

DO WE NEED AUTHORITY CONTROL? INVESTIGATIONS AT THE MILWAUKEE URBAN ARCHIVES

MARK A. VARGAS

ABSTRACT: Although authority control is an issue of growing importance to archivists, little research has been done to investigate its necessity. As an initial step, the Milwaukee Urban Archives conducted an experiment to answer one specific question: how often do names used by the MUA match those already in the Library of Congress Name Authority File (LCNAF) for the same person or organization? The research focused on the potential need to create authority records for personal and corporate names, used either as creators of collections or as added entries. The results show that 85 percent of the names did not exactly match those already in the LCNAF, including 34 percent which conflicted. Only 15 percent of the names matched the LCNAF exactly, meaning that no conflict could occur. The study concludes with observations on the need for further research.

Creating bibliographic records for archival collections in databases such as OCLC or RLIN has forced archivists to reexamine many traditional practices, because the networks required adherence to cataloging standards.¹ Although the issue of whether or not to automate² bibliographic records is now moot, there are still many problems to be resolved. One of these problems is authority control.

This study uses the following definitions:³

Authority control: A bibliographic organization function that ensures the establishment of logical links between authority records and access points to bibliographic records.

Authority file: A set of records maintained that documents the established forms of entries used in public catalogs, including the references that support them.

Authority work: The process of checking headings against an authority file and verifying that the form of entry matches the form already in use in the catalog and that linking references are made for headings requiring them.

Need for the Study

The need for understanding authority control in an archives has become a matter of great importance, because archivists will have to comply with the respective library rules. Or should they? How much effort should archivists put into authority control, and when should authority records be created? Unfortunately, the archival literature is almost devoid of any research data which might offer assistance in answering these questions, and little factual data provides any guidelines.⁴

Some persons might question whether authority control is needed in an era of powerful computers, boolean searching, and truncation. The results of numerous information science studies are mixed, although there appears to be general agreement that some sort of authority control is needed.⁵ In the archival literature, Bearman, Evans, and Szary each recommended the creation of relatively elaborate authority records which could provide extensive information links.⁶ In *Authority Control: A Manual for Archivists*, Elizabeth Black presumed that authority control is necessary.⁷

In the library profession, the whole premise behind sharing online data is to reduce duplication of effort. Similarly, archival authority work should not be done only at the local level, because it will probably lead to greater costs and confusion in the national database.⁸ For example, it is just an assumption that only the Milwaukee Urban Archives (MUA) has some papers of Joseph Johnson (a 19th century farmer). How is the MUA—and a patron—to know if the papers of the Johnson family in some other city do not contain information on the MUA's Joseph Johnson? If archivists cannot, or will not, do the authority work, patrons will have to.⁹ Without authority control, patrons are forced to repeat the steps that the archivists avoided in the first place. Perhaps this is not an entirely bad idea, especially from the perspective of reducing costs, but it is a gamble on the search skills of the patron, and studies have revealed the difficulty patrons have in manipulating online searches and evaluating the results.¹⁰

It is probable that few archives can afford to create an authority record for each new name added to the catalog databases. Archivists will probably have the resources for creating only selective records, largely because of costs. Although the figures are not entirely clear, it certainly takes a great deal of staff time to check the authority files or prepare an authority record. The State Historical Society of Wisconsin (SHSW), University of Wisconsin-Madison, and University of Wisconsin-Milwaukee do not create authority records because of the high costs. Koel stated that the authority work done at Yale took 5-10 times as long as the description of a work.¹¹ A 1986 Library of Congress survey of National Coordinated Cataloging Operations (NACO) participants revealed that the average cost per authority record was \$14.67.¹² The RLIN RECON AMC Project participants found that authority work was the second greatest factor in slowing the creation of bibliographic records.¹³ On the other hand, what is the cost if a patron cannot find the bibliographic record because the catalog lacks name authorities or cross references?

The Research Study

The research project began after the MUA had prepared MARC records for over 600 collections. The staff recognized that some degree of authority work

seemed necessary, but was unsure how much effort should go into it. The staff decided to conduct an experiment in October and November 1991 to answer one specific question: How often do names used by the MUA match those already in the Library of Congress Name Authority File (LCNAF) for the same person or organization? The research focused on the potential need to create authority records for personal and corporate names, used either as creators of collections or as added entries. It did not address issues of creating cross-references, subject or title authority control, nor the use of authority records as extensive reference files, as recommended by Bearman, Evans, and Szary.¹⁴

The MUA staff began the research project with names used in its MARC AMC records. The MUA consists of two archival agencies. The first is the university archives, which acquires the records of the University of Wisconsin-Milwaukee. The second is the Milwaukee Area Research Center (ARC), one of thirteen branches of the State Historical Society of Wisconsin. The MUA staff had just completed creating 523 MARC records for the ARC, a project that took about two years to complete.¹⁵ The SHSW had created an additional 120 MARC records for the Milwaukee ARC, but the MUA staff often augmented them, including a larger number of personal and corporate names as added entries. The staff checked all corporate and personal names against the LCNAF, available through the NOTIS system at UW-Milwaukee. When no previously established heading existed for the name in the LCNAF, the MUA used the fullest form of the name in the MARC records.¹⁶ Corporate bodies created 409 (64 percent) of the 643 collections; individuals or families created 234 (36 percent).

For the purposes of this study, the names were rechecked against the LCNAF, and each fell into one of twelve categories:

Exact LCNAF match, used only *once* in the collections

1. Corporate name
2. Personal name

Partial LCNAF match, used only *once* in the collections

3. Corporate name
4. Personal name

No LCNAF match, used only *once* in the collections

5. Corporate name
6. Personal name

Exact LCNAF match, used *more than once* in the collections

7. Corporate name
8. Personal name

Partial LCNAF match, used *more than once* in the collections

9. Corporate name
10. Personal name

No LCNAF match, used *more than once* in the collections

11. Corporate name
12. Personal name

An *Exact* LCNAF match occurred when the name used by the archives was already established in the LCNAF for the same individual or corporate body. A *Partial* LCNAF match occurred when the name used by the archives was not established by the LCNAF, but did partially match a name for a different person or organization. For personal names, a partial match included a similar surname and forename, or first initial when used. For corporate names, a partial match

included at least the first two words in the name. A *No* LCNAF match occurred when the name used by the archives was not established by the LCNAF, and did not partially match another name. Table 1 shows the number and relative percentage of the names in each category.

The results in Table 1 show that 85 percent of the names did not exactly match those already in the LCNAF. This probably reflects the fact that few materials by or about the organizations or individuals represented in the MUA collections have been cataloged by LC or a NACO contributor. Thirty-four percent conflicted with names already in the LCNAF. Only 15 percent of the names matched the LCNAF exactly, meaning that no conflict could occur. In comparison, Michelson found that 36 percent of the name headings entered into RLIN AMC were "searched and found" in the LCNAF.¹⁷

Interestingly, 83 of the 116 exact personal name matches were for men; 24 were for families; and only 9 were for women. Almost every individual was white. This disproportionate figure raises numerous policy issues, which, unfortunately, are beyond the scope of this study. This may be an issue worth researching in greater detail.

The relative percentages of matches within the corporate and personal name categories also provide interesting statistics. Comparing the information found in Tables 2 and 3, we find that the number of exact matches for personal names is proportionally almost twice as high as for corporate names. Yet the number of partial matches for personal names is almost four times lower than corporate names. The number of no matches for personal names is two-thirds higher than for corporate names. The results suggest that corporate names may cause greater conflicts with the LCNAF than personal names.

The findings illustrated in Tables 2 and 3 call into question the informal principle that names that show up only once in the catalog do not need authority records.¹⁸ Lotka's law predicts that the number of authors who produce only one item is about 60 percent and library research has largely supported it. For example, Potter found 69.33 percent of the authors only once in the catalogs of the University of Wisconsin-Whitewater, and 63.5 percent at University of Illinois

TABLE 1
NUMBER OF MATCHES FOUND IN LCNAF

	Corporate Name Number of Hits	% Total Hits	Personal Name Number of Hits	% Total Hits
Exact Match	94	6	111	7
Partial Match	375	25	87	6
No Match	381	25	355	23
Exact Match (2 or more)	23	2	5	0
Partial Match (2 or more)	28	2	7	1
No Match (2 or more)	29	2	26	2
Total	930	62	591	39

Note: Percentages do not add up to 100 points because of rounding.

at Champaign-Urbana, and Fuller found 68.3 percent in her study.¹⁹ The MUA study found strikingly different figures: 92 percent of the names (1,403 of 1,521) were used only once in the collections. Most importantly, 44 percent of the names used only once either exactly or partially matched names already in the LCNAF. If other studies reveal the same general proportion of exact or partial matches, then it seems clear that there is great potential for confusion in the bibliographic databases.

The Need For Further Research

It should also be noted that, like many applied studies, other archivists who attempt to replicate the results of this one study might find very dissimilar results. Although the same methodology can be used, the fact remains that archival catalogs and holdings are very different. The holdings at the MUA may not reflect what is found in other repositories, and the results reflect the kinds of collections acquired by the institution. At least several more repositories need to

TABLE 2
NUMBER OF CORPORATE NAME MATCHES FOUND IN LCNAF

	Corporate Name Number of Hits	% Total of Corporate Names
Exact Match	94	10
Conflicting Match	375	40
No Match	381	41
Exact Match (2 or more)	23	3
Conflicting Match (2 or more)	28	3
No Match (2 or more)	29	3
Total	930	100

TABLE 3
NUMBER OF PERSONAL NAME MATCHES FOUND IN LCNAF

	Personal Name Number of Hits	% Total of Personal Names
Exact Match	111	19
Conflicting Match	87	15
No Match	355	60
Exact Match (2 or more)	5	1
Conflicting Match (2 or more)	7	1
No Match (2 or more)	26	4
Total	591	100

undertake the same kind of project before any trends could possibly be established. A manuscripts repository specializing in literary manuscripts might find results far different from those at the MUA. State archives, responsible for the administrative records of government agencies and departments, would probably have a greater preponderance of corporate names.

A great deal of fundamental research needs to be done in the online control of archival authority and bibliographic records. The MUA undertook this study to better understand how much effort should be applied towards authority work, but it is intended to be only a first step in learning how to make access to our collections more efficient. The results should not be considered the foundation for a broad principle that should be applied profession-wide. As additional research is conducted, perhaps the profession will find ways to create more effective archival information retrieval systems.

ABOUT THE AUTHOR: Mark A. Vargas has been academic archivist at the Golda Meir Library of the University of Wisconsin-Milwaukee since 1989. He began working as an archivist at the Massachusetts Institute of Technology in 1987.

NOTES

1. Throughout the article, the term "collection" is used to describe a fonds, record group, series, or manuscript collection.
2. Throughout the article, "automate" and "automation" refers to the creation of online bibliographic records.
3. Doris Hargrett Clack, *Authority Control* (Chicago: American Library Association, 1990), 207. See also Barbara B. Tillett, "Considerations for Authority Control in the Online Environment," *Cataloging & Classification Quarterly* 9:3 (1989): 3-4; Jackie M. Dooley, "An Introduction to Authority Control for Archivists," in *Archives and Authority Control: Proceedings of a Seminar Sponsored by the Smithsonian Institution October 27, 1987* (Pittsburgh: Archives and Museum Informatics, 1988), 5. See also the *ALA Glossary of Library and Information Science* (Chicago: American Library Association, 1983), 16.
4. The only major research article on the topic is Avra Michelson, "Description and Reference in the Age of Automation," *American Archivist* 50 (Spring 1987): 192-208.
5. The development of authority control is summarized in Robert H. Burger, *Authority Work* (Littleton, Colo.: Libraries Unlimited, 1985), 4-10; Ritvars Bregzis, "The Syndetic Structure of the Catalog," in *Authority Control: The Key to Tomorrow's Catalog* (Phoenix: Oryx Press, 1982), 19-27; Barbara B. Tillett, "Bibliographic Structures: The Evolution of Catalog Entries, References, and Tracings" in *The Conceptual Foundations of Descriptive Cataloging* (San Diego: Academic Press, 1989), 149-165. Several of the major research studies include Raya Fidel, "Searchers' Selection of Search Keys: II. Controlled Vocabulary or Free-Text Searching," *Journal of the American Society for Information Science* 42:7 (August 1991): 501-14; Elizabeth E. Fuller, "Variation in Personal Names in Works Represented in the Catalog," *Cataloging & Classification Quarterly* 9:3 (1989): 84-5, 90; Alexis J. Jamieson, Elizabeth Dolan, and Luc Declerck, "Keyword Searching vs. Authority Control in an Online Catalog," *Journal of Academic Librarianship* 12:5 (November 1986): 280; and Carol Tenopir, "Full Text Database Retrieval Performance," *Online Review* 9:2 (April 1985): 149-64. A review of the research can be found in C.P.R. Dubois, "Free Text vs Controlled Vocabulary," *Online Review* 11:4 (August 1987): 243-53; and in Elaine Svenonius, "Unanswered Questions in the Design of Controlled Vocabularies," *Journal of the American Society for Information Science* 35:5 (September 1986): 331-40.
6. Max Evans, "Authority Control: An Alternative to the Record Group Concept," *American Archivist* 49 (Summer 1986): 249-61; David Bearman, "Authority Control Issues and

- Prospects," *American Archivist* 52:3 (September 1989): 286-99; David Bearman and Richard Szary, "Beyond Authorized Headings: Authorities as Reference Files in a Multi-Disciplinary Setting," in *Authority Control Symposium* (Tucson: Art Libraries Society of North America, 1987), 69-78.
7. Elizabeth Black, *Authority Control: A Manual for Archivists* ([Ottawa]: Bureau of Canadian Archivists, 1991), 1-6. Hugo Stibbe also accepted the need for authority control without reservations in "Implementing the Concept of Fonds: Primary Access Point, Multilevel Description and Authority Control," *Archivaria* 34 (Summer 1992): 120-3.
 8. Lorene E. Ludy and Sally A. Rogers, "Authority Control in the Online Environment," *Information Technology and Libraries* 3:3 (September 1984): 265.
 9. This idea is based on a statement in Henriette D. Avram, "Authority Control in Its Place," *Journal of Academic Librarianship* 9:6 (January 1984): 333.
 10. Marcia Bates, "Subject Access in Online Catalogs: A Design Model," *Journal of the American Society for Information Science* 37:6 (November 1986): 357-76; Christine L. Borgman, "Why Are Online Catalogs Hard to Use? Lessons Learned from Information-Retrieval Studies," *Journal of the American Society for Information Science* 37:6 (November 1986): 387-400; Charles R. Hildreth, "Beyond Boolean: Designing the Next Generation of Online Catalogs," *Library Trends* 35 (Spring 1987): 647-67; Martha M. Yee, "System Design and Cataloging Meet the User: User Interfaces to Online Public Access Catalogs," *Journal of the American Society for Information Science* 42:2 (March 1991): 78-98.
 11. Ake I. Koel, "Bibliographic Control at the Crossroads: Do We Get Our Money's Worth?" *Journal of Academic Librarianship* 7:4 (September 1981): 220-2.
 12. Randal K. Barry, *Report of the 1986 NACO Cost Survey* (Library of Congress National Coordinated Cataloging Operations, 1987), 9, cited in Jean Dickson and Patricia Zadner, "Authority Control and the Authority File: A Functional Evaluation of LCNAF on RLIN," *Cataloging & Classification Quarterly* 9:3 (1989): 58.
 13. Patricia D. Cloud, "The Cost of Converting to MARC AMC: Some Early Observations," *Library Trends* 36 (Winter 1988): 577. The project, funded by the National Endowment for the Humanities and the Pew Memorial Trust, involved twelve major research libraries and the conversion of about 21,000 bibliographic records to the MARC AMC format in RLIN. The worst problem was the poor quality of finding aids. The Philadelphia Area Consortium of Special Collections Libraries Name Authorities Project does not use any set of standards for establishing the names. Letter of Amy E. McLoughlin to the author, 29 October 1991.
 14. Evans, "Authority Control," 249-61; Bearman, "Authority Control Issues and Prospects," 286-99; Bearman and Szary, "Beyond Authorized Headings," 69-78.
 15. An analysis of the cataloging project can be found in Mark A. Vargas and Janet Padway, "Catalog Them Again for the First Time," *Archival Issues* 17 (Spring 1992): 49-63.
 16. Of the 591 personal names, only 46 required a detailed analysis of the provenance records to help determine the fullest form of the name. The finding aids proved sufficient in the other cases, although the staff noted that biographical data was often scattered throughout the finding aids.
 17. Michelson, "Description and Reference in the Age of Automation," 196.
 18. See Thomas Garnett's comments during the general discussion period of the final session of the Archives and Authority Control seminar, as reported in *Archives and Authority Control: Proceedings of a Seminar Sponsored by the Smithsonian Institution, October 27, 1987*, 48; Marion Matters, "Authority Work for Transitional Catalogs," in *Describing Archival Materials: The Use of the MARC AMC Format*, Richard P. Smiraglia, ed. (New York: The Haworth Press, 1990), 95-6, 109.
 19. William Gray Potter, "When Names Collide: Conflict in the Catalog and AACR2," *Library Resources and Technical Services* 24 (Winter 1986): 3-16; and Fuller, "Variation in Personal Names in Works Represented in the Catalog," 82.

