



Article

Name Authority Challenges for Indexing and Abstracting Databases

Denise Beaubien Bennett

Engineering Librarian and Online Coordinator, Marston Science Library

University of Florida, George A. Smathers Libraries

Gainesville, Florida, United States of America

E-mail: dbennett@ufl.edu

Priscilla Williams

Head, Authorities and Metadata Quality Unit Cataloging and Metadata Department

University of Florida, George A. Smathers Libraries

Gainesville, Florida, United States of America

E-mail: priwill@uflib.ufl.edu

Received: 01 December 2005

Accepted: 22 February 2006

© 2006 Bennett and Williams. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Objective - This analysis explores alternative methods for managing author name changes in Indexing and Abstracting (I&A) databases. A searcher may retrieve incomplete or inaccurate results when the database provides no or faulty assistance in linking author name variations.

Methods -The article includes an analysis of current name authority practices in I&A databases and of selected research into name disambiguation models applied to authorship of articles.

Results - Several potential solutions are in production or in development. *MathSciNet* has developed an authority file. The method is largely machine-based but it involves time-consuming manual intervention that might not scale up to larger or multidisciplinary databases. The use of standard numbers for authors has been proposed. Solutions in practice include author-managed registration records and linking among several authority files. Information science and computer science researchers are developing models to automate processes for name disambiguation, shifting the focus from authority control to access control. Successful models use metadata beyond the author name alone, such as co-

authors, author affiliation, journal name, or keywords. Social networks may provide additional data to support disambiguation models.

Conclusion - The traditional objective of name authority files is to determine precisely when name variations belong to the same individual. Manually-maintained authority files have served library catalogues reasonably well, but the burden of upkeep has made them ill-suited to managing the volume of items and authors in all but the smallest I&A databases. To meet the access needs of the 21st Century, both catalogues and I&A databases may need to implement options that present a high degree of probability that items have been authored by the same individual, rather than options that provide high precision with the expense of manual maintenance. Striving for name disambiguation rather than name authority control may become an attractive option for catalogues, I&A databases, and digital library collections.

Introduction

Indexing and Abstracting (I&A) databases generally have not implemented name authority control as is used in many library catalogues. Most I&A databases burden the searcher with identifying and selecting name variations. The use of widely variant forms of authors' names without reference or linkage to alternatives causes hardship for searchers. End-users' search results may be inaccurate or incomplete, resulting in a decrease in the scientific integrity of the research. This article will explore various approaches to solving these challenging name variation issues.

For many years, across research communities, librarians and researchers have had to deal with the problem of increasing numbers of variant forms of an author's name. Some variants are created and occur over the life of a publishing career; some may be attributed to author preferences while others are created to conform to requirements of publishing guidelines. Variant forms due to misspellings, spacing, cultural norms, and use of initials supply one set of concerns. Name changes, outgrowths of an author's life over time as a result of outside influences involving such personal matters as marriage and legal name changes,

provide a special challenge for database maintainers as well as searchers.

Individual library online catalogues have been capable of applying authority control methods since the implementation of AACR2 (Taylor 224). Personal name authorities bring together works by an author, regardless of the variations in name as identified in the work itself (Tillett "Authority control" 24). Name authorities and related issues tend not to be discussed in the database indexing world to the extent they are discussed in cataloguing and back-of-the-book indexing (Taylor 225; Spink and Leatherbury 143-44).

Name authorities present many challenges for I&A databases beyond those facing maintainers of library catalogues. In addition to variations in language translations and cultural naming customs, publication editors frequently dictate whether authors may use their full names or are restricted to their initials (see Appendix A). Thus, I&A databases receive items that may already contain name variations. I&A databases may choose to exert some sort of name authority control over the variations to ensure that a search on one form of the author's name will retrieve all works by that author. I&A databases tend to develop their own procedures for handling name

authority issues, such as stripping all author names down to initials. Most I&A databases cluster works by the form of author name, but don't provide redirects to other forms of the authors' names. For example, the searcher must note and select all relevant entries such as "Last, F.," "Last, First," "Last, F. M.," "Last, First M.," "Last, First Middle," "Last, F. Middle," "Middle Last, F." where all of these variations are included in the author index. Some I&A databases, such as the *Web of Science*

<<http://scientific.thomson.com/products/wos/>>, both strip author names down to initials and deliberately choose not to exert any authority control, cautioning searchers to try all likely name variations ("Author Names"; *Web of Science 7.0 Workshop* 41).

One particular challenge lies in managing author name changes. Indexing practices recommend appropriate treatment, such as: "But if a person was well-known also under a previous name, cross-references from and to the changed name should be made.... The same treatment applies to married women who become well-known under their maiden names and continued to create literary or artistic works or became otherwise known also under their married names." (Wellisch 360-61). Few databases have chosen to link the variations or name changes to facilitate searching and retrieval of an author's works (see Appendix B). I&A databases may also move all of an author's works from the former name to the current name (see Appendix C), thus altering some records so the author name no longer matches that displayed on the original article.

Regardless of whether I&A databases choose to link author variations, searchers expect the form of name on the retrieved bibliographic records to match the form of the name on the published article. When the names are significantly mismatched between the I&A database and the article

itself, the searcher is likely to be confused. Future researchers may cite an article by copying the form used in the I&A database, thus carrying over the disconnect from the name used on the article. Further chaos ensues when citations are gathered by citation indexes and linking databases, such as the *Web of Science*. Any citation that uses a form of the author's name other than that on the article will not match the correctly identified items already in the *Web of Science* database.

The challenges of coping with name variations multiply when end-users search across multiple databases while formulating their literature searches. Automated or manual de-duplication of identical items becomes more problematic, whether end-users create their own bibliographies or employ bibliography management software to manage their citations, with name changes than with simpler name variations.

Linking services such as CrossRef <<http://www.crossref.org/>> rely on Digital Object Identifiers and other numerically hashed methods of identifying identical citations to link through OpenURL both to full text options and to shared citations. Where the matching and passing algorithms rely only on numbers (such as ISSN, year, volume, issue, starting page), problems with name variations and changes may be reduced from chaotic to merely puzzling. Where the algorithms include author names, variations may reduce the probability of matches and linkages. As long as researchers rely on author names to identify works, I&A databases can assist by clearly identifying the name on the article as well as its variations.

Examples of Problems with Name Changes

One author has published works under two forms of her name: Denise M. Beaubien until mid-1992, and Denise Beaubien Bennett

after mid-1992. A search for her works in WilsonWeb's *Library Literature & Information Science Full Text* database

<<http://www.hwwilson.com/Databases/liblit.htm>> for <denise beaubien> in [All - Smart Search] yields disturbing results (Figure 1).

The author name on the articles of the five oldest items, published 1988-1992, is Denise M. Beaubien. However, only one of the citations [Beaubien, D.M. "The changing roles of online coordinators." *Online* (Weston, Conn.) v. 15 (September 1991) p. 48-50+] displays this form of the author's name. The other four older citations display a form of name that (1) does not appear on

the articles and (2) has never been used by the author but which appears to be an amalgamation of the two forms of her name created by the database indexers: Denise M. Beaubien Bennett. All but one of the citations from 1993 to the present also display the amalgamated form of name, but the initial "M" does not appear on the articles and has not been used by the author in any context, legally or professionally, since mid-1992. The most recent citation [Bennett, D.B., et. al., "A Class Assignment Requiring Chat-Based Reference." *Reference & User Services Quarterly* v. 44 no. 2 (Winter 2004) p. 149-63] uses the form of the name on the article, without the "M."

Current Search: <near>(denise,beaubien) <in> ALL
Records: 11
In: Library Lit & Inf Full Text
Limiting Icons: Full Text PDF Peer Reviewed **Linking Icons:** SFX SFX

60% <input type="checkbox"/>	1	Bennett, D.B., et. al., A Class Assignment Requiring Chat-Based Reference . <i>Reference & User Services Quarterly</i> v. 44 no. 2 (Winter 2004) p. 149-63
60% <input type="checkbox"/>	2	Bennett, D.M.B. Archaeology and evolution: what happens when you clean your office [unearthing old computer equipment]. <i>College & Research Libraries News</i> v. 63 no. 1 (January 2002) p. 30
60% <input type="checkbox"/>	3	Bennett, D.M.B. RUSA review underway: your survey responses . <i>Reference & User Services Quarterly</i> v. 40 no. 4 (Summer 2001) p. 352-60
60% <input type="checkbox"/>	4	Matylonek, J.C., et. al., Managing access to a publisher package: IEEE, IEL, and Xplore (computer file) . <i>Issues in Science & Technology Librarianship</i> no. 30 (Spring 2001)
60% <input type="checkbox"/>	5	Lundgren, J.H., et. al., The art of electronic service [at Marston Science Library, University of Florida]. <i>Florida Libraries</i> v. 41 no. 6 (September/October 1998) p. 120-2
60% <input type="checkbox"/>	6	Bennett, D.M.B., et. al., Campus-wide training in LEXIS/NEXIS, with applications in the sciences [at the University of Florida]. <i>Science & Technology Libraries</i> v. 14 (Winter 1993) p. 89-104
60% <input type="checkbox"/>	7	Bennett, D.M.B. Wilson vs IAC on tape: a comparison [selection factors in choosing a locally mounted database for the university system of Florida]. <i>Database (Weston, Conn.)</i> v. 15 (February 1992) p. 52-6
60% <input type="checkbox"/>	8	Beaubien , D.M. The changing roles of online coordinators [evolving roles of reference librarians]. <i>Online (Weston, Conn.)</i> v. 15 (September 1991) p. 48-50+
60% <input type="checkbox"/>	9	Bennett, D.M.B., et. al., Software for patron use in libraries [special issue]. <i>Library Trends</i> v. 40 (Summer 1991) p. 1-197
60% <input type="checkbox"/>	10	Battiste, A.L., et. al., The University of Florida's moving experience [consolidating branches into a central science library]. <i>College & Research Libraries News</i> no. 6 (June 1989) p. 467-71
60% <input type="checkbox"/>	11	Bennett, D.M.B., et. al., Patron-use software in academic library collections [University of Florida guidelines for purchasing, cataloging, circulating, and preserving]. <i>College & Research Libraries News</i> no. 10 (November 1988) p. 661-2+

Figure 1: Search for author's older form of name in WilsonWeb's All - Smart Search. Copyright © 2006 by the H. W. Wilson Company. Material reproduced with permission of the publisher. Permission granted 2/13/2006.

Title: If you want to evaluate your library— (Book Review)

Personal Author: **McCook, Kathleen de la Pena**, reviewer

Source: [Journal of the American Society for Information Science v. 41 \(January 1990\) p. 72](#)

Figure 2: Citation in WilsonWeb for older item displays author's newer name. Copyright © 2006 by the H. W. Wilson Company. Material reproduced with permission of the publisher. Permission granted 2/13/2006.

Other authors who have changed their names suffer a similar fate. A search for <Kathleen Heim> in <Author> yields 88 items, published from 1984-2004. However, a record in *Library Literature & Information Science* for a publication from 1990 displays as is shown in Figure 2., while the author's name on the article is:

Kathleen M. Heim
Louisiana State University
School of Library & Information
Science

The amalgamated name does serve to draw the author's works together. However, searchers may be lulled into assuming the amalgamated name is used throughout the database. But redirecting a search on the amalgamated name does not retrieve the citations that use the form of name as listed on all of the articles. In the case above, two citations are not retrieved from the redirected search. Examples from both of these authors indicate that authority control is applied incompletely in the database, negating many of its benefits. Unsuspecting searchers will not know they may have missed some hits.

The concept of the amalgamated name may aid in retrieving database records, but the practice can lead to errors or variations in citation functions. Many researchers create bibliographies by copying and pasting citations from databases. This practice is encouraged by database producers, who develop excellent tools for marking, sending, and saving records; and by librarians, who

encourage patrons to use these database features as well as bibliography management software to reduce citation errors. When database citations do not indicate the form of the name as used on the article, errors in proper citing may follow.

The *Web of Science*, the original citation tool, uses the form of author name (and the rest of the citation) exactly as it appears in the citing article, stripping all but the surname down to initials. ISI's long-standing policy is not to over-correct "variations" because its indexers cannot check them all (*Cited Reference Searching* 3) and will not second-guess an author's intentions. To search for cited references in the *Web of Science* to all of the first author's works above, a searcher should only have to enter two strings: <beaubien dm or bennett db>. However, if authors copy the citation from a *WilsonWeb* database, a searcher must add <or bennett dmb> to the *Web of Science* search string to retrieve all the matches. The problem is magnified when searching for citations to a particular work. When the searcher limits to only the name on the article but an I&A database has reformatted an author's name and a citer chooses the name from the I&A database rather than the name on the article, some citations will not be retrieved. Searching the *Web of Science* is challenging enough when accommodating for authors' typos. Accommodating deliberate variations and name changes introduced by I&A databases adds to the complexity and reduces the recall of items retrieved. In addition to retrieval challenges, incorrect

Times Cited**	Cited Author	Cited Work	Year	Volume	Page	Article ID	View Record
3	BEAUBIEN DM	ONLINE	1991	15	48		View record
Times Cited**	Cited Author	Cited Work	Year	Volume	Page	Article ID	View Record
1	BENNETT DMB	ONLINE	1991	15	48		

Figure 3: Cited References in Web of Science displaying how the error in the index results in an error in the Cited Reference matching and counting. Thomson Scientific, Inc. is the publisher and copyright owner of Web of Science®. The screen shots are used with the expressed permission of Thomson Scientific. Permission granted 2/9/2006.

use of an author's name by an I&A database results in the creation of an additional unlinked record in the *Web of Science* plus a failure to increment the "times cited" counter on the valid record for the item. Having one's cited references grouped for easy and complete counting is increasingly important among authors (Monastersky, sec. 2). The first entry in Figure 3 is the valid record. The second entry was created because a citer (Cardina and Wicks 142) copied the author's name from a *WilsonWeb* database rather than from the article. Not only is the author's work not officially or correctly counted in the *Web of Science*, but subsequent searchers cannot view the full record of the original article within the *Web of Science* because the *view record* link fails to form. Most I&A databases force the searcher to generate all variations on the author name to assure high recall of results. *Library Literature & Information Science* and all the *WilsonWeb* databases are rare among I&A databases in deliberately changing an author's name on a citation to correspond to its latest known form (see Appendix B).

Potential Solutions: Overview

Solutions to the problem of identifying and linking author name changes within I&A databases can take many approaches. Solutions both in production and in the research modeling stage are clustered into categories and described below:

- 1) Authority control through the use or linking of Name Authority files
 - a) Uses a file: *MathSciNet*, *WilsonWeb*
 - b) Proposed file: International Standard Authority Name/Data Number
 - c) Linking across files: HoPEC, ANAC Levy Project; LEAF
- 2) Name disambiguation through automated methods
 - a) In practice: *Author-ity*
 - b) Models in development by research teams, including use of social networks

Maintaining name authority files requires a high amount of labor but benefits the end-user with results of both high recall and high precision (Lancaster 131-4) in identifying documents by or about the same individual. Automated methods of name disambiguation may require less manual labor but likely cannot achieve the level of high recall and high precision of well-maintained authority files unless they also employ substantial manual checking.

Potential Solutions: Authority File in the *MathSciNet* Database

The *MathSciNet*

<<http://www.ams.org/mathscinet/>> database creates and maintains a name authority file to control variations. Much of the identification process is automated; however, approximately twenty percent of

the items require manual checking. (TePaske-King and Richert par. 10; [Uniquely Identifying Mathematical Authors](#)). “Authors are *distinct entities* in the MR Database, independent of name variations used in particular publications.” ([MathSciNet Author Database Help](#)). In the Author Database, search results are displayed as a [headline list](#) of authors. The primary listing is the preferred or fuller form of the name. Listed below the headline are the variations on the name as used on articles cited in the database. The searcher immediately sees the name variations and accepts that the variations point to the same author. The headline name serves to group the variations, but the form of name displayed in each citation matches the name on the article. Searchers who mark and save records to import into their bibliographies

will pass along the name variant as used on the article, enabling future researchers to match the citation and article without confusion over the authorship.

A Quick Search in *MathSciNet* for a truncated author search retrieves records that match only that form of name, and many include more than one author, as is typical in most I&A databases. A search in the Author Database rather than Quick Search for <kister, j*> displays two matches on the truncated form (Figure 4). Each match displays the entry from the authority file and all name variations. Author <Kister, Jane> has published using three variant names. The display indicates that all three variations belong to one author, and confirms the preferred form.

AMERICAN MATHEMATICAL SOCIETY
MathSciNet *Mathematical Reviews on the Web*

Matches for: kister, j*
 Number of Matches: 2 Page 1 of 1

[View All Items](#) [Combine With Full Search](#) [View Co-Authors](#)

1. Kister, James M. MR CD
 - ◆ Kister, J. M.
 - ◆ Kister, James
2. Kister, Jane E. MR CD
 - ◆ Kister, Jane Bridge
 - ◆ Bridge, Jane

Figure 4: Author Database entry in MathSciNet. Reprinted with permission by the American Mathematical Society. Permission granted 2/10/2006.

AMERICAN MATHEMATICAL SOCIETY
MathSciNet Mathematical Reviews on the Web
 Navigate MathSciNet Jump to Search or Browse Screens
 View clipboard: 1 item Support Mail | Help | FAQ
 Items Authored by Kister, Jane E. and
 Number of Matches: 8 Page 1 of 1 Select page: 1
 Citations (AMSRefs) Retrieve Marked Retrieve All Unmark All

- [1] [MR2026942](#) [Kister, Jane E.](#); [Sherbert, Donald R.](#) Robert G. Bartle (1927--2003). *Notices Amer. Math. Soc.* 51 (2004), no. 2, 239--240. [01A70](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX
- [2] [MR1788190 \(2001g:00030\)](#) Featured reviews in Mathematical Reviews 1997--1999. With selected reviews of classic books and papers from 1940--1969. Edited by Donald G. Babbitt and Jane E. Kister. [Featured Reviews in Mathematical Reviews. American Mathematical Society, Providence, RI, 2000.](#) viii+740 pp. ISBN: 0-8218-2166-0 [00B60](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX
- [3] [MR1605708 \(98m:00025\)](#) Featured reviews in Mathematical Reviews 1995--1996. Reviews of outstanding recent books and papers. Edited by Donald G. Babbitt and Jane E. Kister. [Featured Reviews in Mathematical Reviews. American Mathematical Society, Providence, RI, 1998.](#) viii+382 pp. ISBN: 0-8218-0771-4 [00B60 \(00A99\)](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX
- [4] [MR1026392](#) [Kister, Jane E.](#) From published paper to MR review: how does it happen? *Notices Amer. Math. Soc.* 36 (1989), no. 10, 1362--1365. [00A99 \(01A80\)](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX
- [5] [MR0915636 \(89e:03001f\)](#) Ω -bibliography of mathematical logic. Vol. VI. Proof theory; constructive mathematics. Edited by Gert H. Müller, Wolfgang Lenski, Jane E. Kister, Dirk van Dalen and Anne S. Troelstra. [Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1987.](#) xlii+405 pp. ISBN: 3-540-15524-4 (Reviewer: Perry Smith) [03-00 \(00A15 01-00 03B40 03Fxx\)](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX
- [6] [MR0881280 \(88g:03079\)](#) [Crossley, John N.](#); [Kister, Jane Bridge](#) Natural well-orderings. *Arch. Math. Logik Grundlag.* 26 (1986/87), no. 1-2, 57--76. (Reviewer: Peter Pöppinghaus) [03F15](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX
- [7] [MR0532719 \(58 #27171\)](#) [Bridge, Jane](#) Beginning model theory. The completeness theorem and some consequences. Oxford Logic Guides. Clarendon Press, Oxford, 1977. viii+143 pp. (Reviewer: James H. Schmerl) [02-01 \(02H05\)](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX
- [8] [MR0398793 \(53 #2644\)](#) [Bridge, Jane](#) A simplification of the Bachmann method for generating large countable ordinals. *J. Symbolic Logic* 40 (1975), 171--185. (Reviewer: S. Feferman) [02D05](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX

Number of Matches: 8 Page 1 of 1 Select page: 1

Figure 5: Records retrieved from selecting radio button for authority name in MathSciNet Author Database. Reprinted with permission by the American Mathematical Society. Permission granted 2/10/2006.

Selecting the radio button next to the chosen author from the Author database and clicking the View All Items button displays a list of all items written by the author, regardless of name variations on the articles

(Figure 5). In contrast, searching the full database by author's earlier name <bridge, jane*> retrieves only those records with that form of name on the article (Figure 6).

AMERICAN MATHEMATICAL SOCIETY
MathSciNet Mathematical Reviews on the Web
 Navigate MathSciNet Jump to Search or Browse Screens
 View clipboard: 1 item Support Mail | Help | FAQ
 Matches for: Author/Related=bridge, jane*
 Number of Matches: 2 Page 1 of 1 Select page: 1
 Citations (AMSRefs) Retrieve Marked Retrieve All Unmark All

- [1] [MR0532719 \(58 #27171\)](#) [Bridge, Jane](#) Beginning model theory. The completeness theorem and some consequences. Oxford Logic Guides. Clarendon Press, Oxford, 1977. viii+143 pp. (Reviewer: James H. Schmerl) [02-01 \(02H05\)](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX
- [2] [MR0398793 \(53 #2644\)](#) [Bridge, Jane](#) A simplification of the Bachmann method for generating large countable ordinals. *J. Symbolic Logic* 40 (1975), 171--185. (Reviewer: S. Feferman) [02D05](#)
 Linked PDF Add to Clipboard Doc Delivery Journal Article SFX

Number of Matches: 2 Page 1 of 1 Select page: 1

Figure 6: Direct search of MathSciNet by non-authority variation retrieves only those matches. Reprinted with permission by the American Mathematical Society. Permission granted 2/10/2006.

MathSciNet's solution is elegant and workable in the relatively small database where its authors come from a size-limited community and where it is possible for human indexers to check and correct problematic entries manually. Although this solution might not scale up to large databases such as *PsycINFO*, *BIOSIS*, *Chemical Abstracts*, or the *Web of Science*, it should be possible to implement in databases covering narrow disciplines such as *Library Literature & Information Science*.

Potential Solutions: More Examples Creating, Using, or Linking Authority Files

Indexing and abstracting databases may follow Library of Congress (LC) practice but might find an additional benefit in perusing the Library of Congress Name Authority File (LC/NAF) to assist in collocating the variant names in author databases.

The Library of Congress Name Authority File contains over 5 million name authority records. Over 2 million of these records are contributed by NACO, the Name Authority Cooperative of the Program for Cooperative Cataloging run by the Library of Congress (<<http://www.loc.gov/catdir/pcc/naco/nacopara.html>> or <<http://authorities.loc.gov/help/contents.htm>>). Institutions become members of the NACO community and participate in the shared environment of authority control by contributing records to the LC/NAF following LC practice. New and changed name authority records are contributed to the file. As the number of contributions increase, the number of available names that can be used increases. Aside from providing controlled author name access, the records in the LC/NAF are rich in a cross-reference structure that links name changes and provides additional information that can be used effectively in compiling author databases. The LC/NAF, designated the "national" resource authority file, is not

strictly national and hasn't been since 1975. An agreement with the National Library of Canada (NLC) to use NLC headings when creating new name authority records for Canadian personal name authors afforded LC the opportunity to pursue its goal of an international authority file. Also, LC is very likely to use personal name author headings already established by the NLC. In addition to NLC headings, the LC/NAF contain British and Australian personal name authors (Kuhagen 132-133).

Although the LC/NAF is created with data from published books rather than from published articles, I&A databases may benefit from the effort that goes into compiling the LC/NAF. *WilsonWeb* databases check the LC/NAF (see Appendix B), but err in changing authors' names rather than pointing to the variations as given on the article. The LC/NAF supports high precision in linking name variations to an individual, but the identification and linking work is largely done by slow and manual, albeit distributed, methods. Several projects build on LC/NAF and other authority files; selected descriptions follow.

The IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) is working to develop a "conceptual model to assist in an assessment of the potential for international sharing and use of authority data both within the library sector and beyond" (G. Patton 41). One charge to FRANAR is "to study the feasibility of an International Standard Authority Data Number (ISADN)" (G. Patton 40) which, if created, might serve as a model for I&A databases as well as for library catalogues, digital libraries, archives, museums, and rights management organizations. At present, the FRANAR draft report titled *Functional Requirements for Authority Records: A Conceptual Model* (IFLA UCBIM) does not yet address the ISADN issue. In a related effort, Snyman and Jansen

Van Rensburg argue for the use of an International Standard Author Number (ISAN) to reduce dependence on identifying author name variations (“NACO vs. ISAN”; “Reengineering Name Authority Control”). Opponents of Standard Number approaches express concerns regarding organizational maintenance costs (Tillett “Authority Control” 30; Delsey 74).

The HoPEc system (Cruz et al. 1-8) controls author records within the RePEc economics library <<http://authors.repec.org/>>. HoPEc implements an author registration component that enables authors to create and maintain their own authority records. HoPEc thus shifts the maintenance burden away from a centralized group. Authors wishing for their papers to be clustered must identify and manage their own name variations. Reliance on authors leads to uneven participation and data quality, but the model offers a distributed solution with low organizational maintenance costs.

Librarians recognized long ago that linking methods could substitute for authorized forms of names (Tillett “Authority Control” 25). In the automated environment, a system does not have to select one “correct” form as long as all the variations link to each other. The Getty *Union List of Artist Names Online* <http://www.getty.edu/research/conducting_research/vocabularies/ulan/> links records that have been created within several separate authority files.

Members of the large-scale Levy Project to digitize a sheet music collection have created an Automated Name Authority Control system (ANAC) based on the LC name authority file to facilitate interoperability (DiLauro et al. sec. 3; Warner and Brown 21-2). The metadata include the statement of responsibility, such as “composer” or “words by.” Probability is based on a model that permits updating

after new data are added. ANAC was successful in establishing matches 58% of the time: 77% when a name existed in LC/NAF and 12% when a record did not exist in LC/NAF. ANAC took about eight seconds per name to perform the classification and is viewed as a complement to human effort (M. Patton et al. sec. 6).

The LEAF project for Linking and Exploring Authority Files creates a “Shared Name Authority File” (Weber 233) that can be used by all participating database producers. LEAF automatically links all authority records that pertain to the same person, based on the automatic linking rules of the project and including birth/death dates (232). LEAF utilizes the Z39.50 protocol for searching across authority files. LEAF does not merge the records into a new entity; it preserves each local file’s practices. Multidisciplinary databases might follow the LEAF lead in linking authority files that may exist within smaller or narrower disciplines.

Barbara Tillett outlines the progress toward building a virtual international authority file in a series of papers (“Virtual International Authority File”; “AACR2 and Metadata”; “Authority Control”). These cooperative efforts are based on linking parallel authority records that will continue to be maintained locally rather than attempting to merge metadata into super records. Tillett favors testing of unique, persistent record control numbers within existing services (“Authority Control” 30) or any method that does not require establishing an international organization to maintain standard numbers. Ki-Tat Lam proposes converting authority files to an XML format and enabling the files as SOAP nodes (93-95) to achieve global name access. Linked authority records may assist efforts at identifying more name variations that point to a single individual. However, name variations occur more frequently in the

journal literature than in library catalogues due to editing and indexing practices. Linked authority records are still limited to the metadata variations included in those records.

Potential Solutions: Alternative Approaches Using Name Disambiguation

Digital Libraries are examining the issues involved in name authority control as well as topical authority control. "Such name ambiguity affects the performance of document retrieval, web search, [and] database integration, and may cause improper attribution to authors." (Han et al. "A Hierarchical Naïve Bayes Mixture Model" 1065). Rather than devising name authority files, researchers are aiming for an outcome of *name disambiguation*, or an automated method of examining more than the author name to determine the likelihood that any two papers with similar author names i.e., last name and first initial, have been written by the same person. The challenges are summed up by Malin, Airoidi and Carley who state, "In the real world, it is not clear if any observed name ever has complete certainty. This suggests probabilistic models of certainty may be useful for disambiguating names when many names are potentially ambiguous." (136).

Eugene Garfield, founder of the *Science Citation Index* (now in database form as the *Web of Science*) long ago acknowledged the need to examine more data than name and initials alone to disambiguate authors. "On the other hand, when using the *Source Index* of the *SCI* to locate articles written by a particular author it is not possible to differentiate between two different men with the same name and initials, unless one knows something about their fields of work." (2)

The term *authority control* is generally restricted to the library world, and is increasingly limited to catalogues. Other disciplines solve similar records-management problems. Digital libraries strive instead to create *access control*, where variations are linked without establishing an official or preferred version (Cruz et al.). Statisticians discuss *record linkage* to match, for example, family members in health care or census files (Bhattacharya and Getoor 12; Fellegi and Sunter 1183-4). Database maintainers use *deduplication* or *citation matching* or *identity uncertainty* (Pasula et al. sec. 1; On et al. 346), which librarians generally consider as a method for identifying entire records that match rather than matching just the author fields in records. All of these fields offer models that assist with fuzzy matching, but many are not geared specifically toward accommodating name changes that incorporate different words.

Authority name issues can be grouped into three categories: (1) multiple name variations that signify the same author; (2) similar or homonymic names that belong to more than one author, and (3) *linear changes* when an author alters his/her name, generally due to changes in marital status or other religious or legal reasons. Terminology is not standardized, even within research teams, and varies whether researchers are discussing the state of pre-processed records or the process applied to organize the records. The following terms may be used outside of the library science discipline to indicate research into authority control issues. Lee et al. (69) define *mixed citations* as authors with similar/homophonic names grouped or mixed together and *split citations* when one author generates name variations; while Hong, On, and Lee (137) define *split* as the process of separating multiple authors with similar names and *merge* as the process of merging one author's name variations into one cluster. Malin,

Airoidi, and Carley (120) use *variation* to indicate one author with many names and *ambiguity* to indicate similar names/many authors. Niu, Li, and Srihari (sec. 1) define *alias association* as the process of managing one author with many names and *disambiguation* as the process of tackling similar names that indicate many authors. Linear name changes generate less attention, probably because the other categories seem more readily solvable without human effort.

The Torvik team is developing “several planned steps toward our long-term goal of completely partitioning MEDLINE into unique authors.” (157). Their model examines MeSH headings, title words, journal names, and coauthors to estimate the probability that a pair of author names refers to the same individual. From this model, the team developed a name disambiguation tool for the Medline database. *Author-ity* <<http://arrowsmith.psych.uic.edu>> provides “a list of articles ranked by decreasing probability that the author name [searched] given on the article [selected] refers to the same individual.”

The teams led by Han are testing various models of machine learning against the DBLP Computer Science Bibliography data (Han et al. “Mining and Disambiguating”; Han et al. “Information Access”; Han et al. “Name Disambiguating”). The models use data from co-author names, keywords in paper titles, and source titles in addition to the solo author name. The various models all point to similar ways to add data to enhance disambiguation. The number of features included and the weight assigned to these features can improve disambiguation performance (Han et al. “Name Disambiguating” 338). Authors with both similar names and similar research interests pose greater challenges for successful disambiguation. Since the keywords present in article and source titles

may be sparse, using word clustering techniques to group research areas (such as reference or cataloguing) may enhance disambiguation. The team could also consider including the author-supplied keywords where present.

Malin, Airoidi, and Carley (136) and Mann and Yarowsky (2) argue for the use of social networks to assist in disambiguation. Social networks provide context surrounding a name, similar to the manner in which coauthors and keywords provide a context for distinguishing among authors. Researchers acknowledge the depth of the problem when a manual examination of the data is insufficient for determining whether a name belongs to one or two individuals (Bekkerman and McCallum 469; Fleischman and Hovy conclusion). These projects do not focus on the narrower problem of disambiguating names when all are known to be authors and where the metadata reside in tagged author fields, but techniques resulting from these projects may apply to structured bibliographic databases.

The Pasula team admit “[W]e do not currently model the fact that erroneous citations are often copied from reference list to reference list ...” (8) indicating a rare acknowledgement of the copying problem and perhaps a promise to include the chaining of error-filled citations in future models.

The selected disambiguation projects described above, share similar attributes. All use metadata beyond the author name alone. Most have proven that adding more data elements to their models can serve to disambiguate names in a faster manner and with a higher probability of success than in relying on single author names alone. All models are tested on databases of limited subject scopes (music, medicine, computer science, economics) and thus each group of researchers is uncovering similar successes

and challenges. None have yet tested their models on data from multidisciplinary or extremely large databases. Merging the techniques of adding data elements and relying on disciplines to maintain their own linked name files may result in long term success for large, multidisciplinary databases such as I&A databases.

Conclusion

Most I&A databases place the burden on the searcher to identify and select author name variations. The *WilsonWeb* databases impose authority control by altering author names, but this practice causes the index entries to fail to match the name on the article. Maintaining an authority file to manage name variations, such as the *MathSciNet* approach, is an effective service for the searcher but is not likely to scale well for larger databases. Alternative solutions must be implemented to assure access, retrieval, and proper crediting of authors' works. Without control or linkage to name variations, searchers may retrieve incomplete or inaccurate results.

The traditional objective of name authority files is to determine precisely when name variations belong to the same individual. Manually-maintained authority files have served library catalogues reasonably well, but the burden of upkeep has made them ill-suited to managing the volume of items and authors in all but the smallest I&A databases. To meet the access needs of the 21st Century, both catalogues and I&A databases may need to implement options that present a high degree of probability that items have been authored by the same individual, rather than options that provide high precision with the expense of manual maintenance. Striving for name disambiguation rather than name authority control may become an attractive option for catalogues, I&A databases, and digital library collections.

I&A databases may soon have many automated options for facilitating name disambiguation. We encourage I&A database producers to examine and implement options researched by the Digital Library community. Developing automated methods can reduce the searcher's burden of determining author name variations while ensuring that the author index entries match the names on the article and that the end-user can successfully retrieve all of an author's works from that database.

Works Cited

"About Library of Congress Authorities." Library of Congress Authorities Help Pages. Washington, DC: Library of Congress. 1 December 2005. <<http://authorities.loc.gov/help/contents.htm>>.

About Name Authority Control in H.W. Wilson's Indexing Services. New York: H.W. Wilson, 2005. 1 December 2005. <http://www.hwwilson.com/Databases/names_authority_control.htm>.

ASIST Digital Library. New York: Wiley Interscience. 1 December 2005. <<http://www3.interscience.wiley.com/cgi-bin/jtoc/27981/>>.

"Author Names." Web of Science Help. Philadelphia: Thomson Corporation, 2005. 1 December 2005. <http://wos17.isiknowledge.com/searcha/id/help/h_author_names.htm>.

Author-ity. Arrowsmith Project Home Page. University of Illinois at Chicago. 1 December 2005. <<http://arrowsmith.psych.uic.edu>>.

Bekkerman, Ron, and Andrew McCallum. "Disambiguating Web Appearances of People in a Social Network." Proceedings of the 14th International

- Conference on World Wide Web. New York: ACM Press, 2004. 463-470. 10 February 2006. <<http://portal.acm.org/>>.
- Bhattacharya, Indrajit, and Lise Getoor. "Iterative Record Linkage for Cleaning and Integration." Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. Ed. Gautam Das, Bing Liu, and Philip S. Yu. New York: ACM Press, 2004. 11-18. 1 December 2005. <<http://portal.acm.org/>>.
- Cardina, Christen, and Donald Wicks. "The Changing Roles of Academic Reference Librarians over a Ten-Year Period." Reference & User Services Quarterly 44 (Winter 2004): 133-142.
- Cited Reference Searching: An Introduction. A Tutorial using Web of Science. Philadelphia: Thomson Corporation, 2004. 1 December 2005. <<http://scientific.thomson.com/tutorials/citedreference/crs3.htm>>.
- CrossRef. Lynnfield, MA: Publishers International Linking Association Inc. (PILA), 2003. 1 December 2005. <<http://www.crossref.org/>>
- Cruz, José Manuel Barrueco, Markus J.R. Klink, and Thomas Krichel. "Personal Data in a Large Digital Library." Research and Advanced Technology for Digital Libraries: 4th European Conference, ECDL 2000, Lisbon, Portugal, September 2000. Lecture Notes in Computer Science. 1923. Ed. Jose Borbinha and Thomas Baker. Berlin: Springer, 2000. 127-134. 1 December 2005. <<http://openlib.org/home/krichel/phoenix.letter.pdf>>.
- DBLP Computer Science Bibliography. Trier, Germany: Universitat Trier. 1 December 2005. <<http://www.informatik.uni-trier.de/~ley/db/>> and mirrored at <<http://www.sigmod.org/dblp/db/index.html>>.
- Delsey, Tom. "Authority Records in a Networked Environment." International Cataloguing and Bibliographic Control 33.4 (October/December 2004): 71-74.
- DiLauro, Tim, G. Sayeed Choudhury, Mark Patton, James W. Warner, and Elizabeth W. Brown. "Automated Name Authority Control and Enhanced Searching in the Levy Collection." D-Lib Magazine 7.4 (2001). 1 December 2005. <<http://www.dlib.org/dlib/april01/dilauro/04dilauro.html>>.
- Fellegi, Ivan P., and Alan B. Sunter. "A Theory for Record Linkage." Journal of the American Statistical Association 64.328 (December 1969):1183-1210.
- Fleischman, Michael Ben, and Eduard Hovy. "Multi-document person name resolution." 10 February 2006. <http://www.mit.edu/~mbf/ACL_04.pdf>
- Garfield, Eugene. "A Suggestion for Improving the Information Content of Authors' Names." Current Contents 6 (Feb 11, 1970). 1 December 2005. <<http://www.garfield.library.upenn.edu/essays/V1p077y1962-73.pdf>>.
- Getty Research Institute. Getty Union List of Artists Names. Los Angeles: J. Paul Getty Trust. 1 December 2005. <http://www.getty.edu/research/conducting_research/vocabularies/ulan/>.
- Han, Hui, Lee Giles, Hongyuan Zha, Cheng Li, and Kostas Tsioutsoulis. "Supervised Learning Approaches for Name Disambiguation in Author

- Citations." JCDL 2004 : Proceedings of the Fourth ACM/IEEE Joint Conference on Digital Libraries : Global Reach and Diverse Impact : Tucson, Arizona, June 7-11, 2004. Ed. Hsinchun Chen, Michael Christel, and Ee-Peng Lim. New York: ACM Press, 2004. 296-305. 1 December 2005. <<http://portal.acm.org/>>.
- Han, Hui, Wei Xu, Hongyuan Zha, and C. Lee Giles. "A Hierarchical Naive Bayes Mixture Model for Name Disambiguation in Author Citations." Proceedings of the 2005 ACM Symposium on Applied Computing. Ed. Lorie M. Liebrock. New York: ACM Press, 2005. 1065-1069. 1 December 2005. <<http://portal.acm.org/>>.
- Han, Hui, Hongyuan Zha, C. Lee Giles. "Name Disambiguation in Author Citations Using a K-way Spectral Clustering Method." Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries: Denver, June 7-11, 2005. New York: ACM Press, 2005. 334-343. 1 December 2005. <<http://portal.acm.org/>>.
- Hong, Yoojin, Byung-Won On, and Dongwon Lee. "System Support for Name Authority Control Problem in Digital Libraries: OpenDBLP approach." Research and Advanced Technology for Digital Libraries: 8th European Conference, ECDL 2004. Lecture Notes in Computer Science 3232. Ed. Rachel Heery and Elizabeth Lyon. Berlin: Springer, 2004. 134-144.
- IFLA UBCIM Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Functional Requirements for Authority Records: A Conceptual Model. Draft 2005-06-15. 1 December 2005. <<http://www.ifla.org/VII/d4/FRANAR-Conceptual-M-Draft-e.pdf>>.
- Kuhagen, Judith A. "Standards for Name and Series Authority Records." Cataloging & Classification Quarterly 21.3-4 (1996): 131-54.
- Lam, Ki-Tat. "XML and Global Name Access Control." OCLC Systems & Services 18.2 (2002): 88-96.
- Lancaster, Frederick Wilfrid. Vocabulary Control for Information Retrieval. 2nd ed. Arlington: Information Resources Press, 1986.
- Lee, Dongwon, Byung-Won On, Jaewoo Kang, and Sanghyun Park. "Effective and Scalable Solutions for Mixed and Split Citation Problems in Digital Libraries." Proceedings of the 2nd International Workshop on Information Quality in Information Systems, IQIS 2005. Baltimore, June 17, 2005. New York: ACM Press, 2005. 69-76. 1 December 2005. <<http://portal.acm.org/>>.
- Library Literature & Information Science Full Text. New York: H.W. Wilson, 2005. 1 December 2005. <<http://www.hwwilson.com/Databases/liblit.htm>>.
- Library of Congress Authorities Help Pages. Washington, DC: Library of Congress, 2005. 1 December 2005. <<http://authorities.loc.gov/help/contents.htm>>.
- Malin, Bradley, Edoardo Airoidi, and Kathleen M. Carley. "A Network Analysis Model for Disambiguation of Names in Lists." Computational & Mathematical Organization Theory 11.2 (2005): 119-139.
- Mann, Gideon S., and David Yarowsky. "Unsupervised Personal Name Disambiguation." Proceedings of the 7th Conference on Natural Language

- Learning. Edmonton, Canada, May 31-June 1, 2003. Ed. Walter Daelemans and Miles Osborne. 10 February 2006. <<http://www.cs.jhu.edu/~gsm/publications/pd.gsm03.pdf>>.
- MathSciNet. Providence: American Mathematical Society, 2005. 1 December 2005. <<http://www.ams.org/mathscinet/>>
- MathSciNet Author Database Help. Providence: American Mathematical Society, 2005. 1 December 2005. <<http://www.ams.org/msnhtml/authidhelp.html>>
- Monastersky, Richard. "The Number that's Devouring Science." Chronicle of Higher Education. 52.8 (2005): A12 (October 14). 1 December 2005. <<http://chronicle.com/weekly/v52/i08/08a01201.htm>>.
- NACO - The Name Authority Component of the PCC. Washington, DC: Library of Congress, 2005. 1 December 2005. <<http://www.loc.gov/catdir/pcc/naco/nacopara.html>>.
- Niu, Cheng, Wei Li, and Rohini K. Srihari. "Weakly Supervised Learning for Cross-document Person Name Disambiguation Supported by Information Extraction." Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics ACL 2004, Barcelona, Spain, July 2004. 598-605. 1 December 2005. <http://acl.ldc.upenn.edu/acl2004/main/pdf/372_pdf_2-col.pdf>.
- On, Byung-Won, Dongwon Lee, Jaewoo Kang, and Prasenjit Mitra. "Comparative Study of Name Disambiguation Problem using a Scalable Blocking-based Framework." JCDL 2005: Proceedings of the Fifth ACM/IEEE Joint Conference on Digital Libraries: Denver, Colorado, June 7-11, 2005. New York: ACM Press, 2005. 344-353. 1 December 2005. <<http://portal.acm.org/>>.
- Pasula, Hanna, Bhaskara Marthi, Brian Milch, Stuart Russell, and Ilya Shpister. "Identity Uncertainty and Citation Matching." Advances in Neural Information Processing Systems 15. San Mateo, CA : M. Kaufmann Publishers, 2003: 1 December 2005. <<http://books.nips.cc/papers/files/nips15/AP01.pdf>>.
- Patton, Glenn E. "Extending FRBR to Authorities." Cataloging & Classification Quarterly 39.3/4 (2005): 39-48.
- Patton, Mark, David Reynolds, G. Sayeed Choudhury, and DiLauro, Tim. "Toward a Metadata Generation Framework: A Case Study at Johns Hopkins University." D-Lib Magazine 10.11 (2004): 1 December 2005. <<http://www.dlib.org/dlib/november04/choudhury/11choudhury.html>>.
- RePEc Author Service. Storrs, CT: University of Connecticut, Department of Economics. 1 December 2005. <<http://authors.repec.org/>>.
- Snyman, Marieta M. M., and Marietjie Jansen Van Rensburg. "NACO versus ISAN: Prospects for Name Authority Control." The Electronic Library 18.1 (2000): 63-68.
- — —. "Reengineering Name Authority Control." The Electronic Library 17.5 (October 1999): 313-322.
- Spink, Amanda, and Maurice C. Leatherbury. "Name Authority Files and Humanities Database Searching."

- Online & CDROM Review 18 (June 1994): 143-148.
- Taylor, Arlene G. "Variations in Personal Name Access Points in OCLC Bibliographic Records." Library Resources & Technical Services 36 (April 1992): 224-241.
- TePaske-King, Bert, and Norman Richert. "The Identification of Authors in the Mathematical Reviews Database." Issues in Science & Technology Librarianship 31 (Summer 2001). 1 December 2005. <<http://www.istl.org/01-summer/databases.html>>.
- Tillett, Barbara B. "A Virtual International Authority File." 67th IFLA Council and General Conference, August 16-25, 2001, Boston. The Hague: International Federation of Library Associations and Institutions, 2001. 1 December 2005. <<http://www.ifla.org/IV/ifla67/papers/094-152ae.pdf>>.
- — —. "AACR2 and Metadata: Library Opportunities in the Global Semantic Web." Cataloging & Classification Quarterly 36.3/4 (2003): 101-119.
- — —. "Authority Control: State of the Art and New Perspectives." Cataloging & Classification Quarterly 38.3/4 (2004): 23-41.
- Torvik, Vetle I., Marc Weeber, Don R. Swanson, Neil R. Smalheiser. "A Probabilistic Similarity Metric for Medline Records: A Model for Author Name Disambiguation." Journal of the American Society for Information Science and Technology 56.2 (2005): 140-158.
- Uniquely Identifying Mathematical Authors in the Mathematical Reviews Database. Providence: American Mathematical Society, 2005. 1 December 2005. <<http://www.ams.org/mr-database/mr-authors.html>>.
- Warner, James W., and Elizabeth W. Brown. "Automated Name Authority Control." Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries JCDL '01, June 24-28, Roanoke, VA. New York: ACM Press, 2001. 21-22. 1 December 2005. <<http://portal.acm.org/>>.
- Web of Science. Philadelphia: Thomson Corporation, 2004. 1 December 2005. <<http://scientific.thomson.com/products/wos/>>.
- Web of Science 7.0 Workshop. Philadelphia: Thomson Corporation, 2004. 1 December 2005. <http://www.thomsonscientific.com/media/scpdf/wos7-workshop_en.pdf>.
- Weber, Jutta. "LEAF: Linking and Exploring Authority Files." Cataloging & Classification Quarterly 38.3/4 (2004): 227-236.
- Wellisch, Hans H. Indexing from A to Z. 2nd ed. New York: H. W. Wilson, 1995.

Appendix A: Sample of variations in instructions to authors for formatting names in the “references” section of submissions

Journal of Academic Librarianship (published by Elsevier): JAL follows the 15th edition of the *Manual of Style*, published by the University of Chicago Press. Examples: **Article from a Journal:** Paul Metz, ...

Guide for Authors. 1 December 2005,

http://www.elsevier.com/wps/find/journaldescription.cws_home/620207/authorinstructions

Information Processing & Management (published by Elsevier): You are referred to the *Publication Manual of the American Psychological Association*, Fifth Edition ... Examples: Fox, E. A. & Marchionini, G. ...

Guide for Authors. 1 December 2005,

<http://authors.elsevier.com/GuideForAuthors.html?PubID=244&dc=GFA>

Reference Services Review (published by MCB Press/Emerald): References to other publications should be complete and in Harvard style. (c) for articles: surname, initials,... e.g.Fox, S....

Author Guidelines. 1 December 2005,

<http://www.emeraldinsight.com/info/journals/rsr/notes.htm>

Appendix B: The presence/absence of name authority control in databases

Chart 1 includes several large library-subscribed and library-managed databases. Chart 2 describes databases that are not managed in traditional library environments.

Chart 1: traditional library-based databases	Is name authority control used?	How is control applied, or how must searcher identify and select name variations?
ABI/INFORM	no	identify and select from author index
ACM Digital Library	no	no author index
Chemical Abstracts	no	identify and select from author index
Compendex	no	identify and select from author index
EconLit	no	identify and select from author index
ERIC	no	no author index
GeoRef	no	identify and select from author index
Library Literature & Information Science	yes, but inconsistent	changes some author names to current version
INSPEC	no	identify and select from author index
MathSciNet	yes	mostly-automated name authority file
PAISInternational	no	identify and select from author index
PsycInfo	no	identify and select from author index
PubMed	no	truncates to initials, except post-2002 if full name on article
Web of Science	no	identify and select from author index; initials only used

Chart 2: non-library-managed databases	Is name authority control used?	How is control applied, or how must searcher identify and select name variations?
ANAC (Levy Project)	yes	automated name authority file
ArXiv.org	no	no author index
Author-ity	no	disambiguation based on probability
CiteSeer	no	no author index
DBLP Bibliography	no	select from author index
Getty Union List of Artist Names	yes	links among separate authority files
Google Scholar	no	no author index
HoPEc	yes	author-maintained registry
LEAF	yes	links among separate authority files

Illustrated below is a typical view of an author index that includes name variations. A searcher might select all the "CL" variations, but no searcher would know to scroll through to "Lee" without having noticed or known that Dr. Giles emphasizes his middle name.

GILES, C. L.
GILES, C. LEE
GILES, C. O.
GILES, C. R.
GILES, C. RANDY
GILES, C.A.
GILES, C.G.
GILES, C.H.
GILES, C.L.
GILES, C.LEE
[skip 144 lines].
GILES, L.J.
GILES, LEE

Appendix C - WilsonWeb

http://www.hwwilson.com/Databases/names_authority_control.htm

About Name Authority Control in H.W. Wilson's Indexing Services

H.W. Wilson controls names used as subjects. No user should have to search under multiple forms of a name. Personal names are cited consistently across all the Wilson indexes and databases.

Names are established according to the latest revision of AACR2, so H.W. Wilson names are consistent with conventional library cataloging. (The Names Department staff—who are responsible for maintaining the Wilson Names Authority File—are all professional librarians.)

New names are routinely checked against the Library of Congress's LCWeb names authority file, to ensure consistency with national cataloging standards. Chances are, names will be cited in H.W. Wilson files the same way as they appear in a library's own online catalog, if they are indeed the same person.

All personal name subjects are carefully checked against the individual periodical databases, including retrospective files, to avoid duplication and to distinguish between similar but different instances of names.

Similar but distinct names are distinguished from one another by expansion (e.g. inclusion of a full name instead of initials) or the addition of dates.

In cases where the form of a name is uncertain, H.W. Wilson Names Authority staff will search for an authoritative form in appropriate dictionaries, encyclopedias, and directories. The specific sources depend on the discipline, and on the dates and nationalities of the person in question.

H.W. Wilson Names Authority staff routinely establish cross-references from variant forms of a name to the form we cite. WilsonWeb users will be automatically switched from variants to preferred forms of names.

accessed 1 December 2005.

Copyright © 2006 by the H. W. Wilson Company. Material reproduced with permission of the publisher. Permission granted 2/13/2006.