “Ave Maria” in most catalogs would retrieve a record for an item that contains Palestrina’s “Sicut Cervus” and Robert
Parsons' "Ave Maria" because both terms occur in that record. The machine finds the author "Palestrina" and the title "Ave Maria" and registers a false hit.

This will remain a problem until newer software such as Horizon, Voyager, and others make it possible to limit the search for an author and a title to the same field, excluding all other hits. Because "Palestrina" and "Ave Maria" in the example given above are not in the same 700 field, such an advanced system would avoid the false drop. Even that, however, will not solve the problem of coauthors, editors, performers, conductors, and the like. In bibliographic records that stand for single-work items, or "standalone works" as we could call them, such names, entered in separate fields, are ipso facto linked to the work. The moment two or more works are described in the same bibliographic record those relationships become blurred. Consider the case of the RCA Victor Gold Seal recording 7709-2-RG. The OCLC record has a 245 subfield a "showpieces." There are several added entry fields, two of them for the conductors (see figure 9). This CD contains five pieces, four of which are designated by analytic 700 fields. Even if authors' names and work titles in these analytic 700 fields were linked, there is nothing in any indexed field that would allow a library user to focus on the works for which each conductor is responsible. A search for conductor (in most online catalogs, conductors' names are included in the author index) "Solomon" and title "Zigeunerweisen" will retrieve something, but it will not be that work conducted by Izler Solomon.

There are still other complications. Works often appear in translation. Calderon's Life Is a Dream, when published as a separate book, is represented by a bibliographic record that automatically links the English title with the original La Vida Es Sueño in a separate field. The same goes for nicknames and alternate titles. If a search for "Eine kleine Nachtmusik" brings up a standalone work titled "Serenade in G," we can be sure that we are dealing with one and the same work.

But when the work is one of many in an anthology and there is only one bibliographic record for the item and its contents, the links between such separate fields and the works to which they belong are more difficult to establish. The catalog retrieving Chekhov's "Vishnyovy sad" should equate that work with "The Cherry orchard." However, as long as all titles contained in a multiwork item are in separate fields, the catalog has no precise way to establish and display this link.

One other detail needs to be considered. When an added entry in the 7xx fields names a work that is contained in the item named in field 245, the second indicator is given the value of "2" to show that the added entry is an analytic one. This indicator is not as helpful for the analytic retrieval of embedded works as one would hope. In the first place, there is, at this point, no library software available that would, on retrieving such an embedded work, create a display such as the one found in figure 10.

But even if there were such software, it would work only if the 7XX fields contained the actual title of the work sought. In many situations this is not true. A certain recording (EMI CDC 7 49656 2) contains Beethoven's Symphony no. 5 in C minor, op. 67. A title search for this work would, in most libraries today, not retrieve the EMI recording because the relevant analytic added entry uses the collective uniform title "Symphonies." The second indicator, according to the rules, marks this 700 field as one that designates an analytic work contained in the item in hand. But in this case subfield t designates a class of works, not the specific work sought. Needless to say, a search for Symphony no. 5 finds no match.

It would appear that, for works contained in collections and anthologies, not all author/title/subject searches can be handled by MARC 700 added entry fields, even if analytic indicators and "in-the-same-field" strategies are employed, unless the field in question is expanded to contain coauthors, editors, performers, translated titles, subject headings, etc.—all the information that in standalone works appears in separate fields.

Conceivably, added entry analytics could also be constructed using separate
Figure 10. Bibliographic Record for an “In” Analytic Structured in Accordance with AACR2 rule 13.5.
IN" ANALYTICS

The "In" analytic, described by rule 13.5A of AACR2R, is a third possibility. In an online catalog, such analytics consist of separate bibliographic records for all component works, whether contained in collections or anthologies, each with its own MARC 245 and 1xx field, as appropriate. Such a record will also contain a "Host item entry" MARC 773 field, a field that is described in MARC manuals as containing information concerning the host item for the constituent unit described in the record (vertical relationship). As the scope statement for 773 explains, this field is provided in order to enable the user to locate the physical piece that contains the component part being described.

For the example shown above, Eight Plays by Tennessee Williams, eight separate "daughter" records would be created (see figure 11). This stratagem produces reliable retrieval because the author and title are unequivocally linked. In an author/title search, there is no possibility of linking an author and a title that do not belong together, because there is only one title in the bibliographic record. If this method is used consistently, the catalog, searched by author and title, will retrieve all manifestations of a work in one pass. There can be no false drops. And the display will clearly collocate all retrieved manifestations of a work without intervening titles that are not those of the work sought (see figure 12).

The "In" analytic also solves the problem of linking joint authors, performers, alternate titles, uniform titles, subject headings, etc., to the works they pertain to because each work is represented by its own bibliographic record. All fields in that record are, by definition and without special programming, linked to the work named in field 245. A search for author "Beethoven" and title keyword "Octet" will retrieve all and only records that carry this author's name and this title keyword in one of the fields. A false drop such as the London 421 093-2 disc, mentioned above, is not possible. The reader looking for Euripides' Medea will not ever be confused again by being presented with Alcestis instead. If you are looking for Schnitzler's Game of Love, you will not only find it, but find it on the first try.

The subject approach to component works should not be forgotten either. Many studies in library literature confirm the importance of the subject approach, especially in academic libraries. Larson

1. Williams, Tennessee
   Summer and smoke.......... 1
2. Williams, Tennessee
   Summer and smoke.......... 1
3. Williams, Tennessee
   Summer and smoke.......... 1

Figure 12. Brief Display of Analytic Records.
(1991), for example, concluded that subject searches are not only most used but are also most likely to fail. Without investigating in detail the several reasons for failure, it is easy to see one of them. When many works are packaged together in one container and each work deals with a different subject, searchers are likely to miss important information. The primary reason for this is that multiwork items are often assigned only a heading that summarizes the subject content of all the component works. The Analytic Spirit: Essays in the History of Science (Cornell University Press, 1981), for example, was assigned the summary heading Science—History. In the contents field, the titles and authors of all 15 essays, ranging from a discussion of Lavoisier’s theory of the gaseous state to the supernova of 1054, are listed. Will a reader interested in the Crab Nebula look under the broad heading “Science”? Probably not.

Librarians at libraries where users have a strong interest in astronomy, remembering Ranganathan’s fourth law (“Save the time of the reader”), might want to make it easier for readers to access interesting materials by adding essay-specific subject headings to a book like this. As long as the essays are listed only in a 505 field, or even in a 7XX field, it is impossible to tell which subject heading goes with what essay. In such a case, the “In” analytic seems indicated. For the essay by L. Pearce Williams, “The supernova of 1054,” a separate bibliographic record might be created, complete with author, title, specific subject headings, and link to the mother record.

The reader searching by subject will now retrieve all relevant “standalone” as well as component works. The search is precise and exhaustive, requires only one pass, and the retrieved materials will be neatly and unequivocally collocated on the screen.

CONCLUSION AND RECOMMENDATION

There are three major methods employed in libraries to catalog component works—contents notes, added entries, and “In” analytics. Under the third method, separate bibliographic records for all component works are created, and these are linked by a common key to their host records, the collections and anthologies that contain the works. If all collections and anthologies on a library’s shelves are consistently cataloged by this method, a search for a specific work will:

1. Retrieve all manifestations of that work,
2. Retrieve only that work without false drops,
3. Not require a second pass, and
4. Clearly collocate all retrieved work titles.

If librarians plan to offer their readers reliable analytic catalogs that perform well on these four points, then the choice of record format will have to be given serious consideration. It appears that “In” analytics linked to their host item records, a method so far neglected in most online catalogs, might be the choice of the future.

WORKS CITED


