An Introduction to Authority Control for Archivists

Jackie M. Dooley

Was Zimbabwe once part of the same nation as the territory now occupied by Malawi? What was John Paul II's name before he was elected Pope? Why am I unable to locate the papers of the novelist Gordon Ashe? Was UNESCO ever known by another name? Does the Library of Congress really own no manuscripts pertaining to the Vietnam War?

Any good reference archivist can find the answers to these questions in a variety of encyclopedias, directories, bibliographies, dictionaries, and finding aids. It may come as a surprise, however, that the answers can also be made available through a single source: an archives catalog that has been organized using authority control.

Few people are strangers to authority control concepts, though they may not realize it. The terminology of "authority control" comes from the library world, but its tenets are used to promote *consistency of language* in many nonlibrary environments. Consider some circumstances in everyday life in which we puzzle over the ambiguous or synonymous terminology used in information sources: searching for a particular product in the telephone book; looking up a recipe in a cookbook's index; wondering where a recent study on photographic conservation was filed; or determining where the local record store shelves its anthologies of Baroque dance music. Imagine how each of these problems could be alleviated through consistency of language and a good system of cross-references, as we explore the elements of authority control.

This paper outlines the basic concepts of authority control, provides a rationale for the use of authority files in an archives catalog, explains new opportunities afforded by automation, and suggests authority control applications used in libraries that may also be appropriate for archives.

DEFINITIONS AND OBJECTIVES

Bringing the record descriptions in an archives under authority control is a time-consuming, exacting, and on-going task. Any repository must, therefore, begin by asking: "Is the value of authority control worth the expense?" The following definitions will help lay the groundwork for a rationale for the use of authority control and help us begin to answer this important question.

Definitions

Authority control provides standardization of terminology; you might think of it as quality control for concepts and vocabulary. Avram says simply: "authority control is a process for insuring consistency ... and consists of the following elements: distinguishing names ... showing relationships ... [and] documenting decisions."[1] As we will see later, this is accomplished through three basic techniques: vocabulary control, collocation of related material, and establishment of linking relationships among concepts.

Libraries have used authority control in card files and book catalogs for many years. Automation now offers new possibilities for simplifying and expanding authority control, but the basic concepts are identical in both manual and automated environments.

In MARC AMC records,[2] every name, subject, and title in an access point field (MARC fields 1xx, 4xx, 6xx, 7xx, 8xx) is a candidate for authority control. Online catalogs index these fields for retrieval, so it is this terminology which must be made consistent if the catalog is to perform effectively.

Authority work is the process of determining the form of name that will be used in a bibliographic record, determining all necessary and useful cross-references, determining relationships of the name to other names in the file, and documenting the decisions made.

Many archivists may not realize that they already do a great deal of authority work. The detailed agency histories and biographical statements that evolve during the process of arranging records and

manuscripts contain most of the information needed to devise authority records. Unfortunately, this information normally is stored in a single MARC AMC record, though the agency or organization name and history may be relevant to many different series or collections.

Authority records are the data sets that record the decisions made during authority work. As stated above, these data are simply the names and relationships established, together with documentation of the decisions made. Fig. 1 is the authority record for Pope John Paul II. The established heading appears on line 2. "Wojtyla, Karol," the Pope's secular name, is the first of several cross-references (line 3).

```
1  010       n 78078345
2  100 10    John Paul II, Pope, 1920-
3  400 10    Wojtyla, Karol, 1920-
4  400 10    Joannes Paulus II, Pope, 1920-
5  400 10    Juan Pablo II, Pope, 1920-
6  400 10    Jean Paul II, Pope, 1920-
7  400 10    Jan Pawel II, Pope, 1920-
8  400 10    Giovanni Paolo II, Pope, 1920-
9  400 10    Vojtyla, Pope, 1920-
10 400 10    Gruda, Stanislaw, 1920-
11 510 20    Catholic Church. Pope
             (1978- : John Paul II)
12 670       Palumbo, C.E. Cuestiones de
             doc. soc. de la iglesia,
             1982. p. 7 (Juan Pablo II)
```

Figure 1. Authority record for Pope John Paul II.

An authority record carries only the data that applies to a person or organization in the abstract sense, not to any particular printed records, manuscripts, or other artifacts. That is, an authority record is not tied to a bibliographic record, i.e., the description of a particular record series, such as in a MARC AMC record; rather, it applies to *all* records that include information by or about that person or organization. Once the authority record for John Paul II has been created by one archivist, the authority work need never be repeated by another archivist who has access to the same authority file.

An *authority file* is simply a group of authority records arranged in logical order (i.e., alphabetically), and searchable by all established headings and cross-references. In any particular repository or network of repositories, the authority file comprises all of the authority records made by those repositories.

By extracting provenance data from MARC AMC bibliographic records and in-house finding aids and transferring this information into authority records, archivists can create shared authority files and thereby avoid considerable duplicative work. In addition, they can greatly improve the consistency and usefulness of each individual archival catalog, as Szary has noted.[3]

Objectives of the Archives Catalog

To implement an effective authority control system, an archives must first establish the objectives of its catalog. What is the catalog intended to accomplish? What are the significant relationships among the concepts represented in its bibliographic records? Of the voluminous data in collection descriptions, which is most important for retrieval?

The objectives of a library catalog traditionally have been:

- to locate a particular work by author or title,
- to locate all works of an author,
- to locate all editions of a particular work, and
- to locate all works on a particular subject.

Although these objectives offer a useful point of departure, archivists must question whether or not they are also valid for an archives catalog, in which provenance often has a more powerful role than authorship. What types of questions must the catalog be able to answer if it is to serve its users effectively?

Is it true, for example, that most researchers want *all* of the records issued by a particular agency or written on a particular subject? Or are a few random examples usually good enough? Do they most often want a particular "known item," such as the manuscript of Thoreau's *On Walden Pond* or Lincoln's earliest penciled draft of the Gettysburg Address? How important is it to be able to locate all archival materials in a particular form of material (e.g., account books, birth certificates, or ships' logs)? Or those in a particular physical format or medium (e.g., daguerreotypes, handmade paper, engravings, or cloth)? What about all of the records maintained or housed under the authority of a particular government agency, perhaps an agency that has had several different names? Might the activities of a specific *type* of agency sometimes be of interest, particularly to those unfamiliar with agency names and hierarchies? Might certain researchers wish to compare all of the versions of a literary work or law as they were altered over time, including those known by a variety of titles?

The answers to such questions help to determine the types of data that must be included in the archives catalog, as well as to identify ways in which authority control can help make data retrieval more consistent and comprehensive. Although little

research has been done to answer such questions through empirical study,[4] archivists can instinctively define the most frequent types of research questions they are asked in order to begin analyzing ways to improve their catalogs.

Once the objectives of an archives catalog have been determined, the catalog's effectiveness can be evaluated in terms of its ability to deliver high levels of precision and recall. Perfect *precision* results in retrieval only of relevant material. That is, no irrelevant material is retrieved. For example, a precision failure occurs when ambiguous, nonunique names or concepts are not rendered unique by establishing a mutually exclusive name for each. Perfect *recall*, on the other hand, results in retrieval of all relevant documents, or an exhaustive search. Recall failures occur when variant forms of name, such as synonyms, are not linked.

Two examples will help illustrate these concepts. Fig. 2 displays the many versions of the name "John Gardner" found in a particular authority file, and reminds us that each separate heading must be established in a unique form in order to differentiate it from all the others.

```
TO SEE TITLES, TYPE LINE #
AND PRESS RETURN.

1  Gardiner, John.
2  GARDNER JOHANN VON (2)
3  GARDNER JOHN (3)
4  GARDNER JOHN A (2)
5  Gardner, John Consaul.
6  GARDNER JOHN E (2)
7  Gardner, John Edmund Garratt
8  Gardner, John L.
9  Gardner, John Louis, 1929-
10 Gardner, John N.
11 Gardner, John Pugh.
12 GARDNER JOHN WILLIAM (3)
```

Figure 2. Browse display from an online authority file, showing the variant forms of the name "John Gardner" found in a library catalog. Note that headings listed in all capital letters indicate those with multiple occurrences in the file.

The researcher interested in the works of a particular "John Gardner," but unaware of his exact middle initial or date of birth, appreciates the high recall value of this browse display and can proceed down the list of names until the correct Gardner is located. Someone specifically in search of "John William Gardner," on the other hand, finds the browse display carefully subdivided to distinguish this man's works from all others. This searcher can immediately select line 12, and view only the works of the author of interest; this search results in perfect precision. Without authority control, neither researcher could feel confident of dependable results.

Fig. 3 is the authority record for the novelist John Creasey, who wrote under many pseudonyms and in a variety of literary genres. Once this authority record is added to the catalog, a researcher interested in finding the papers of the pseudonymous "Gordon Ashe" can easily determine the author's true identity and locate all of the published and unpublished works of Creasey's numerous literary personae (high recall). Without authority control, the researcher would probably retrieve the published works by Ashe (high precision), but fail to locate both the Creasey manuscripts and those works published under his real name or other pseudonyms (low recall).

```
1  010       n 80057230
2  100 10    Creasey, John.
3  400 10    Marric, J. J.
4  400 10    Halliday, Michael
5  400 10    York, Jeremy
6  400 10    Morton, Anthony
7  400 10    Ashe, Gordon
8  400 10    Hunt, Kyle
9  400 10    Deane, Norman
10 400 10    Frazer, Robert Cain
11 400 10    Manton, Peter
12 400 10    Martin, Richard
13 400 10    Ranger, Ken
14 400 10    Riley, Tex
15 400 10    Reilly, William K.
16 670       His First came a murder, 1935?
17 678       b. 1908, d. 1973.
```

Figure 3. Authority record for the novelist John Creasey, who wrote under numerous pseudonyms.

Again, what do archivists and their users really need? Michelson writes that "known item" searches are much less important in archives than in libraries, so archives should work toward systems that achieve high recall.[5] If this is true, it speaks for a need for systems with excellent links among various forms of name, as well as linked authority and bibliographic files. It is critical that research be done to test this hypothesis, so that archives can with confidence devote their often limited automation resources to developing the most effective, efficient catalogs possible.

How Does Authority Control Work?

Let us now examine the specific means by which authority control improves a catalog's effectiveness:

- vocabulary control,
- collocation, and
- linking relationships.

Vocabulary control seeks to improve the consistency of language in a catalog, so that users can find the materials they seek by using terminology familiar and comfortable to them. Human language is so complex and variable that few concepts are universally known by a single name. Similarly complex and variable are the multiplicity of identities that a single person can assume, as Figs. 1 and 2 amply illustrate.

Svenonius waxes eloquent in her rationale for vocabulary control: "Since the time of Plato, linguists and philosophers have dreamed of a systematic language where every thing is referred to by one word and every word refers to only one thing. Authority control in information retrieval languages is an attempt to realize just such a dream, to untangle the criss-cross relationships between words and their referents. An information retrieval language is, thus, an artificial language; its semantics is considerably more *systematic* than that of a natural language ... any data element which can be ambiguously or variously named is a candidate for authority control."[6]

By systematically identifying the relationships between synonyms and other name variants, authority control results in *collocation,* or gathering together of like materials, such as all the works of an author or all the works on a particular subject. A researcher who inputs any of Creasey's pseudonyms (Fig. 3), for example, will find all of his works arranged in a single file, rather than having to identify and search each name separately.

An authority file also must *establish linking relationships* among separate but related concepts. The Zimbabwe example (Fig. 4) shows the authority file links established among seven different African nations that have occupied the same physical territory at various moments in history, as well as a lengthy note explaining the changes over time.

```
1  010       n 80089993
2  151  0    Zimbabwe.
3  451  0    Zimbabve
4  551  0    Southern Rhodesia
5  551  0    Nyasaland
6  551  0    Northern Rhodesia
7  551  0    Rhodesia and Nyasaland
8  551  0    Malawi
9  551  0    Zambia
10 665       Southern Rhodesia, Northern
             Rhodesia, and Nyasaland were
             united in 1953 to form the
             Federation of Rhodesia and
             Nyasaland; the Federation was
             dissolved in Dec. 1963. On
             July 6, 1964, Nyasaland was
             granted independence and name
             changed to Malawi; on Oct. 24,
             1964, Northern Rhodesia was
             granted independence and name
             changed to Zambia ...
             Southern Rhodesia adopted the
             name Rhodesia and, on April
             18, 1980, the name Zimbabwe
             ... Works by the other 3
             jurisdictions published be-
             fore the changes in 1964 are
             found under Nyasaland,
             Northern Rhodesia, Rhodesia
             and Nyasaland. Works by
             Nyasaland and Northern
             Rhodesia published after the
             changes in 1964 are found
             under Malawi, Zambia.
             SUBJECT ENTRY: Works about
             these jurisdictions are
             normally entered under Malawi,
             Zambia, and/or Zimbabwe,
             regardless of period covered.
             However, works limited in
             subject coverage to ... the
             1953-1963 period are entered
             under Rhodesia and Nyasaland
             ...
11 670       Tel. call to Bur. of African
             Affairs,, State Dept., 4/8/80.
12 670       Wash. Post, 4/18/80
             (Zimbabwe gains independence)
```

Figure 4. Authority record for Zimbabwe (edited for brevity). Note the *SUBJECT ENTRY* note that explains the difference in usage for works about Zimbabwe, as opposed to publications or records "by" Zimbabwe.

The UNESCO record (Fig. 5) includes links made between name and subject authorities. A researcher interested in the subject of "Intellectual cooperation" (line 10) or "International cooperation" (line 11) learns that UNESCO is an organization closely allied with these topics. In both examples, the authority file structure helps researchers locate potentially important new sources of documentation by providing links to concepts related to those being searched in the catalog.

```
1  010        n 79054523
2  110 20     Unesco.
3  410 20     U.N.E.S.C.O.
4  410 20     Unesko
5  410 20     Educational and Cultural
              Organization of the United
              Nations
6  410 20     United Nations Educational,
              Scientific, and Cultural
              Organization
7  410 20     United Nations. Educational,
              Scientific, and Cultural
              Organization
8  410 20     Organisation des Nations
              Unies pour l'education, la
              science et la culture
9  510 20     United Nations Preparatory
              Education, Scientific, and
              Cultural Commission
10 550  0     Intellectual cooperation
11 550  0     International cooperation
12 670        Conference of allied
              ministers of education,
              London. Draft proposals for
              an educational and cultural
              organization of the United
              nations, 1945.
```

Figure 5. Authority record for UNESCO (edited for brevity). Note the cross-references from a higher level of the corporate hierarchy (line 7), as well as the linking reference from an earlier form of name (line 9) and from subjects with which UNESCO is closely linked (lines 10–11).

A word about *keyword searching* is also in order, not because it plays a central role in authority control, but because it is often touted as a panacea that makes authority control unnecessary. In a keyword search of an automated retrieval system, the user inputs one or more terms that are matched against each word in the name, title, and/or subject fields of an online database. Inputting the user's own language may achieve high precision, but if the terms input have variable meanings, problems will ensue. For example, inputting "camouflage" may retrieve materials about both military and biological camouflage, whereas the user may have desired only one or the other.

Even more serious is the fact that keyword retrieval offers no guarantee whatsoever of high recall. Inputting "Vietnam War" retrieves materials with those precise words in their descriptions, but misses those indexed under the official name, "Vietnamese Conflict." Inputting "Opiates," one misses the many materials indexed under the synonym "Narcotics." Keyword searching can work to great advantage if the user's terminology is linked to an authority file; otherwise, it delivers results that are incomplete at best, and highly misleading at worst.[7]

TYPES OF AUTHORITY RECORDS

As mentioned earlier, the data describing each concept placed under authority control resides in an individual authority record for that concept. Each record includes the following essential elements:

1. The *heading*, or "official" form of name established for use in the catalog.
2. *Cross-references* to the established heading from alternative forms of the name (often referred to as "see" references).
3. *Linking references* to related headings that are also established for use in the catalog (often referred to as "see also" references).
4. *Notes* documenting the research and decisions that occurred as the record was created.

In the examples discussed earlier, these elements can be recognized by their respective MARC field tags,[8] as follows:

1xx Heading (e.g., 100, 110)
4xx Cross-references (e.g., 400, 410)
5xx Linking references (e.g., 500, 550)
6xx Notes (e.g., 665, 670)

Authority records are most frequently made for proper names, titles of works, and subjects, but authority files are also becoming more common for more specialized terminology such as forms of material and physical characteristics of library and archival materials. The four basic elements listed above apply to all of these, but certain aspects of the data change from one type to the next.

Proper Names

Personal name authority records are established for names of individual people. Although many people are known by a single name throughout their

lives, a surprising number of variations in a person's name may exist, including:

- multiple linguistic forms or reanimations (Fig. 1),
- a change in status, such as from priest to pope (Fig. 1),
- the use of pseudonyms or nicknames (Fig. 3),
- variant spellings (Fig. 6),
- variations in fullness of name (Fig. 6), or
- a change of name due to marriage, use of a stage name, or any other reason (Fig. 7).

All of these circumstances call for creation of an authority record to link all variant forms in the catalog. Without such links, a researcher's chances of finding all, or possibly any, of the material sought is greatly reduced.

```
1   010        n 79045085
2   100 10     Ford, Ford Madox, 1873-1939.
3   400 10     Hueffer, Ford Madox, 1873-1939
4   400 10     Hueffer, H. Ford, 1873-1939
5   400 10     Huffer, Ford, 1873-1939
6   400 10     Chaucer, Daniel, 1873-1939
7   400 10     Hueffer, Ford Hermann, 1873-
               1939
8   400 10     Hueffer, Ford M. (Ford
               Madox), 1873-1939
9   670        His The Marsden case ...
               1923.
```

Figure 6. Authority record for the author Ford Madox Ford, who lived and published under a variety of names, some with variant spellings.

```
1   010        n 79054611
2   100 10     Ali, Muhammad, 1942-
3   400 10     Clay, Cassius, 1942-
4   400 10     Cassius X, 1942-
5   400 10     X, Cassius, 1942-
6   400 10     Muhammad Ali, 1942-
7   670        His I am the greatest!
               [Phonodisc] 1963.
9   670        Kaletsky, R. Ali and me, c1982
               (a.e.), p. 11 (Cassius
               Marcellus Clay)
```

Figure 7. Authority record for the boxer Muhammad Ali, who changed his name from Cassius Clay.

Corporate name headings are established for political jurisdictions, government agencies, business corporations, churches, clubs, and all other types of organizations. Some of the most common variations in forms of corporate names match those already mentioned for personal names, but they occur for somewhat different reasons:

- the use of abbreviations and acronyms (Fig. 5),
- an official change of name (Figs. 8–9),
- variations in fullness, such as by omission of portions of the corporate hierarchy (Figs. 8–9), or
- the use of popular names of organizations (Fig. 10).

```
1   010        n 81015980
2   110 20     National Museum of American
               History (U.S.)
3   410 10     United States. National
               Museum of American History
4   410 20     Smithsonian Institution.
               National Museum of American
               History
5   410 10     Washington (D.C.). National
               Museum of American History
6   510 20     National Museum of History
               and Technology
7   550  0     Historical museums —
               Washington (D.C.)
8   551  0     Washington (D.C.) — Museums
9   670        Phone call to Museum, 2/12/81
               (name changed in Oct. (?) 1980
               from National Museum of
               History and Technology; still
               considered part of the
               Smithsonian).
```

Figure 8. Authority record for the Smithsonian's National Museum of American History. Note cross-references from the corporate hierarchy (lines 3–4) and geographic place (line 5), as well as the linking reference (line 6) and history note (line 9) explaining the museum's earlier name.

1	010		n 79007487
2	110	20	National Museum of History and Technology.
3	410	20	Smithsonian Institution. National Museum of History and Technology
4	410	10	United States. National Museum of History and Technology
5	410	10	Washington (D.C.). National Museum of History and Technology
6	510	20	Museum of History and Technology (U.S.)
7	510	20	National Museum of American History (U.S.)
8	550	0	Historical museums — Washington (D.C.)
9	550	0	Industrial museums — Washington (D.C.)
10	550	0	Science museums — Washington (D.C.)
11	670		Phone call to Museum, 2/12/81 (name changed in Oct. (?) 1980 to National Museum of American History)

Figure 9. Authority record for the National Museum of History and Technology, former name of the National Museum of American History.

1	010		n 50072583
2	110	20	Watch Tower Bible and Tract Society.
3	410	20	Watchtower Bible and Tract Society
4	410	20	Jehovah's Witnesses
5	670		Encyc. Brit., 15th ed. Jehovah's Witnesses is the popular name of the Watch Tower Bible and Tract Society)

Figure 10. Authority record for the Jehovah's Witnesses. The established heading is the form of name used by the Society, not the popular name.

Corporate names tend to be much more complex than personal names, so it is particularly important that cross-references, linking relationships, and history notes be made as accurate and extensive as possible. These techniques compensate for the fact that libraries often do not include the full organizational hierarchy in many corporate name headings. Many archivists have found this practice irksome, but it should be recognized that thorough cross-referencing renders this problem much less serious than it may seem on the surface.

Geographic place names can also be quite complex, as the Zimbabwe example (Fig. 4) illustrates. Variants of geographic place names occur most commonly through changes in political jurisdiction (Fig. 4) or variant linguistic forms (Fig. 11).

1	010		sh 86001807
2	151	0	Wienerwald (Austria)
3	451	0	Vienna Woods (Austria)
4	451	0	Wiener Wald (Austria)
5	550	0	Forests and forestry — Austria
6	670		Lippincott (Wiener Wald)
7	670		Web. geog.
8	670		Old catalog heading (Wienerwald)

Figure 11. Subject authority record for Vienna's Wienerwald, with cross-references from the English form of its name and a variant spelling in German.

For well over a century, proper names for people, corporate bodies, and geographic places have been established by library catalogers using standard rules of description. At this time, the most widely-used source of rules for establishing proper names for library catalogs is the *Anglo-American Cataloging Rules*, 2nd edition, commonly referred to as AACR2[9]. The usefulness of AACR2 for archival cataloging is discussed later in this paper.

Titles of Works

Library catalogers establish *uniform titles* for works that are known by more than one title. Although authority control of titles is sometimes considered irrelevant for archives, it would be more accurate to say that it is necessary much less frequently in archives than in libraries. Materials in archives and manuscript repositories that would benefit from uniform title authority control might include:

- literary manuscripts whose authors' changed their titles, or that were known by different titles before publication or in ain a revised edition,
- translations,
- laws, bills, and other documents known by popular names,

- newspapers that changed titles, and
- works of art known by more than one title.

Library catalogers establish uniform title headings using AACR2, just as they do for proper names. An example of a uniform title authority record is shown in Fig. 12.

```
1  010       n 79029194
2  110 10    United States. Declaration of
             Independence
3  430  0    Declaration of Independence
4  510 10    United States. Laws, statues,
             etc.
```

Figure 12. Authority record for the Declaration of Independence, with the governmental jurisdiction (United States) entered as "author" to indicate that this an official U.S. document.

Subjects

The usefulness of subject indexing of archival collections is currently of great interest to many archivists. The issues are complex. How to select which concepts to index from among the thousands represented in a large collection? How specific or broad should the concepts be? Is there a reasonable maximum number of terms that should be assigned to a particular collection? How will archivists decide which thesaurus to use? If a library-oriented thesaurus is selected, how can it be adapted to formulate headings that are meaningful for indexing original historical documents?

Answering such questions is far beyond the scope of this paper, but a brief outline of the nature of subject authority records is not. Subject authority records contain the same record elements already seen for names, i.e.:

150Heading
450 Cross-references
550 Linking references
6xx Notes

Additionally, a standard thesaurus notation has been recommended by the American National Standards Institute (ANSI)[10] for use in displays of subject authority records, as follows:

- USE leads from unused synonyms and other variants to the term used as the established heading;
- UF (used for) is the reciprocal of the USE reference and accompanies the term to which the USE reference refers;
- BT (broader term) refers from a term for a member of a class to the term for the class (a hierarchical relationship);
- NT (narrower term) refers from a term for a class to the term for one of its members (a hierarchical relationship, and the reciprocal of BT);
- RT (related term) is used between terms with a relationship that is not hierarchical when it is helpful to link two closely related concepts.

Fig. 13 is a subject authority record for the topic Narcotics.

```
Narcotics
UF Opiates
NT Cocaine
   Hashish
   Heroin
RT Sedatives
   Stimulants
```

Figure 13. Subject authority record for Narcotics.

Is authority control of subject terminology necessary? The answer must be an emphatic "yes." The terminology of subjects, as well as the relationships among them, are vastly more complex than those of names. Our language is riddled with synonyms and near synonyms that make subject information retrieval extremely difficult. Consider, as a very simple example, the many ways in which a researcher might approach an online catalog in search of documents relating to World War II: World War II, World War 2, the Second World War, and WWII are just some of the variant forms given this topic that must be linked if searchers (and indexers) are to expect high recall.

The general subject heading list used most widely in U.S. research library catalogs is the *Library of Congress Subject Headings*,[11] commonly known as LCSH. The usefulness of LCSH for archival cataloging is discussed later in this paper.

Form of Material and Physical Characteristics

In addition to proper names, titles, and topical subjects, two additional types of terminology must be brought under authority control in archives catalogs.

Form of material terms, sometimes called "genre" headings, denote categories of materials that are

interesting as examples of an intellectual, literary, or pictorial form. Sample terms include Advertisements, Broadsides, Cartoons, Greeting cards, Logbooks, Passion plays, Sermons, and Yearbooks. (These terms are coded in MARC bibliographic records as field 655.)

Physical characteristics terms, on the other hand, apply to the features of an actual physical object. What is it made of? By what means was it produced? Does it represent a finished product or a stage in production? Who has owned it? What condition is it in? Sample terms include Authors' copies, Daguerreotypes, Facsimiles, Galley proofs, Scrolls, Tintypes, and Watermarks. (These terms are coded in MARC bibliographic records as field 755.)

Fig. 14 is an authority record display for the form of material Advertisements, and Fig. 15 is the record for the physical medium Tintypes.[12] A variety of relevant thesauri are mentioned later in this paper.

```
Advertisements
PN Public notices of the availability of
   goods or services through purchase or
   other means.
CN Also index under BROADSIDES, HANDBILLS,
   POSTERS, or other appropriate form.
BT Ephemera
NT Advertising cards
   Advertising mail
   Display cards
RT Fashion photographs
   Fashion prints
   Premiums
```

Figure 14. Thesaurus notation display of authority record for the form of material "Advertisements." "PN" is a public note, and "CN" is a cataloger's note.

```
Tintypes
PN Direct-image photographs in which the
   collodion negative, supported by a
   dark-lacquered thin iron sheet, appears
   as a positive image. Popular mid-1850s
   through 1860s; in use through 1930s.
   Usually small portraits.
CN Used in a note under PHOTOGRAPHS.
UF Collodion positive photographs
   Ferrotypes
   Iron photographs
   Melainotypes
BT Photographs
```

Figure 15. Thesaurus notation display of authority record for the physical characteristic term "Tintypes." "PN" is a public note, and "CN" is a cataloger's note.

Archival Possibilities

Archivists are beginning to consider new possibilities for archival authorities that go beyond those used in libraries, both in terms of using authority files as reference files, and in considering new types of access points that are important in an archives.[13]

Two types of data that have received great attention are occupations and functions. The MARC AMC format includes fields for entering such data (fields 656 and 657, respectively), but might this data not also be appropriate for inclusion in authority records? The MARC authorities format would have to be expanded to carry such data, but as archivists develop and implement thesauri for indexing occupations and functions, enhancing the format may become part of the national archival automation agenda.

AUTOMATION OPPORTUNITIES

As mentioned earlier, authority control was in widespread use before the age of automation, but computerized information retrieval systems now offer more flexible and efficient opportunities for authority control than those possible in manual systems.

Automation's most significant achievement to date may well be that it provides strong *incentives for the use of standards.* Automated systems make it much more feasible for repositories to share data, and as they do begin to do so, archivists frequently are finding it impossible to interpret another archives' records if they do not follow the same descriptive

standards. Since most computer searching is utterly literal, the slightest variation in terminology, spelling, or the form of a person's name means that a heading meaningful to one archives cannot be retrieved online by its neighboring repository. The incentive exists, therefore, for shared authority systems in addition to shared databases of MARC AMC records.

Let us examine a variety of techniques available in automated catalogs for increasing the efficiency of authority control systems:

Linking files: By providing conceptual links between authority records and bibliographic records, corrections or changes in authority file headings can automatically identify the bibliographic records that also must be changed. Even more important, user search success increases greatly when mapped through the authority file instead of leading directly to the bibliographic file.

Validation of headings: As new bibliographic records are added to an online catalog, an authority control system can generate reports that identify all headings appearing for the first time. These may include errors (such as conflicts with existing cross-references, or simple misspellings), as well as legitimate new headings for which authority records must be created. Catalogers must still perform much of the intellectual work of evaluating the nature of the new headings, but the system provides them with a tremendous head start.

Global change: Some changes in authority records generate a need for hundreds or even thousands of bibliographic record changes. Imagine, for example, how many library records were affected when usage changed "Aeroplanes" to "Airplanes." Global change capability allows a cataloger to identify all relevant records and update them quickly.

Browse screens: Some online catalogs display all search results in a browse screen such as that shown in Fig. 2. This makes it possible for both catalogers and users to see and compare many headings at a glance. It also makes certain types of cross-references unnecessary.

Transparent authorities: Through an online system's interpretation of cross-references, a user's search query is translated into the established form of a name or concept. Existing systems usually perform this translation and then switch to display the established form. In a transparent authority system, however, the form input by the user is displayed, thereby eliminating any need on the user's part to understand that one form is considered authoritative. One longs to see this approach adopted widely so that catalog users no longer have to be befuddled at the appearance of the "best" form of heading as established in the catalog.

General reference files: Online dictionaries of common spelling variants can eliminate the need for many cross-references and reduce user frustration that results from inputting misspelled words, legitimate variant spellings, abbreviations, acronyms, and the like.[14] Some examples of common variants include "U.S." vs "United States," "Mac" vs "Mc," and "Alan" vs "Allen."

Unfortunately, most online catalog systems do not yet offer all of these features. Many libraries and archives are in a period of transition from manual to automated systems, and systems must be marketed and sold long before they are "perfect." Numerous institutions are opting to purchase an imperfect system, however, rather than continue pouring resources into antiquated, labor-intensive activities such as typing and filing cards. Archivists should not let the imperfections of existing systems discourage them. Rather, they should insist upon development of additional features in online catalogs so that they can help archives meet their goal of helping researchers successfully locate the materials they need.

WHAT CAN ARCHIVES BORROW FROM LIBRARIES?

Library catalogs have long incorporated principles of authority control to make their catalogs useful to researchers, and a tremendous infrastructure of automation formats, cataloging rules, subject thesauri, and bibliographic networks are in place and working effectively.

These library applications do not necessarily translate perfectly into the archives environment, but it is important that archivists learn and borrow what they can in order to avoid "re-inventing the wheel" regarding information retrieval. Building systems from scratch is a daunting task; if archivists can save their considerable energies for adapting and improving library applications for archival use, perhaps sufficient human and institutional resources will remain for creating *new* types of applications, such as a functions vocabulary, which do not exist in the library world.

With these thoughts in mind, let us examine some library applications relating to authority control.

MARC Format for Authorities

The MARC AMC (Archival and Manuscripts Control) format has gradually gained wide acceptance among archivists as the standard format for computerized storage and exchange of bibliographic record descriptions or archival and manuscript materials. Authority records are *not* stored in the MARC AMC format; the *USMARC Format for Authority Data*[15] is used for storage and exchange of authority records, serving as a companion to

MARC AMC and the other MARC bibliographic formats.

Just as the MARC AMC format consists of a set of field names and numerical field tags for bibliographic record descriptions, the authorities format consists of the same elements for authority records, as already seen in the section on authority records and in the figures that appear throughout this paper. The same format is used for all types of library authority records, including names, subjects, titles, forms of material, and physical characteristics.

A variety of descriptive standards (such as AACR2, discussed below) may be used for formulating data input into the authorities format. Codes are assigned to widely-used standards, making it possible for a repository to know whether or not the cataloging agency that input a particular authority record uses the same standard as the home repository.

The authorities format includes fields for notes explaining the source of a heading or cross-reference, background or history information, scope, and usage. Libraries tend to use these fields somewhat sparingly because of the labor-intensive effort involved in researching and composing such notes. For archives, information such as an agency history note is important enough to warrant the additional effort. Archivists could potentially make a significant contribution to network authority files by adding history notes and additional cross-references to existing authority records, particularly if they adopt existing standards such as AACR2.

AACR2

The *Anglo-American Cataloguing Rules,* 2nd edition (AACR2), provides detailed instructions for formulating the data that catalogers enter into bibliographic records and authority records. This is quite different from the purpose of the MARC formats, which simply provide standard field names and numerical field tags, but offer no guidance in selecting the main entry for a collection, devising a scope and contents note, formulating a name, or any of the other descriptive tasks that catalogers undertake. Hensen[16] argues persuasively for the use of standards such as AACR2 for formulating the data that is supplied in MARC AMC records.

AACR2 does *not* include instructions for establishing subject headings or classification numbers. It is, however, a nearly universal standard for description of bibliographic materials in research libraries throughout the English-speaking world, and its widespread application greatly enhances the consistency of the catalog records and authority records available in these many institutions.

AACR2 includes rules for formulating headings, selecting the names and titles in a bibliographic record that should be indexed, and describing various forms of material. For archives cataloging, the most important rules are those in Chapters 22-26 for formulating names for persons, geographic places, corporate bodies, uniform titles, and cross-references. AACR2 provides guidelines for resolving dilemmas such as:

- distinguishing between identical names in order to create unique headings,
- selecting the most predominant or common form of name,
- treatment of fictitious names and pseudonyms,
- choice of entry element for persons with multiple surnames,
- standard forms for complex names of royalty, clergy, and government officials,
- special rules for treatment of names in particular languages,
- special rules for distinguishing government bodies from other organizations, and
- devising cross-references to relate variant forms of the same name.

Some archivists have complained bitterly about aspects of AACR2, such as its rule on the use of the most common form of a name instead of the full name (e.g., "Mozart, Wolfgang Amadeus" instead of "Mozart, Johann Chrysostom Wolfgang Amadeus"), as well as direct entry of corporate names under the most distinctive aspect of the name instead of using full corporate hierarchies (e.g., "Anacostia Neighborhood Museum" instead of "Smithsonian Institution. Anacostia Neighborhood Museum").

The form of a particular heading, however, is much less significant than the researcher's ultimate ability to locate required documents. How many researchers have the need to know an organization's full corporate hierarchy? Some certainly do, but many others require only the records of a particular subagency and are not interested in the full organizational hierarchy. Use of a standard set of rules for devising headings, in concert with a carefully devised system of cross-references (in which the full hierarchy *can* be stored, for those who need it) and linking relationships, makes it possible for a user to approach the catalog with *any* form of a name — and be led to relevant material. A good authority control system is designed to compensate for the impossibility of everyone agreeing on the same "best" form of any name.

Rather than pressing for development of an entirely new set of rules for establishing names, or adopting no uniform standard at all, archivists might consider participating in the ongoing AACR2 revision process. AACR2's rules are based primarily on the types of data found in published books. Archives and

manuscript materials, as well as other media such as graphics, films, maps, and computer software, are presenting challenges to traditional library practice, but no concerted effort has yet been launched to alter the rules in response to the need.

AACR2 also includes rules for *selecting* which names should be used as access points for a particular bibliographic record. These rules are completely book oriented, however, and most are not relevant for cataloging collections of archives and manuscripts materials. Archivists should consider, however, whether development and use of a similar set of rules would improve the consistency of archives cataloging.

Most archivists do not find AACR2's rules for *description* of materials very useful, but instead prefer Hensen's *Archives, Personal Papers, and Manuscripts,*[17] which expands and improves upon AACR2, Chapter 4, for cataloging of original historical materials. For catalogers of visual materials such as original photographs, prints, and drawings, Betz's *Graphic Materials*[18] is the standard source of rules for description.

LCSH

The *Library of Congress Subject Headings,*[19] or LCSH, is a general subject heading list of enormous depth and breadth. It is edited and published by Library of Congress experts in subject cataloging and is available in bound volumes, microfiche, and computer tape. LCSH is used almost universally by U.S. research libraries; researchers can therefore depend on finding similar terminology used in many different institutional catalogs throughout the country.

LCSH does have its problems. Librarians and archivists both have many complaints about choice of terminology, the speed with which new concepts are added to the list (often very slowly), and its confusing guidelines for use. For archivists, LCSH has an additional drawback: its rules of application are oriented toward the indexing of single published texts, not collections of original historical documents. Drawing on concepts introduced earlier, however, consider several reasons why archives may find it desirable to adopt LCSH for their own subject indexing:

- LCSH includes thousands of terms for subjects in all disciplines.
- Although agreement on specific terminology is next to impossible, cross-references can be enhanced locally to respond to this problem.
- Development and maintenance of a local thesaurus is a huge undertaking that few archives can afford.
- LCSH is in widespread use, is well supported by the Library of Congress, is updated regularly, and is easily available in a variety of formats.

As already mentioned in the context of AACR2, it may be much more effective for archivists to devote their energies to developing a set of guidelines for applying LCSH that would serve the content indexing needs of archival and manuscript collections and to lobby for the addition of appropriate new terms and subdivisions to LCSH. Starting from scratch would be a daunting proposition, and it is probably quite unnecessary.

In addition to LCSH, many specialized thesauri exist that focus on a single discipline, subject area, or medium. One such subject thesaurus of possible interest to archivists, especially those with extensive collections of visual materials, is the *LC Thesaurus for Graphic Materials.*[20] Other thesauri are available for indexing forms of material[21] and physical characteristics of published texts[22] and graphics.[23]

Bibliographic Networks

Throughout the library world, networks have become indispensable for sharing bibliographic data and authority records associated with cataloging, as well as for exchanging collection holdings information for reference purposes. The days when librarians eagerly awaited monthly, quarterly, or annual supplements to the *National Union Catalog* for updates on the holdings of neighboring institutions seem like ancient history; those libraries which participate in online bibliographic networks have been thoroughly spoiled.

There are numerous advantages to membership in a bibliographic network, be it one of the major national networks (such as RLIN, OCLC, and WLN) or a smaller regional, local, or institutional network. Some of these advantages are:

- access to shared bibliographic data for cataloging and reference;
- the opportunity to see how other institutions use and interpret the MARC AMC format, AACR2, and other standards;
- availability of derivative cataloging, supplied by other archives and libraries, for microforms and other nonunique media;
- access to online authority control standards such as the Library of Congress name authority file and LCSH;
- lower cataloging costs due to use of derivative cataloging and authority records;
- user groups for discussion of mutual problems and solutions; and
- access to a variety of user support services, including development and maintenance of hardware and software, system documentation, and staff training programs. (The advantages of these services

should not be underestimated; each is extraordinarily expensive when handled completely in-house.)

And, of course, there are the trade-offs:

- the inevitable compromises that occur when more than one group of users share the same system;
- the major networks have many more libraries than archives as members, although they usually are willing to be responsive to demands voiced by their members (RLG's RLIN system, for example, has been very responsive to archivists' needs);
- the priorities given to particular system enhancements may not always be the same as those preferred by a single institution;
- networks usually designate particular bibliographic standards that all members must use, such as the MARC formats, AACR2, and LCSH; and
- membership and telecommunications fees may be high (but are usually much lower than the costs of building and maintaining an independent in-house system).

CONCLUSION

The ultimate goal of the archives catalog is, of course, to improve service to the users of archival repositories. By providing consistent terminology, collocating similar materials, and establishing links between related materials, authority control helps achieve this goal.

Consider the following as you work toward a decision on whether or not your archives can "afford" to implement authority control: Without authority control, the burden is on the user to know when a search has failed — and how does the archives user know whether the system has failed, or whether the material sought simply is not in the repository? How does the user know when to end the search? As Malinconico reminds us: "It seems to be a common human tendency to assume that if one has retrieved a number of like items, but not the one sought, this is so because it simply does not exist."[24]

Why not place the burden of authority control on catalogers rather than on researchers? What researcher comes to the archives catalog armed with a literary author's every pseudonym, a country's history of changing political boundaries, or a government agency's seemingly whimsical changes of name? Avram states it well: "Rather than each user having to individually establish the relationship (or fail to do so and not obtain material), the cataloger does it once."[25]

The likelihood of failure in a catalog search is high if authority control is left to the researcher, who is unlikely to consider all possible forms of name or synonyms. The user certainly cannot be expected to know the catalog's "rules." For both researchers and archives staff, there is great value in predictability, in knowing that one will be able to find related material described consistently throughout a repository's catalog.

And why in only a single repository's catalog? As authority control standards and files become widely shared among multiple archives, each cataloger's work will become increasingly more cost effective, and each researcher's success will come with less frustration and wasted time. Time will prove authority control an excellent investment.

NOTES

1. Avram, Henriette. "Authority control and its place." *Journal of Academic Librarianship 9* (Jan. 1984), p. 331. Avram provides an excellent introduction to authority control concepts, as do the following: Auld, Larry. "Authority control: an eighty-year review." *Library Resources and Technical Services 26* (Oct/Dec 1982), p. 319-330. McDonald, David R. "Data dictionaries, authority control, and online catalogs: a new perspective." *Journal of Academic Librarianship 11* (1985), p. 219-222. Schmierer, Helen F. "The relationship of authority control to the library catalog." *Illinois Libraries 62* (Sept. 1980), p. 599-603.

2. *USMARC Format for Bibliographic Data: Including Guidelines for Content Designation.* Prepared by the Network Development and MARC Standards Office of the Library of Congress. Washington, D.C.: Cataloging Distribution Service, 1988. The MARC format provides a standard way of coding bibliographic data for automated storage and retrieval; the AMC format is the MARC format used for storing archives and manuscripts records.

3. Szary, Richard V. "Expanding the role of authority files in the archival context." Paper presented at Society of American Archivists annual conference, November 1985, p. 16.

4. Michelson, Avra. "Description and reference in the age of automation." *American Archivist 50* (Spring 1987), p. 193.

5. Michelson, ibid., p. 199.

6. Svenonius, Elaine. "Authority control: what and for whom." In: *Authority Control Symposium, Occasional papers no. 6.* Papers presented during the 14th Annual ARLIS/NA Conference, New York, N.Y., February 1986. Karen Muller, ed. Tucson, Ariz.: Art Libraries of North America, 1987, p. 3.

7. For a more detailed discussion of the nature of

keyword searching, refer to: Jamieson, Alexis, Elizabeth Dolan, and Luc Declerck. "Keyword searching vs. authority control in an online catalog." *Journal of Academic Librarianship 12* (Nov. 1986), p. 277-283.

8. *USMARC Format for Authority Data.* Prepared by the Network Development and MARC Standards Office of the Library of Congress. Washington, D.C.: Cataloging Distribution Service, 1987. Full MARC coding has not been used in Figures 1-15 for the sake of brevity. Refer to the *USMARC Format* for additional details of coding, as well as for an explanation of codes shown in figures but not explained in the text.

9. *Anglo-American Cataloguing Rules.* 2nd ed. Chicago: American Library Association, 1978.

10. *American National Standard Guidelines for Thesaurus Structure, Construction, and Use.* New York: American National Standards Institute, 1980.

11. Library of Congress. Subject Cataloging Division. *Library of Congress Subject Headings.* 10th ed. Washington, D.C.: The Library, 1986.

12. The record displays for both terms are from: Zinkham, Helena and Elisabeth Betz Parker. *Descriptive Terms for Graphic Materials: Genre and Physical Characteristics Headings.* Washington, D.C.: Cataloging Distribution Service, 1986.

13. These three articles are particularly recommended: Bearman, David, and Richard Lytle. "The power of the principle of provenance." *Archivaria 21* (Winter 1985/86), p. 14-27. Bearman, David, and Richard Szary. "Beyond authorized headings: authorities as reference files in a multi-disciplinary setting." In: *Authority Control Symposium, Occasional papers no. 6.* Papers presented during the 14th Annual ARLIS/NA Conference, New York, N.Y., February 1986. Karen Muller, ed. Tucson, Ariz.: Art Libraries of North America, 1987, p. 69-78. Evans, Max. "Authority control: an alternative to the record group concept." *American Archivist 49* (Summer 1986), p. 249-261.

14. Taylor, Arlene G. "Authority files in online catalogs: an investigation of their value." *Cataloging and Classification Quarterly 4* (Spring 1984), p. 16.

15. *USMARC Format for Authority Data,* 1987.

16. Hensen, Steven L. "The use of standards in the application of the AMC format." *American Archivist 49* (Winter 1986), p. 31-40.

17. Hensen, Steven L. *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries.* Washington, D.C.: Library of Congress, 1983. (A 2nd edition of Hensen is in progress.)

18. Betz, Elisabeth W. *Graphic Materials: Rules for Describing Original Items and Historical Collections.* Washington, D.C.: Library of Congress, 1982.

19. Library of Congress, *Library of Congress Subject Headings,* 1986.

20. *LC Thesaurus for Graphic Materials: Topical Subjects.* Comp. by Elisabeth Betz Parker, intro. by Jackie M. Dooley. Washington, D.C.: Cataloging Distribution Service, 1987.

21. The following form of material thesauri are available: *Genre terms: A Thesaurus for Use in Rare Book and Special Collections Cataloguing.* Compiled by the RBMS Standards Committee. Chicago: Association of College and Research Libraries, 1983. Hickerson, Thomas and Elaine Engst. *Form Terms for Archival and Manuscripts Control.* Stanford, Calif.: Research Libraries, Group, 1985. Zinkham and Parker, *Descriptive Terms for Graphic Materials,* 1986. In addition, Harriet Ostroff, editor of the *National Union Catalog for Manuscript Collections,* has begun work on an extensive form of material thesaurus for archives and manuscripts terms.

22. *Printing and Publishing Evidence: Thesauri for Use in Rare Book and Special Collections Cataloguing.* Compiled by the RBMS Standards Committee. Chicago: Association of College and Research Libraries, 1986.

23. Zinkham and Parker, *Descriptive Terms for Graphic Materials,* 1986.

24. Malinconico, S. Michael. "Bibliographic data base organization and authority file control." In: *Authority Control: The Key to Tomorrow's Catalog,* Mary W. Ghikas, ed. Phoenix, Ariz.: Oryx Press, 1982, p. 4.

25. Avram, "Authority control," p. 333.

AUDIENCE DISCUSSION

AUDIENCE SPEAKER:

I have a practical question about the fixed fields at the top of MARC records. Can you do anything with them?

JACKIE DOOLEY:

All MARC records are chock full of data which may on the surface appear to be absolutely useless, but they are, in fact, important. In bibliographic records, for example, there are codes to indicate whether or not a publication consists of conference proceedings; whether or not it is fiction; what languages the text is written in, and what the dates of publication are. Such codes are most frequently used in retrieval systems to limit searches. It is much more efficient for a system to do so using fixed field data than it would be reading variable-length data.

In the authorities format, the fixed field codes are used to store a variety of codes in shorthand form. There are codes to indicate which cataloging rules were used to formulate the heading, the nature of its cross references, whether or not the heading is valid for use as a subject, and many others. David, can you think of additional ones?

DAVID BEARMAN:

I was going to suggest a partial answer to that.

One of the things that Jackie mentioned, and I think is going to be a theme throughout today's sessions, is the trade-off that is involved in trying to share information of any sort.

Obviously, the reason for sharing information is the promise that somebody else will do your work for you. If, in fact, we deal with the same universe of institutions, of forms of material, of functions, of people, of subjects, of physical types, then there is really some promise to that hope.

But on the other hand, there is an overhead involved, and the overhead here is the overhead of maintaining automated files. Most of the unreadable stuff that people put in that makes us wish we didn't have to deal with something like the MARC format is for machine manipulation of the data into files in which it will reside so that somebody else can use the program.

That overhead is real. And, quite frankly, if the overhead is greater than the benefits of the sharing, then we would all be damned stupid to be sharing authority data.

That's one of the things that we have to determine, not just in the course of today as we work on questions of what authority control might be for, but in the course of actually planning the use of these systems. Is the payoff big enough. Very fundamental question.

RICH SZARY:

I think when we talk about overhead, I agree with the distinction David made about having other people do your work for you. If you can use an authority file that someone else has constructed and maintained, then you don't have to mess with most of that information. People that maintain it will take care of the fixed field information for you. Most users, whether catalogers or patrons, are interested primarily in the headings, possibly some of the note material, and some of the explanations.

ROBERT HARDING:

In the Gardner record, I noticed that some of the headings were capitalized and some of them weren't. We have the same situation on SIBIS. I am wondering why? Is this a universal problem?

JACKIE DOOLEY:

No, it's not a problem, it's just an OCLC display technique which lists all of the headings your search retrieved. Where headings are capitalized, this is to indicate an additional level of listing below the one you see on the screen. In other words, there are multiple occurrences of that heading in the index. It's just a devise to cue you.

DAVID BEARMAN:

In this case, the user interface clues you, through capitalized headings, to the fact that there are multiple occurrences of one form.

But I think the second point is that vocabulary control deals both with the trivial and the profound from misspellings, miscapitalizations, mispunctuations, misspacing, to true intellectual variance. The areas that we're dealing with here in talking about authority control span that entire range.

One of the things that came out in the survey of the SIBIS users, was a realization that it might be possible to solve some of the problems that afflict the catalog through the resolution of some relatively trivial types of control, which require a concomitantly low intellectual investment. One of the issues before us is how much authority control work and how much intellectual investment in the building of reference files needs to take place to achieve what level of improvement in ultimate retrieval.

This is one of these issues that will be with the group as it starts to look at practical steps to take but the order of those steps clearly is that if you can clean up a substantial number of problems by correcting your misspellings or by linking misspellings, collocating misspellings, then you can achieve some benefits without substantial investments. And that's an important payoff.

JACKIE DOOLEY:

One thing that's worth mentioning, and I don't think I said, was that display conventions are system specific at present. There are efforts being made to devise standardized display conventions so that from one system to the next you can expect similar displays.

The MARC format does not mandate display conventions. You have many choices. Some systems do it much more clearly and better than others.

JOHN FLECKNER:

Is there software available that runs on small computers, microcomputers, that controls thesauri?

JACKIE DOOLEY:

There are a variety of them. One would be hard-pressed to control a file the size of the Library of Congress authority file on a small system, but, there are software packages emerging for controlling subject thesauri online.

The Prints and Photographs Division of the Library of Congress and the General Accounting Office both use one called LEXICO, which does automatic validation of headings.

It's a reasonably decent system, although it has a long way to go in terms of flexibility of input, and the ease with which you can modify records. And, unfortunately, not compatible with the MARC format. So, at present, the thesaurus used by the Prints and Photographs Division cannot be linked online with the division's bibliographic records in the MARC Visual Materials format.

DAVID BEARMAN:

We are running out of time here for questions. Let me just make one last comment because I think it carries us into the rest of the morning.

As we move along this morning and think about systems to do authority work, it's important to recognize that they can help you in a lot of different ways. They can help you to maintain a vocabulary, which is different than helping you necessarily collocate, which is different, in turn, from actually linking within the authority, which is certainly different from the kind of transparent referencing, user assistance, intellectual aid kinds of capabilities that we would want to see in a fully linked and automated authority file.

So, we are going to keep sorting out as we move through the day the different levels of system functionalities, as well as the different types of things that we're trying to do with authorities.