The Reliability Factor in Subject Access

Constance McCarthy

For truly effective subject access, it is essential that books on any given topic be brought together consistently under the same subject heading. With the advent of online catalogs, this goal has assumed new importance but has also become easier to achieve.

As the library community has gained experience with online catalogs, subject access has assumed new importance as a topic of discussion. Experts from several constituencies within librarianship are bringing their skills and experiences to bear on an old problem in a new setting.

The old problem is that of creating the best subject indexing for a large collection of items that is never static. New considerations arising with online capabilities include the possibility of overcoming some of the limitations formerly imposed by the presentation of the catalog on cards as well as the desire to take advantage of any other improvements to subject access that computers might make possible. The recent series of articles in American Libraries by Pauline Cochrane, to which many others contributed, served well to delineate the present range of opinions within the profession. Carol Mandel and Judith Herschman, in an article on possible ways of improving subject access in online catalogs, provide a review of the literature up to February 1983.

For the most part, the issues being debated thus far concern the online catalog itself and the features it has or might have; the adequacy of the Library of Congress Subject Headings as a thesaurus for the online catalog; and the possible enhancement of the catalog record by adding more subject headings or key words from tables of contents or elsewhere. An issue that has not been addressed in this context, as far as I can discover, is that of uniformity or consistency in the application of subject headings (as distinct from consistency among headings, the lack of which produces such often-noted pitfalls as French Literature, but Philosophy, French). To avoid confusion, I have given the name "reliability" (a less-than-ideal term) to this particular kind of consistency.

Before discussing this matter further, it may be useful to consider briefly the purpose of subject headings or descriptors and the ways in which a search based on their use differs from one using free text or key words. The purpose of subject headings is to make it possible to find all documents about the same subject by consulting the same term. The key concept in this oversimplified definition is of course all. Enthusiasts describing free-text searching invariably give examples of the many useful documents retrieved; they seem not to realize that—for all they know—the most useful documents may have eluded them.

Controlled subject heading structure serves the same purpose as its counterpart, name authority control; they each make possible a comprehensive search.

As another prefatory note, I want to explain that I am a reference librarian. My viewpoint on subject headings is that of a consumer, not a producer; my ideas have been arrived at through my experience in...
assisting library users in their subject searches as well as through my own individual experience.

The aspect of subject access I wish to discuss is the reliability of the choices made by catalogers when assigning subject headings. Regardless of the merits of the subject term itself, can we find grouped under it all the books on the subject it identifies? Or will some catalogers have chosen other headings? Trade-unions may not be an idiomatic term for contemporary Americans, but reliability will be served if all the books on labor unions are given that subject heading and not dispersed under, say, Labor and Laboring Classes or Industrial Relations. I have chosen to use the term "reliability" for this concept, because the other terms used by catalogers who discuss it have become so identified with other topics that they might be misleading.

It has, of course, nothing to do with the form of the catalog; however, I believe that in the online environment it has an even greater importance than in the card catalog. Further, online catalogs have certain features that should make reliability easier to achieve.

By reliability then, I mean simply the basic goal of bringing all books on the same topic together under the same subject heading. A. C. Foskett, speaking of "consistency," says it means that once the user has found out how a subject has been indexed, he will find all similar documents, which will presumably also be relevant, indexed in the same way.4 Lois Chan, who refers to "uniform heading," says "a true subject entry lists all works on the same subject together regardless of the authors' choice of terminology."5 David Haykin, using the term "unity," says "a subject catalog must bring together under one heading all the books which deal principally or exclusively with the subject, whatever the terms applied to it by the authors of the books and whatever the varying terms applied to it at different times."6

In the present context, the problem is not, usually, that the authors' terminology has misled the subject cataloger. Rather, it may be a national cataloging policy that places undue emphasis on specificity and on finding an exact match for the scope of the entire book, whatever the cost in dispersal of books among many headings. Or, it may be the presumably constant increase in the complexity of topics that appear in the literature. Certainly the problem has been exacerbated by a policy of restricting the number of subject headings used. This restriction was presumably necessary because of problems related to the amount of space in the card catalogs. Another factor is undoubtedly that subject catalogers at the Library of Congress, although they are working with an online catalog, do not have easy and immediate access to a consolidated list of all established subject headings, including those most recently formulated.

Whatever the reasons, books in many subject areas are simply not reliably brought together by our present catalogs. If the subject of a book can be encompassed in a proper name, then that book will be handsomely treated by the subject catalogers at the Library of Congress, and it is almost certain to find its place with other books on the same person, place, or corporate entity. If it is about a chemical compound, a single identifiable literary genre, a historical period, or any of many other definite and discrete concepts, it will fare nearly as well. But books devoted to less indisputably definable topics, or to subjects too new to be familiar, seem to stand a poor chance of ending up together with others on the same topic under the same subject heading.

One of the very few places where this situation has been discussed is in John M. Christ's book on the lack of correspondence between the terminology of the social sciences and library subject headings. In describing the large number of different subject headings assigned to books on the sociological topics "role" and "culture," he points out that "there appears to be a high degree of dispersion in the variety of subject headings assigned by the Library of Congress."7 For 109 bibliographical entries in his study, on six different subjects, 188 subject headings (fewer than two per title) were assigned, of which 121 were different!

But it's not that the problem concepts are all as arguably amorphous as those
studied by Christ. Last year, for example, the reference collection at Northwestern University Library had five different directories of online databases. Among them they had eight different subject headings, but no subject heading was common to all five. This would have been less of a problem if they had not used, among them, five different Dewey classification numbers. Imagine the reference librarian who is looking for any one of them without being able to remember the title or the author’s name. And then imagine a library user hoping to find information from such a directory, and trying to select the best one from the library’s holdings!

There was nothing wrong with any of these subject headings and, to some extent, they reflected differences in content. But any library user would have expected to find all five directories under the same subject heading. I use the past tense because the Catalog Department at Northwestern, freed from the burden of card filing and maintenance, has been willing to provide us with additional headings for some titles when we call serious inadequacies to their attention. This was an instance where we requested, and got, the same two subject headings applied to each of the directories. As grateful as we are for these interventions on our behalf, it is too time-consuming a procedure in both departments for us to resort to it often. And if every reference department trying to keep track of its directories of online databases were to follow the same path, the expenditure of time would be truly monumental.

What catalogers and others having infrequent contact with the public may not fully realize is the extent to which users depend on the catalog to use the same heading consistently for books on the same subject. Users may find some of our terminology to be peculiar, but once they have found a subject heading that produces some books on their topic they normally assume that all the books in the library on that topic will have that same subject heading. I talk to users every day who are having problems with subject searching. I would not attempt to generalize about their problems (terms too broad, too narrow, unfiltered through LCSH), but one thing they all have in common is their complete faith in the total reliability of the library’s subject headings.

My concern about reliability arises primarily from my experience as a reference librarian using subject headings both to locate reference books and to assist users in their subject searches. The examples I use here all concern reference books, but subject access to the general collection is not, I think, dissimilar in any significant respect. It seems to me that the experience of reference librarians has value at a time when we are considering changes in the way our subject catalogs are created. We are sometimes referred to as “sophisticated” or “skilled” catalog users, but we could also be called “stubborn” catalog users. I believe that one of the chief differences between a librarian and a naïve catalog user is that the librarian is likely to be looking for something that he or she knows, or is nearly certain, must be in the catalog, whereas the user is far more likely to give up a search, assuming that there is nothing available on the topic or that everything available has been found—thus the most relevant or most basic titles may remain undiscovered. I believe it is now the responsibility of reference librarians to become articulate, as well as stubborn, catalog users, and to participate in the discussion of online subject catalogs more actively than they have. It is to this end, and not with any wish to carp about the generally good subject cataloging provided by the Library of Congress, that I record these instances of professional user frustration.

Take, for example, *The Book of American City Rankings*, by John Tepper Marlin (New York: Facts On File, 1983). Where would you expect to find this book, which ranks cities on everything from books loaned per capita by public libraries to homicides as a percentage of all deaths? Certainly not an easy book to categorize, and perhaps the single subject heading Cities and Towns—United States—Handbooks, Manuals, etc. is not too bad. At least, it wouldn’t be too bad if you hadn’t become accustomed to finding *Places Rated Almanac* by Richard Boyer and David Savageau (Chicago: Rand McNally, 1981) under a subject heading that well de-
scribes the content of both books: *Quality of life—United States—Statistics* or under one of its other subject headings: *Social Indicators—United States or Cities and Towns—United States—Statistics*. However, the *Book of American Rankings* by Clark S. Judge (New York: Facts On File, 1979), which ranks the states in the same way that the Marlin book ranks cities, has only *United States—Statistics* as a subject heading, which might have led you to rely on *Statistics* as a subdivision of place when looking for similar books in the future. The first edition of the *Book of World Rankings* by George T. Kurian (New York: Facts On File, 1979), on the other hand, had only *Social Indicators* as a heading. The new edition, *The New Book of World Rankings* (1984), however, has *Statistics, Economic Indicators*, and (hurrah!) *Quality of Life—Statistics* as well. But wait! A new edition (1985) of *Places Rated Almanac* has appeared, and this time it has four subject headings instead of three! But, alas, *Quality of Life* is no longer among them, although the other two headings from the 1981 edition have been retained, and two new headings, *United States—Statistics* and *Almanacs, American* have been added.

Presumably the subject catalogers of these books didn't search for other works on the same topic before choosing a subject heading. But then, it is also unlikely that they will ever be trying to find any of them in the future to answer a reference question or to satisfy a client who knows of the existence of these books and expects to find them, if not under the same headings, at least under headings that are consistent with one another. Users of many online catalogs get subject tracings with each catalog entry they see. It should at least be possible for them, knowing of the existence of any one of these books on rankings of places, to enter its author or title into the online catalog and get a display of subject tracings that would produce all of the others. And when they have located a book through the use of a subject heading, their reasonable assumption that all other editions of that book have been found also, having been given the same subject heading, should not prove wrong.

An example of a different kind of problem is *America's Working-class Women: A Historical Bibliography* by Susan Estabrook Kennedy (New York: Garland, 1981). The only subject heading assigned to this book by the Library of Congress was *Working-class women—United States—History—Bibliography*. This was the first time *Working-class Women*, a newly established heading, had appeared in our reference department subject catalog. If we were looking for bibliographical information about the history of women at work, which is the predominant element in this book, our experience and instincts would tell us to look under *Women—Employment* or a similar heading (for a secondary element, the heading *Labor and Laboring Classes* might come to mind); we would not think of *Working-class Women*. The correctness of the heading assigned, which is without doubt the best single heading for the book, is not in question here—and of course we could have found the heading. To do so we could have gone to the 1981 supplement of *LSCH* and found a see also reference under *Women* (not, of course, *Women—Employment*). How likely would we be to do that? The answer is simple: we would never do that—first, because it would take too much time, given that we would be searching without any real conviction of need (isn't *Women—Employment*, a heading under which we would already have found other, related, titles, good enough?), and secondly, because we would have many other avenues of approach to the topic, through indexes, bibliographies, and sometimes, proximity of classification number. Being unaware of the existence of the ideal bibliography for our needs, we would not feel deprived by its absence.

One of the very few advantages of a card catalog over an online catalog is that headings and guide cards that are several inches away and even more distant labels of drawers are all within the field of vision of a user looking at a single heading. With a computer catalog, even if there is an index or thesaurus of terms, only a single screen of data can be seen at once. This means that there may be less awareness of the extent of the subdivisions in a file or even of headings that vary only slightly from those being viewed. This limited
view imposed by the terminal screen bestows added importance on each heading, as alphabetically related terms are less likely to be seen.

LUIS, our online computer catalog at Northwestern, has as yet no subject cross-references. But let us assume for a moment that we do have up-to-date online subject authority records incorporated into the online catalog. Would reference librarians and users, having once found Women—Employment in our subject guide screen, go back several screens to look at the cross-references under Women? Probably not. What we may need is a more generous allotment of see also from or broader term headings referring to Working-class Women. In addition to Women, Labor and Laboring Classes, and Social Classes, the xx references in LCSH, we might do well to have Women—Employment, Women—Social Conditions, and perhaps others. It has been suggested that cross-references—including broader, narrower, and related terms—be displayed on the screen automatically whenever the user requests a subject term. Another solution in this case would be simply the more generous assignment of subject headings. The bibliography under consideration contains more information on Women—Employment than do some others devoted entirely to that topic; why may it not have a subject heading bringing out that aspect of its contents?

Another instance in which the cataloger has presumably tried to choose the best single heading for a book is the World Press Encyclopedia edited by George T. Kurian (New York: Facts On File, 1982). The only subject heading given to this two-volume work was Government and the Press—Handbooks, Manuals, etc. Certainly much of the information contained in this comprehensive nation-by-nation survey of the press in 180 countries relates to the relationship between news media and governments. However, it is also the best available single source on the characteristics and history of the world’s press. Why should we hide it away from library users hoping to find this kind of information under such a heading as Press, Journalism or even Newspapers? Why is there a tendency to ration subject headings for large, important books useful to many, while squandering them on tiny or special interest works?

Some other instances of subject heading inconsistency that have caused problems for Northwestern reference librarians concern certain frequently used directories. We have found that we cannot rely on finding consistent subject headings for directories of colleges and universities, of graduate programs, of academic scholars and university faculty members, or of authors’ guides to publishing in various fields. These instances come readily to mind because they are the titles of handlists we have produced for our own use and to give to students looking for this kind of information; we could probably use many more such lists if we had the time to compile them. In the case of directories of graduate education, for example, many directories have a subject heading following this pattern: Sociology—Study and Teaching (Higher)—United States. Others, however, omit (Higher) or add Directories or omit United States, until there is no way of predicting just what the subject heading will be. The problems with the other directories mentioned are similar.

To summarize the difficulties discussed in the preceding examples:

1. Different terms may be chosen by subject catalogers for books dealing with the same topic, as with the online directories and the books about ranking of places. Presumably this is particularly a danger with new topics.

2. The level of specificity may be such as to divide a book from others on closely related topics, as exemplified by our bibliography on working-class women. Presumably this is also a hazard especially characteristic of areas of expanding scholarship.

3. Not enough headings are assigned to bring out the contents of the book. This was the case with The World Press Encyclopedia and also with America’s Working-class Women.

4. As with the directories for graduate education, subdivisions may have been applied to different levels of specificity.
5. Cross-references, as with *Working-class Women*, are made too sparingly, and at too high a subdivision level.

Now let us consider possible solutions to the difficulties—ways of insuring to the best of our limited ability that materials on the same subjects may be located under the same subject headings. Many would immediately mention keyword access, free-text searching, and Boolean logic as the best hope for overcoming some of the limitations of our catalogs. For the bibliography on working-class women, we need only have entered the word *women*, the word *work*, truncated, and the word *bibliography*. True, we would have retrieved that bibliography, but we would have missed others in our reference collection whose titles do not include *work* or *women*. To get all of them, we would need considerable skill in searching and—the greatest flaw—we could never be *sure* that we had all of them, or (what we really want) all the best of them. This situation would be alleviated somewhat by enrichment of the records with phrases from the table of contents or index, but it could never completely be overcome by such means. Perhaps the most important function of subject cataloging, with the human intelligence and forethought it entails, is to try to make possible a true comprehensive search of the database to which it provides an index. A glance at the other examples will show that even greater difficulties would attend an attempt to retrieve them with certainty through free-text searching. Keyword searching, although of the greatest usefulness for narrowing a search when too many books have been retrieved or in locating materials for which existing subject headings are inadequate or insufficiently current, will never solve the reliability problem for us; in fact, it is very nearly the antithesis of reliability! In any case, before counting too much on our users' abilities as online searchers, we need to do more study of transaction tapes from online catalogs and of reports from librarians working with end users. 

Other, more pedestrian approaches show more promise. The simplest—and perhaps the best—would be to encourage subject catalogers to provide more subject headings for most books. (The average number at present has been calculated by different studies at between 1.4 and 1.9 headings per book.) This approach, which has been advocated by many, would greatly improve a user's chances of finding materials. Surely libraries using card catalogs and worrying about space limitations could simply refrain from using all the headings? If necessary, perhaps LC could designate an optional break-off point, as they do with Dewey numbers.

In the same vein, a loosening of the prohibition against assigning headings that represent less than the entire book would be helpful. Not that the catalog should try—at least not at this point in its history—to take on the functions of indexes and bibliographies, but when a substantial component of a book can be bought out by additional subject headings, it should be done.

Other promising approaches involve the cross-reference structure. First, we should make it a top priority to start using the cross-references we already have. "To paraphrase a famous quotation, except in the LC Public Catalog, the Library of Congress subject heading system has not been tried and found wanting; it has never been tried." Most large libraries long ago gave up maintaining an adequate structure in their card catalogs, at what expense to users it is painful to imagine. The experience at Northwestern, unlike that reported by some other libraries, was that the advent of the online subject catalog caused users to turn to *LCSH* with astonishing diligence. We have bought several additional sets, many of which have had to be rebound. But compared to having the cross-references in the catalog where they belong, and recorded in an easily understood format, this is a mere makeshift.

Next, we can improve the entry vocabulary. There is no need to use synonymous subject headings, as some have suggested, to try to match our users' vocabularies; but in the computer catalog we could increase the number of see references almost without limit and greatly improve the usefulness of the catalog to the public.

To achieve reliability, however, it is the
see also references directing users to broader, narrower, and related terms that are of greater significance. The new cataloging rules for see also references recently implemented by the Library of Congress should, in time, make LCSH equivalent to a thesaurus, with better logic and a greatly improved system for assisting users to move up and down the ladder of specificity. This encourages us to speculate about the possibility of manipulating the subject authority files in ways that would be predictable and easy. One of the worst enemies of reliability is the tendency of the user to approach the catalog on a different level of specificity from that of the subject cataloger or the existing literature. What if LC’s new rules eventually enabled us to approach their subject headings as we do the tree structures in MESH (the National Library of Medicine’s Medical Subject Headings), exploding the online search into a more general subject heading or an array of related terms when too little is found under a narrow heading? If that is too visionary, we should at least be able to overridesubject subdivisions at will, producing a broader result. In years past, as a matter of policy, some types of subject heading files were subdivided only when they reached unmanageable proportions. However, it was not normally considered possible to go back and add subdivisions, where appropriate, to the previously cataloged titles. This policy itself thus produced a great deal of “unreliability” and confusion. In recognition of this problem, each term is now more likely to be subdivided as minutely as possible to allow for infinite expansion, a policy that often results in the separation of titles users would hope to find together. At this juncture, it should be possible for the user to elect the use of a broader heading. In any case, now that it is possible to use Boolean logic and other searching techniques in online catalogs, the terror of the long file should soon be alleviated. The user who does not want everything on a particular topic can limit by date or combine the heavily used term with one or more free-text words.

But what can be done about the most serious of the impediments to reliability—the catalogers’ choice of different terms for books on the same topic? Manipulation of cross-references alone will not solve this problem. The only solution, it seems to me, would be for catalogers to examine the literature on the topic under consideration to see what subject headings have been used in the past. This would involve not only checking to see how a particular subject heading has been used but would also sometimes necessitate the use of bibliographies, indexes, or online searches to locate books on the topic by means other than the Library of Congress subject headings. (Perhaps this is sometimes done now, but I assume that it is not common practice.) The books so located could then be examined for similarity of subject matter and the subject headings brought into alignment either by matching the new book to the old or by adding headings to the old books. This would, of course, be time-consuming and expensive, but the improvement in the quality of subject cataloging would be invaluable. Also, or at least so Robert Holley and Robert Killheffer maintain, such efforts are fully justifiable if one considers the attention expended on name authority work, despite the fact that the chances of the same name being used more than once are only one in four.

The other suggestion I have is that subject catalogers try to think more of the catalog as a whole—how its headings interact together—and somewhat less, perhaps, about finding the ideal subject heading for each book. According to Janet Swan Hill, head of cataloging at Northwestern, this is easier to do when the catalog is online. “You are better able to consult it,” she says, “and to consult it creatively; and you begin to think more in terms of the catalog as your creation, rather than just the individual records. This rather monstrous shift in attitude has certainly happened here, and is almost totally attributable to the switch from the card catalog to the online one.”

The Reference Department at Northwestern meets almost every week on Friday mornings. As the last item of business, new reference books of interest are introduced and briefly described. Two im-
important questions will be asked about many of these books: "What subject headings have they given it?" and "Where is it classified?" We need to know because, in our reference collection of nearly fifty thousand volumes, the call number and subject headings are essential if we hope to find the books again. For many new reference books, such questions are not necessary: as mentioned earlier, subject access for proper names and easily identifiable concepts are excellent. For others, we nod our heads approvingly when we hear the subject headings—"that's just where I'd look for it." But in all too many cases, everyone simply groans—"they've done it to us again."

Can we, as a profession, consider the distress catalog users would feel if they knew how many books they may never find because of inadequacies in the application of subject headings? Can we summon up the commitment and the resources necessary to produce really adequate subject access? Such a task would require a greatly increased application of human time and intelligence in an era when we turn all too readily to computers for solutions. But for catalog users, the results would be of inestimable value.

REFERENCES AND NOTES

5. Lois Mai Chan, Library of Congress Subject Headings: Principles and Application (Littleton, Colo.: Libraries Unlimited, 1978), p.24. Chan points out here that consistency does not necessarily imply uniform entry—brining together all the books under a single heading. Theoretically, it would be possible to use several synonyms for a topic and still maintain consistency, as long as all the books were given all the headings. Economy, however, has so far required that we instead make see references from the unused synonyms.
8. The headings are Information Storage and Retrieval Systems—Directories; Machine-readable Bibliographic Data—Directories; Information Networks—Directories; On-line Bibliographic Searching—Directories; On-line Data Processing—Directories; Computer Industry—Directories; Electronic Data Processing; and Business—Data Processing.
10. The Frankenstein Catalog, by Donald F. Glut (Jefferson, N.C.: McFarland, 1984), for example, has the following subject headings: Shelley, Mary Wollstonecroft, 1797-1851; Frankenstein—Bibliography; Shelley, Mary Wollstonecroft, 1797-1851—Adaptations—Bibliography; Frankenstein Films—Catalogs; Monsters in Literature—Bibliography; Monsters in Mass Media—Catalogs.
11. The fact that in the case of professional schools, the formula is different, producing Medical Colleges—United States, Law Schools—United States, but Schools of Architecture—United States; although a parallel inconsistency, this is an example of inconsistency among subject headings—far less damaging, it seems to me, than the inconsistency in the application of subject headings I have been describing, as it can more easily be mitigated by cross-references.
12. For the skeptical reference librarian's point of view on some of these matters, see Anne G. Lipow,


15. Mandel and Herschman, p.149.

