

Notes on Operations

Creating Organization Name Authority within an Electronic Resources Management System

By Kristen Blake and Jacquie Samples

Staff members at North Carolina State University (NCSU) Libraries have identified the need for name authority control within E-Matrix, a locally developed electronic resources management (ERM) system, to support collection intelligence, the process of collecting, collocating, and analyzing data associated with a collection to gain a sophisticated understanding of its qualities for strategic planning and decision making. This paper examines the value of establishing authority control over organization names within an ERM system in addition to describing NCSU's design for conducting name authority work in E-Matrix. A discussion of the creation of a name authority tool within E-Matrix is provided along with illustrations and examples of workflow design and implementation for the assignment of authoritative headings. Current practices related to authority control and ERM systems in academic libraries and within organizations such as the Online Computer Library Center (OCLC) are also investigated and summarized to provide context for this project. Future possibilities for the use of this type of authority control on the part of librarians, vendors, and standards bodies are explored.

As electronic resources management (ERM) systems become more advanced and their use more widespread, libraries have begun to consider the potential of these systems to aid in collections decisions by performing advanced data analysis functions. Name authority control is of critical importance if ERM systems are to be put to this use because information drawn into a system from different sources must be collocated to produce accurate and useful analyses and reports. Throughout the development of E-Matrix, a homegrown ERM system, North Carolina State University (NCSU) Libraries has focused on the application's potential to facilitate effective collection intelligence, the process of collecting, collocating, and analyzing data associated with a collection to gain a sophisticated understanding of its qualities in order to strategically plan and make decisions. E-Matrix centralizes information from the library's catalog, link resolver, and assorted flat files within a single database and can perform analysis functions that include data from all of these sources. A challenge presented by this process is the identification and collocation of data elements imported to E-Matrix in a multiplicity of uncontrolled formats.

Data about organizations, such as the names of publishers, vendors, providers, and licensors of serials and electronic resources, has been the most difficult element to normalize within E-Matrix. Because organization names are imported from unformatted fields created for outside applications, the data in E-Matrix naturally lacks consistency. The names of organizations appear in dozens of variant and erroneous forms, with neither any indication of connections between entities that indicate business relationships nor authorized forms of names. In aiming to use E-matrix as a sophisticated reporting and collection intelligence tool, NCSU

Kristen Blake (kristen_blake@ncsu.edu) is Electronic Resources Librarian, Metadata and Cataloging Department, and **Jacquie Samples** (jacquie_samples@ncsu.edu) is Continuing and Electronic Resources Librarian, Metadata and Cataloging Department, North Carolina State Libraries, Raleigh.

Submitted June 16, 2008; tentatively accepted August 10, 2008, pending revision; revised and submitted September 4, 2008, and accepted for publication.

Libraries came to the conclusion that the application must apply authority control to its organization name data to correct these inherent irregularities. Following that decision, library staff members have implemented a project to create a set of singular authorized headings to control and normalize the organization name data stored in E-Matrix.

This paper reports on the process of creating authoritative data for organization names within E-Matrix at NCSU Libraries. The discussion begins with a brief literature review and an analysis of how libraries and library organizations have been using electronic resource management systems to manage organization names. It then describes the planning and implementation of an organization name authority at NCSU Libraries. The paper concludes with an analysis of future possibilities for data use and control within ERM systems.

Literature Review

Over the past decade, ERM systems have emerged as the accepted tool for storing and managing complex data about serial and electronic resources, including information about the organizations that publish, sell, and host those resources. As early as 2004, the Digital Library Federation's Electronic Resource Management Initiative (DLF ERMI) supported the use of ERM systems for tracking organization data, defining them, in part, as tools that would centralize data from disparate areas of large libraries to aid in the selection and evaluation of electronic resources.¹ The DLF ERMI report stresses the need to define standards and best practices for data elements stored within an ERM system, but does not delve into these features on a practical level.

Since the publication of the DLF ERMI report, no further literature has been published that examines

conceptually or practically the work needed to establish a data collection and analysis function within an ERM. The data management facet of ERM functionality has largely been eclipsed in practice by the more urgent need to implement licensing and workflow functions. Recent articles on ERM systems tend to mention data collection and reporting only briefly and as a corollary to broader processes, such as connecting support staff with needed management information or creating general tools for bibliographers.² Most often, the challenges of using the data collection functions of ERM systems are simply assigned to the realm of the future. While the need for ERM tools to facilitate collection intelligence and reporting remains in the professional consciousness, it has not yet been explored in a meaningful way.

Despite the lack of targeted discussion about the role of data collection within ERM systems, the general process of creating useful administrative metadata has been touched upon in other contexts and proves useful here. Hawthorne, as well as the DLF ERMI report, addresses the need for standards to avoid labor duplication within and between libraries working with ERM systems, but advises staying focused on broader design principles rather than addressing the nature of the data that will be collected and manipulated.³ Gorman, while not specifically addressing ERM systems, adds a layer of insight to the equation when he discusses the need for all metadata content to be subject to the same stringent requirements as bibliographic content. What use is a set of standardized fields when the data within those fields can vary so broadly?⁴ To be truly useful as a tool for access and collocation, he argues, the content of metadata fields must be subject to some level of authority control.⁴ The application of this concept to ERM data presents strong support for NCSU Libraries' decision incorporate a name authority into E-Matrix.

Current Practices in Authority Control and ERM Systems

To provide context for NCSU Libraries' organization name authority project, an informal survey of academic libraries known to have begun ERM system implementation was conducted to gauge the use of organization name data within these systems. Telephone interviews were conducted in October and November of 2008 with nine professionals from nine institutions, including Patrick Carr from Mississippi State University; Jill Emery from the University of Texas–Austin; Diane Grover from the University of Washington; Patricia Martin from the California Digital Library; Kim Maxwell from the Massachusetts Institute of Technology; Ophelia Payne from the University of Virginia; Clara Ruttenberg from Johns Hopkins University; Barbara Weir from Swarthmore College on behalf of the Tri-College Consortium; and Paoshan Yue from the University of Nevada–Reno. These discussions identified the duration and extent of each library's experience with its ERM; the functions each library supported, or planned to support, with its ERM; and the uses, if any, each library had found for data related to organizations.

The survey revealed that these libraries cover a wide spectrum in the extent of their ERM development. Of the nine libraries contacted, eight owned an ERM system, and one was in the process of evaluating a system for purchase after rejecting a previously purchased product. Six owned commercial systems, and two were transitioning from homegrown systems to commercial products. The products represented by the surveyed libraries included Verde by Ex Libris, Electronic Resources Management by Innovative Interfaces, and 360 Resource Manager by Serials Solutions. MIT was the most experienced ERM library, having created the homegrown system Vera in 1999 as a

FileMakerPro database. The four least experienced libraries had owned their current ERM systems for less than one year. Of the eight libraries that owned ERMs, four still considered themselves to be in the implementation stage. The other four considered their ERM systems functional, but indicated that they were still adding new features and hardly considered their systems done.

The functions these libraries supported with their ERMs were as varied as their stages of development. Librarians reported using their systems for managing licenses, maintaining holdings data, storing contact information for customer representatives, tracking orders, managing workflow, generating usage statistics, batchloading e-journal metadata to an online public access catalog, and storing information about product trials. The libraries that only recently acquired their ERMs tended to have implemented only one or two of these functions, primarily in the areas of licensing and holdings data, while the more experienced libraries had branched out into additional functions. Grover, electronic resources coordinator at the University of Washington, which has been using Innovative Interfaces' Electronic Resources Management since 2003, named at least a half dozen creative ERM functions under development at her library, including database and e-book management, Internet Protocol address range tracking, and SUSHI-compliant usage data feeds.⁵ (SUSHI stands for the *Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol Standard*, which defines an automated request and response model for the harvesting of electronic resource usage data through a Web services framework.)⁶

While an authority file of organization names could be beneficial to several of the above functions—specifically producing usage reports and storing publisher contact information—most of the libraries surveyed had not made use of an organization

name authority within their ERM systems. Of the librarians surveyed, five reported that their institutions were not yet far enough along in the implementation process to give the idea serious consideration. Martin, director of bibliographic services at the California Digital Library (CDL), which was still exploring new options for ERM systems at the time of the interview, said CDL's implementation team had expressed a desire for greater organization name control and would probably examine the situation more closely once it had selected a product.⁷ Weir, of Swarthmore College, said that the Tri-College Consortium (Swarthmore, Bryn Mawr, and Haverford colleges) was still involved in the basics of implementing Verde's workflow features and had not yet gotten to the point where it could think about reporting functions.⁸ All of the librarians who had not yet considered the use of organization names within their ERM systems acknowledged that the practice could be useful at some point in the future.

Three librarians said that their institutions had given some level of consideration to the problem of organization name authority and had decided that a solution was not necessary at the present time. Grover said the University of Washington Libraries had looked at using fixed fields within Innovative's system to differentiate between publishers and access providers. Ultimately, the staff decided that control of organization names could be useful, but didn't warrant an elaborate solution at the time.⁹ Emery echoed that perspective, saying that University of Texas–Austin's primary focuses for its ERM system were public access and workflow functions, not descriptive metadata.¹⁰ Carr, serials coordinator at Mississippi State University, said his institution has not needed the functionality of a name authority because it primarily uses organization names within its ERM system to store contact information for customer representatives, and

organization names supplied by Serials Solutions so far have been sufficient to support that function.¹¹

Of the librarians contacted, only Maxwell, serials acquisitions librarian and associate head of acquisitions and licensing services at MIT, said that her library had developed a fully realized solution for tracking organization names through its ERM system.¹² Maxwell described that solution, which dates back to the late 1980s. At that time, a librarian at MIT created a local database called Commitments to track serials pricing by publisher. As each new publisher was added to the list, an authoritative name was decided upon and maintained with each new entry. Relationships between publishers were also tracked as companies were bought and sold. As the needs of the library regarding electronic resources evolved, MIT integrated Commitments into Vera, its home-grown ERM system, and has kept the list up-to-date over the years. In 2008, MIT planned to abandon Vera in favor of Ex Libris's Verde system. Maxwell said that Verde would not have the same organization name capabilities as Vera, and would rely instead on a central knowledgebase maintained by Ex Libris. Maxwell anticipated the maintenance of publisher name data in Verde to be different and more confusing than the MIT's current system and said MIT will likely rely on the existing Commitments database until a new solution can be developed.

Speaking with colleagues in the academic library profession allowed a number of useful conclusions to be drawn about the role of organization name data in ERM systems. First, the control and manipulation of organization name data is not a project that many libraries have considered, often because of a lack of resources or expertise in the ERM implementation process. Second, organization name control holds varying degrees of importance for libraries. While some institutions may see it as an important component of their reporting and

evaluation practices, others regard it as optional or outside their focus. Finally, based on MIT's example and NCSU Libraries' own experience, organization name control is an issue that tends to be an enterprise venture originating in the library, since commercial systems do not usually facilitate the creation or maintenance of that data.

OCLC as a Source for Organization Name Authority

In addition to investigating the roles that name authority has played in electronic resources management in the academic library field, NCSU Libraries also sought context for its project from the examples of two recent authority-based initiatives originating at the OCLC Online Computer Library Center (OCLC). The WorldCat Registry (<http://oclc.org/registry>), a directory of institutional data about libraries and consortia and the services they provide, aims to function as a global authority file and may have value as a source for authoritative data about the institutions with which libraries interact. OCLC's publisher name server (<http://oclc.org/research/projects/publisherns>) is a research project to build a service that will normalize publisher names and provide users with other relevant metadata. Each project illustrates the importance of authority control in managing organization data, highlights some advantages and drawbacks of current approaches, and informs the process undertaken at NCSU Libraries.

The WorldCat Registry, which debuted in February 2007, has been marketed as a product that will allow libraries to manage the details of their identities (i.e., names, aliases, parent-child relationships, and IP addresses, among other elements) and make them available through a centralized database to a variety of third parties, including vendors, consortia, and other libraries.¹³ This product hits on the critical concept that metadata

about organizations is essential in a field where libraries and library service providers deal with many different groups on a daily basis. Just as libraries benefit from tracking information about content-providing organizations, they also have an incentive to ensure that subscription agents, vendors, and other service providers receive consistent and accurate data about the libraries themselves. In addition to providing data about libraries, the WorldCat Registry allows entries for publishers and other groups working in the library sphere. The data contained in those entries suggests the registry may be of some use in creating organization authority files within ERM systems.

Unfortunately, the WorldCat Registry suffers many of the same drawbacks that plague ERM data. Because organizations are responsible for keeping up their own entries, inconsistencies and inaccuracies often appear in the data. For example, a search for "North Carolina State University" returns nine results, including one heading for the university's main library and separate headings for each of its four branch libraries. No parent-child or other linking relationships have been established between them, even though a central body governs all five. While the existence of less than a dozen variations on the NCSU Libraries may be a step up from the scores of variant names that crop up in some ERM data, the standard of control needed for effective ERM functions is not met. In addition to concerns about the consistency of its data, the WorldCat Registry's self-maintenance policy also presents a challenge because without formal review and enforcement of content standards, many records for library-related organizations may remain incomplete or fall out of date. While it may serve as a useful reference tool, it cannot provide the level of detail and consistency needed by those wishing to create name authority within an ERM system.

OCLC's publisher name server, which is still in the research phase, presents a more tailored solution to the problem of organization name authority. The service aims to resolve variant publisher names to a single, authorized form and make available relevant data about each publisher, including its location, language, genre and format, subject areas, and parent and subsidiary companies. Lynn Silipigni Connaway, head of the project's research team, said that she and her associates originally viewed the project, much of which is being done algorithmically, as a data mining exercise.¹⁴ As their work progressed, she realized the advantages the server could offer to many areas of librarianship, chiefly collection intelligence and analysis. Additionally, the service has the potential to facilitate quality control in library catalogs, and may be of use to catalogers sometime in the future. While no prototype has yet emerged, Connaway reported that the publisher name server project has already generated a great deal of interest, including weekly inquiries from members of the library and publishing communities.

While OCLC's publisher name server offers many features that could be helpful to ERM system users in establishing name authority, it cannot be considered a full or viable solution at this time. As of late 2007, Connaway said that the service would only resolve book publishers, in keeping with OCLC's focus on projects and services that address monographic titles and holdings.¹⁵ Without attention to serial publishers, users will be left with an incomplete data set. Equally important, the publisher name authority is still under development, and many libraries might not be able to wait until a publicly accessible version of the application is released to begin exploring the use of name authority. While it may not be immediately compatible with efforts to establish authority control over organizations related to serials, OCLC's publisher

name authority server nonetheless demonstrates a need for organization name authorities and may provide context for librarians whose methods and research have already prompted similar projects.

Organization Name Authority at NCSU Libraries

At NCSU Libraries, institutional needs and the unique requirements of E-Matrix have required the development of an organization name authority tool fully integrated into the ERM system. The need for this tool was recognized nearly five years ago, when the original E-Matrix development team discussed the product's capacity to produce sophisticated evaluative reports. The team realized that to produce, for instance, accurate reports of money spent sorted by publisher, vendor, or provider, the names of those organizations needed to be consistent across all instances. Unfortunately, because organization name data is imported into E-Matrix from sources without authority control, the names found throughout the application would vary widely, as seen in figure 1, which illustrates the multiplicity of names associated with Elsevier. During that early period, creation of a name authority was proposed and approved as a solution to the problem of consistency, but implementation was postponed because of uncertainty about that implementation and a lack of precedent for organization authority in traditional library data sources.

As development of E-Matrix continued, additional justifications for a name authority arose. Design of the licensing module required the capability to link a license to specific titles through the licensor field. As with the reporting module, accuracy in the licensing module required a list of unique, authorized names from which users could choose when mapping a new license. For that list to exist,

authoritative names would need to be assigned to all organizations imported into E-Matrix.

Organization roles also strongly suggested the need for a name authority. The DLF ERMI report originally defined the concept of roles by describing how an organization could occupy a number of them, such as vendor, provider, publisher, licensor, and so on.¹⁶ By assigning one or more roles to an organization, an ERM system will avoid the needless duplication and confusion that might result from creating separate entities for each organization in each role.¹⁷ Again, this feature of E-Matrix could only function properly if organization names were assigned consistently throughout the data. Assigning multiple roles to an organization would make no sense if several variants of that organization's name could be found elsewhere. In short, authority control was crucial for clean relationships between organizations and roles.

In time, the concerns about data consistency raised by these aspects of E-Matrix made clear that the development of a name authority was not optional. A primary goal of E-Matrix was to bring together works related to one organization regardless of the form that name took in the original bibliographic descriptions. Without authority control, such correlation would be impossible. In light of those requirements and despite the lack of precedent in the field, NCSU Libraries deemed name authority creation a top priority and assigned its implementation to the library's metadata and cataloging department under the supervision of the continuing and electronic resources librarian. The project began in the fall of 2006 and still continues. The name authority

Quick Search > Results

81 items found, displaying 1 to 25
[First/Prev] 1, 2, 3, 4 [Next/Last]

Name
Elsevier Engineering Information
Pergamon Press - An Imprint of Elsevier Science
Elsevier Biomedical Press (Elsevier)
Elsevier Scientific Pub. Co (Elsevier)
Elsevier Science Pub. Co., etc (Elsevier)
Elsevier Scientific Pub (Elsevier)
Publications Elsevier (Elsevier)
Elsevier on behalf of the Federation of European Biochemical Societies
Elsevier Science Publishers, Biomedical Division (Elsevier)
Elsevier Publications
Elsevier Scientific Publishing Co
Elsevier Science Publishing Co
Elsevier Publishing Company
Elsevier Science Publishers (Biomedical Division)
Elsevier North-Holland
Published for the International Union of Biochemistry by Elsevier
Elsevier for the Trade Policy Research Centre
Elsevier Science Pub
âEditions Scientifiques Elsevier

Figure 1. Display of Elsevier Variants within the Organizations Module of E-Matrix

project has passed through planning, design, and implementation phases, each of which will be described subsequently along with a summary of results to the present.

Getting Started

Because no library or library-related group had previously attempted to compile an authoritative list of organization names within an ERM tool, NCSU Libraries' name authority project had to be developed in-house from scratch. The continuing and electronic resources librarian, Jacquie Samples, working as part of the E-Matrix product committee, developed a model for creation of the authority through a pilot project. By taking a small subset of organizations from the ERM data and assigning them authoritative names, she determined important specifications for the larger project, including a preliminary analysis of how assigning authorities would affect the library's ERM data, an estimate of the project's timeline, and expectations for the type of work that would be required to complete the authority.

The name authority pilot project assigned authoritative headings to 483 vendor names extracted from the library's integrated library system (ILS) in November 2006. Vendor names were chosen because they were deemed most useful for developing the licensing module, which had been identified as a priority around the same time. Working from a spreadsheet, the continuing and electronic resources librarian evaluated each vendor and chose a preferred name. Altogether, 444 authoritative vendor names were selected on the basis of predominant usage within the library community and assigned to the group of original names. Thirty-nine names (8 percent of the original sample) were variant names that could be linked through the use of an authoritative heading. Cross references were also established during the pilot project to represent business relationships between vendors—for example, companies that had been purchased by larger entities—as well as variant names not native to the data set. Additional roles were also noted for many of the entries, although they could not be incorporated into the ERM data at that time. The prevalence of variant names, complex business relationships, and multiple roles all provided strong evidence that the entire E-Matrix data set required authority control of organization names.

At the conclusion of the pilot project in April 2006, Samples estimated that approximately forty hours had been spent assigning names and cross references, as well as making notes about interesting or unusual circumstances. Using the estimated number of organization names in E-Matrix at that time (about seven thousand) she determined that to assign authoritative names to the entire data set would take 580 hours—about three and a half months of full-time work for one person. Given the project loads of the librarians in the technical services departments at NCSU Libraries, the

project likely would take the better part of a year for one librarian to complete.

In addition to the sheer amount of time that would be required to create a fully functioning name authority, the pilot project also revealed that much of the work would be composed of manual functions, including collocating names that represented the same entity and determining authoritative headings for each group of names using library and serials industry resources. These tasks would require the expertise of a library staff member familiar with the nature of serials and common library-based information sources. To determine authoritative names, the individual performing the work would require a strong sense of the relationships that exist between libraries and organizations. The librarian also would need the ability to follow a set of guidelines that combine accepted authority practices with the needs of the university and the library. The guidelines would not be prescriptive, and the selection of names would require a strong element of judgment and a fluid and intuitive use of available tools. The ultimate goal would be to determine headings for each organization that best fit with the library's existing practices and the overall structure of the authority. In light of the nature of the name authority work, the possibility of selecting names in an automated fashion was rejected because it was uncertain if an automated system could effectively make the judgments needed to choose correct authoritative names.

The demonstrated importance of the name authority project, along with the volume and intensity of the work it would require, convinced the E-Matrix committee and library administration that the authority tool deserved top priority. The continuing and electronic resources librarian and a programmer from NCSU Libraries' IT department were assigned as part of their E-Matrix work the tasks of

designing a system to store and manage name authorities and of using that system to assign authoritative names to each organization in the licensor, provider, publisher, and vendor roles within E-Matrix. An NCSU Libraries Fellow, Kristen Blake, was appointed to aid in the determination and assignment of authoritative names. Together, this group made up NCSU Libraries' E-Matrix name authority team. With the support of library administration and increased resources, the name authority had moved into the realm of the possible.

Defining Structure

Before the name authority project could be fully implemented, a module had to be designed within E-Matrix to manage preferred names and other authority data. The name authority team created a framework that could be integrated into the ERM system functions that had already been designed while remaining true to the vision of name authority established by the librarian working on the product's conceptual development.

The original plan for the organization name authority within E-Matrix envisioned a record-based structure to link together related organization names and store descriptive data about each organization. Each authorized name would be stored on a record and connected to the work-level records of all resources featuring one of its variant names. The authorized name record would have the ability to store data relevant to the authorized heading as well as its variants. Variant names would remain in place on the work-level records to which they originally belonged, and those records would be used to store data unique to each variant. Additionally, the variant names on each work-level record would be assigned one or more of the roles available in E-Matrix: licensor, provider, publisher, and vendor.

License records would also be linked to an authoritative name record through the work-level record and the nonauthoritative names associated with that work. All of these data—titles of works, unauthorized names, and organizational roles—would loop back and display on the authoritative record. Figure 2 shows the relationships between records that link name authority headings to work-level records in E-Matrix.

In addition to establishing associations between resources and organizations, the initial vision of the name authority tool also included the ability to store useful data on organization records. Authoritative name records would serve as the primary storage location for business history notes, internal remarks, vendor contact information, product trial details, and other information as needed. Nonauthoritative name records could also be used to store notes specific to a single variant or imprint. By incorporating a detailed record structure into the name authority, the library hoped to mimic the structure of established name authorities that serve as repositories of historical and local information, as well as guides for consistent and accurate data creation.

The realities of implementing E-Matrix forced the library's programmers and planners to apply a phased approach to the design of the name authority application. In the interest of getting the project started as quickly as possible, some of the features in the original proposal were assigned to later phases, and a name authority system was designed that included essential functions but did not yet incorporate more robust design features. The system links organization name data through unique identifiers in a relational database. A member of the name authority team can assign a name authoritative status, which places a property in the database record for that name, indicating that status. Once a nonauthoritative name

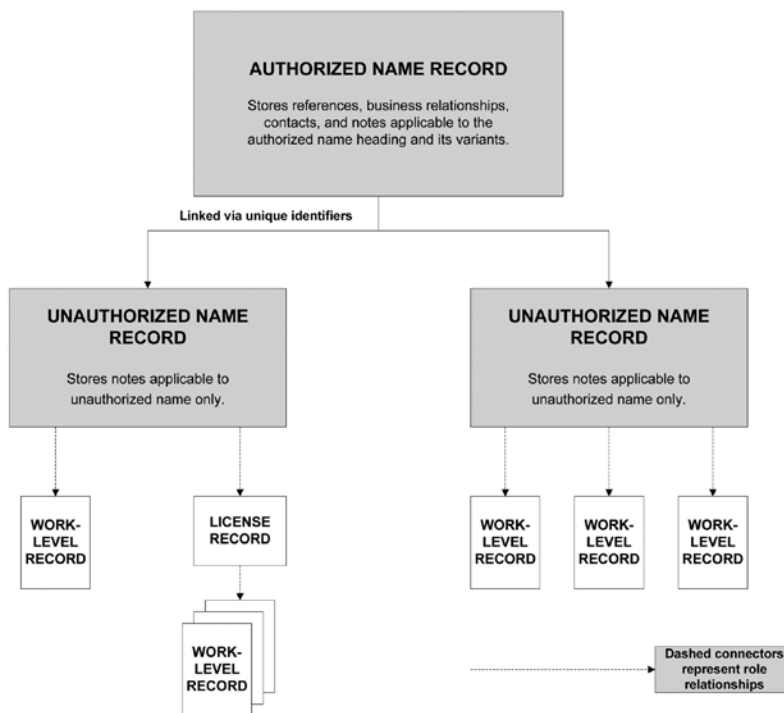


Figure 2. Envisioned Structure for the Name Authority Tool

is associated with an authoritative heading, database records for nonauthoritative names and their roles link to that name algorithmically. When an authorized heading has been assigned to a certain organization name, all future instances of that name will be automatically subsumed into the existing hierarchy. This essential feature will decrease the amount of maintenance necessary once the authority project has been completed.

The inclusion of a formal record structure for storing historical and internal data was also transferred to a later phase in the design of the authority tool. In its current incarnation, instead of using formal records like those a cataloger might find familiar, the names within the authority function more like related nodes without a formal record structure. E-Matrix searches the authority database on the fly and, when data intersects, the relationships between names and roles are displayed within the E-Matrix user

interface or on a report.

Within this design scheme, the system of roles assigned to each organization acts as a temporary substitute for the hierarchy of business relationships envisioned for the authority. While the authority team cannot currently assign, for example, Elsevier as the current owner of Pergamon Press, the roles assigned to each organization can be manipulated to produce reports that would reflect that same relationship. Because Pergamon may be assigned to a resource in the publisher role, and Elsevier assigned to that same resource in the provider role, a report listing all titles for which Elsevier is the provider along with their additional roles will indirectly display the relationship between Elsevier and publishers, like Pergamon, that it has acquired.

The current system allows E-Matrix to accomplish the primary task of collocating all resources associated with a certain organization,

Organizations > Name Authority > Select Organizations (Step 1 of 4)

Organization	Authoritative Organization	Roles	Selected
Duke University School of Law Journals			Yes
Duke University. Rare Book, Manuscript, and Special Collections Library			No
Duke University Press	Duke University Press		No
Published by Duke University Press under the editorial direction of Mediterranean Affairs Inc			No
Duke University School of Law			Yes
School of Law, Duke University			Yes
Published for Hopscotch Inc. by Duke University Press			No
Duke University Press, etc	Duke University Press		No
Slavic and East European Language Resource Center, Duke University			No
The Duke University Press	Duke University Press		No
Duke University, Dept. of Geology			No
Duke University, School of Law			Yes
Duke Journal of Gender Law & Policy			No

Figure 3. Initial Interface Display of the Organization Name

though the lack of formal records delays the library's goals of using its ERM system to capture the syndetic structure of business relationships and use the ERM as a primary location for storing these data. The eventual addition of that functionality to E-Matrix will allow the name authority project to move forward and more strongly fulfill the local needs identified early in the development of E-Matrix.

Designing an Interface

Once this structure was in place, the next step in the implementation process was the design of a convenient and intuitive way to assign and manage authoritative names. Since the start of the project, the name authority team has worked to design an interface within E-Matrix that would allow the authority to be easily accessed and manipulated. For this phase of the project, the technical services librarians contributed ideas for the design and functionality of the interface, while the programmer translated this vision into a series of interfaces, each enhanced and refined with feedback based on actual use.

The first authoritative organization

names were initially stored in a local database because the user interface for the E-Matrix authority had not been completed at the start of the project. Instead, the librarians assigning authoritative names used a Microsoft Access database for storage as they began working through a list of 7,858 publisher names culled from the library's holdings in the SFX (Ex Libris' link resolver) knowledgebase. These publisher names were chosen not only because they provided sufficiently complex test data for the developing interface, but also because they represented the names most desperately in need of authority control. (Publisher data from the library catalog were not initially included because much of it was duplicated in the SFX data, and the remaining publisher names could be controlled later by building on the headings already determined during the SFX phase.) The librarians evaluated each name and collocated it with others representing the same publisher, recorded their decisions in the database, and made notes specifying the justification for each decision and any problems that might need to be investigated when the E-Matrix interface became functional. This evaluative

period resulted in the determination of authoritative headings that could be used later, as well as suggestions for the design of the integrated interface planned for E-Matrix.

Throughout this early stage, designing a name authority interface shared top priority status with assigning names. Based on input from librarians who had tested the interface, the E-Matrix programmers produced a rudimentary beta interface, which was then adopted on a trial basis as part of the process of assigning names. The interface allowed the name authority team to select, on one screen, an authoritative organization name, as well as a group of names that should fall under its domain, and record that relationship within E-Matrix. This interface allowed the authority team to transition from recording their decisions only in the local database to actually entering them into E-Matrix, where they could be used experimentally by library staff. The beta interface had many limitations, however, and the local database was maintained as a backup and a place to record comments and problems.

The authority team continued working within E-Matrix, and the

Organizations > Name Authority > Select/Create Authoritative Organization (Step 2 of 4)

Select Authoritative Organization

Organization Select

Create Authoritative Organization

Authoritative Organization Name Create

[Previous Step](#)

Organizations > Name Authority > Confirm Organizations (Step 3 of 4)

Duke University, School of Law will be assigned as the Name Authority for the following organization(s):

- Duke University School of Law Journals
- School of Law, Duke University
- Duke University School of Law
- Duke University, School of Law

[Previous Step](#) [Finish](#)

Organizations > Name Authority > Confirm Organizations (Step 4 of 4)

Duke University, School of Law was assigned as the Name Authority for the following organization(s):

- Duke University School of Law Journals
- School of Law, Duke University
- Duke University School of Law

New Name Authority

Figure 4. Remaining Steps in Assigning an Authoritative Organization Name

team's observations helped the programmers develop an improved name authority interface, which was released with E-Matrix 1.0 in December 2007. This interface allows names to be assigned using a simple four-step process. In step 1, the browse or search feature is used to identify and select names that will be collocated under a common authorized heading. Figure 3 shows the results of a search for "Duke University" and lists the names that need to collocate.

In step 2, an authoritative name can be chosen from among those selected or a new name entered manually. In step 3, the user can perform a final review of selections and assignments and submit them. The fourth and final step confirms the assignments made and offers the user a link

to begin the process again. Steps 2–4 are shown in figure 4.

The new interface made the team's work easier because it included expanded display options that clarified whether an organization already had an authoritative heading assigned to it and what that heading was. The authoritative relationships were also displayed in more places within E-Matrix, making the product useful as a source of organizational data for the authority team and all library staff using the ERM. In comparison with the beta version, the first production interface was more intuitive, incorporating browse functionality, a cleaner design, and the ability to correct errors by reassigning authoritative headings. With the addition of a notes feature in a future release, the team will

transition their workflow completely into the ERM system.

Crafting Policies

At the same time that an interface was being designed, the continuing and electronic resources librarian was working to create the practical guidelines and specifications needed to choose authoritative names and assign them to organizations within E-Matrix. The most important deliverables were guidelines that ensured accuracy and consistency in the selection of authoritative names and the identification of organizations that will fall under them and a set of tools to help apply those guidelines.

Creating clear, accurate authoritative headings that were useful for the NCSU Libraries staff has been the primary consideration in deciding how names should be assigned within E-Matrix. In the interest of local policies, the first step taken in establishing name selection guidelines was to consult the library's collection managers. Because they were the people who would be making use of the authority data most frequently for collection evaluation, the authority needed to reflect their preferences and standards. The collection managers indicated that their chief goal was to preserve the organization name that most directly reflected the intellectual content of a work. Most commonly, this directive affected the assigning of authoritative publisher names, which often are more directly tied to specific content areas than a vendor or provider name. Thus, to maintain that connection, the original publisher of a title is almost always chosen as the authoritative heading, even if that publisher has since merged with or been acquired by another entity. In other words, the "statement of responsibility" takes precedence over any current business arrangements. Journals that have been published by Academic Press, for example, are kept under the

Academic Press heading, even though Academic Press has long been an imprint of Elsevier.

In many cases, publisher statements that have been imported from the library catalog are long and convoluted. Large societies often issue publications on behalf of smaller societies and, in these cases, collection managers presume that the larger society is acting as a kind of benefactor, while the smaller society represents the creator of the journal's content and, therefore, the better choice for an authoritative heading. For example, a publisher statement reading, "Published for the International Union of Biochemistry by Elsevier" would be assigned the authoritative name "International Union of Biochemistry" because that group entity is presumed to be responsible for the biochemistry-related content of the titles associated with the organization. Elsevier's role in the creation of the material is not lost because it can be assigned to those resources as a provider and licensor.

In another common scenario, groups of small societies publish a title jointly, and the publisher statement contains the names of two or three discrete entities with each given equal weight. E-Matrix's current functionality allows for only one authoritative organization heading to be assigned to each work. In these instances, every effort is made to determine which organization is chiefly responsible for the work in question and to assign authority accordingly. For example, considering the publisher statement "American Society for Environmental History and Forest History Society," the librarian determining the authoritative name must conduct a thorough investigation that includes identifying and viewing publications linked to this statement, then reading front matter and publisher information to determine which publishing group is the primary contributor to the content of the title.

Because NCSU Libraries holds a

fair number of foreign language titles, the formatting of the authoritative names for these titles is also important to the library's collection managers. They requested that all foreign language authorities using Roman script remain in the original language. Authority headings for titles using non-Roman scripts have been translated into English rather than transliterated. E-Matrix does not allow for the inclusion of non-Roman characters, and translations were determined to be clearer and easier to assign than transliterations.

Beyond these few specific requests, NCSU Libraries' collection managers and the E-Matrix development team felt comfortable relying on traditional library resources for the determination of names. The Library of Congress Name Authority File (LCNAF) (<http://authorities.loc.gov>) has been used whenever possible as a source for preferred names, as long as they do not conflict with local customizations. If no heading is available from the LCNAF, trustworthy serials databases such as Ulrich's Periodicals Directory (<http://ulrichsweb.com>) and the ISSN Portal (<http://portal.issn.org>) have been used as secondary sources. NCSU Libraries' sources correspond almost exactly to those made by Connaway and Dickey of the OCLC's monograph-focused publisher name server.¹⁵ That project uses the LCNAF as the chief source of authorities, followed by Books in Print (<http://books.inprint.com>) and the International ISBN Agency (<http://isbn.org>), monographic counterparts of Ulrich's and the ISSN Portal, respectively.

Unlike OCLC, which has formalized its choice of sources for name selection, NCSU Libraries has the flexibility to put its local needs first. Rather than make a hard and fast rule about the order of sources consulted, the librarians at work on the name authority project have the freedom to make decisions on the basis of what will best fit the library's needs. The

decision to disregard name changes, mergers, and other business-related changes until richer syndetic functions can be incorporated into E-Matrix illustrates a fundamental application of the local needs principle. Another example is the decision to apply *Anglo-American Cataloguing Rules*, 2nd ed., rules for formatting of complex names to all headings, regardless of their source.¹⁹ This practice ensures consistency and, just as importantly, enhances the browsability of organization data. When all departments, research centers, publishing arms, and other offshoots of a major institution are grouped together using consistent formatting, finding and grouping resources emanating from a particular group becomes easy, while still retaining more detailed information about the subgroups involved in its production. E-Matrix's browse features, as well as the sorting capabilities of the reports module, are greatly enhanced by this practice.

Librarians assigning authoritative names are also encouraged to use their knowledge of local practices to make case-by-case exceptions when necessary. Often, these types of decisions are used to resolve small quirks that might never be addressed by a stricter set of rules. For example, the authoritative name chosen for the Institute of Electrical and Electronics Engineers is IEEE, even though the LCNAF has authorized the full version of the name. While working through the list of organizations, the continuing and electronic resources librarian recognized, for example, that the acronym IEEE was simply more recognizable to library staff than the organization's rarely used full name. The full version of the name will not be lost from the name authority data, however, because it is stored as a searchable cross reference. On both the conceptual and practical levels, the flexibility to tailor the name authority specifically to NCSU Libraries' interests has been essential to its success.

Equally important as the authoritative sources are other sources that provide insight into unique or problematic names not addressed in the traditional library databases. Many of the publisher names came in the form of obscure initialisms, and, in these cases, the website Acronym Finder (<http://acronymfinder.com>) suggested leads that eluded typical research sources. In general, the Web search engines proved vital to researching the kind of obscure names that are routinely not found in the LCNAF. Often, use of E-Matrix itself was necessary to find the name of one or more journal titles associated with a specific publisher, and then a Web search engine was used to trace those titles back to the primary source. Viewing the publisher name in a table of contents or on an authoritative website provided an extra level of confidence in the decision-making process for publishers not found in an established source.

The combination of an E-Matrix interface designed with user input and a set of fluid guidelines for name selection has made the day-to-day work of assigning name authorities a smooth and intuitive process. While unexpected and unusual names can slow what has become a speedy process, these are usually resolved through discussion and attention to the established guidelines and the library's needs. Again, the flexibility of the assigning process and the value placed on librarians' judgment have been essential to the implementation of this original and complex project.

Organization Name Authority at Work

The tools and procedures needed to create a functional organization name authority within E-Matrix were put into place by summer 2007. As of this writing, practical implementation of the name authority project has been underway for nearly a year. During that period, the NCSU Libraries

has focused on the establishment of authoritative names for organizations stored in E-Matrix and in the use of those names to enhance data integrity, licensing procedures, and reporting capabilities.

As of March 2008, the name authority team had evaluated 1,319 organization names, not including those names originally evaluated in the pilot project. After authority control was applied, this group of names was reduced to 532 authorized organization names, a 59 percent reduction. These results were significantly more dramatic than the 8 percent drop seen in the pilot project, but that discrepancy can be explained by the post-pilot focus on normalizing the library's big deal packages. For instance, 58 Elsevier variants were reduced to only one authoritative name, a 98 percent reduction in the number of variants and errors, contributing to a more dramatic decrease overall. Similarly, the 532 assigned authoritative names relate to 21,672 titles, a substantial percentage of the more than 35,000 unique manifestations of resources currently managed through E-Matrix, further confirming the widespread effects of controlling the names of major organizations.

On the broadest level, the introduction of authority control of names into E-Matrix has demonstrated progression toward cleaner, more usable implementation of data on the basis of the concept of roles as defined in the DLF ERMI. By eliminating duplication, variation, and errors in the organization data, E-Matrix enables use of role relationships. Titles that share a common publisher, vendor, or provider can now be identified using the authoritative name that links them to a single organization. Without authority control, an individual would have to manually account for each organization variant every time he or she worked with ERM data.

The E-Matrix licensing module also benefits from the use of clean

data. As the library begins its license mapping process, human data entry will be used to map the details of a license into structured data elements within E-Matrix. Each organization name entered into the license form must correspond to the correct serial resources. To ensure that a license is correctly applied to all resources whose terms it dictates, the licensing module will allow only authoritative names to be entered into the licensor field. Limiting the available licensors within E-Matrix preserves the appropriate use of roles and relationships throughout the data.

The benefits of clean role relationships can be seen even more substantially in E-Matrix's reporting module. The name authority team, in collaboration with collection managers and programmers, has begun to test the capabilities of the module to incorporate authoritative names in ways that enhance the comprehensiveness and flexibility of reports. An authoritative publisher report displays every serial title in E-Matrix associated with a publisher. This very basic report serves mainly as a test object to illustrate how authoritative names have been incorporated into the data. Within E-Matrix, users can link from this report to detailed displays of publisher and resource information, facilitating discovery of related organizations and titles. Using the report module's export tool to transfer the data to a spreadsheet or database, all resources published by the same entity can be easily identified, and groups of similar publishers explored. A similar test report displays each authoritative organization along with its related titles. This report expands the authoritative publisher report across all roles, allowing for a more complete picture of how organizations relate to works and highlighting implicit business relationships by illustrating the multiple organizations associated with a single resource through their respective roles.

In addition to these preliminary

reports, the name authority team also has conceived more advanced reporting using authoritative names. For example, by leveraging the subject categories that have been assigned to every resource in E-Matrix, reports can be generated listing the most common authoritative publishers within any given subject area. That data could be combined with the money spent on each resource to provide a comprehensive picture of the amount spent per publisher in a certain subject area. Without the authoritative name data, such advanced reporting would be much more difficult because the proliferation of variant organization names would make the process of identifying and collocating all instances of a publisher tedious and prone to error. The name authority team hopes to see staff from other departments take advantage of the clean relationships between organizations and resources to produce analogously sophisticated custom reports.

To ensure continuing data integrity across all modules of E-Matrix, the organization name authority project will continue to be maintained and enhanced once the primary authority project has been completed. Monthly maintenance reports will list any new organization names that have come into E-Matrix since the authority was last updated. Organization names with existing authorities will automatically be subsumed under the proper heading. In this way, the library will retain control over organization data within E-Matrix.

Conclusion: The Future of ERM Systems and Name Authority

The name authority project marked the start of the NCSU Libraries' application of authority control on data within its ERM system. The demonstrated need for authority control at NCSU Libraries and indications of

similar thinking at other institutions make clear that authority control is poised to emerge as an important issue in the serials and electronic resources field. As this issue evolves, NCSU Libraries aims to improve local practices and participate in initiatives that span the library community.

The name authority team plans to pursue the enhancement of the existing tool and its functions within E-Matrix. An ongoing development priority is the expansion of the tool's structure to more closely align it with the vision established at the outset of the project, namely the creation of a full record structure that would allow for the storage of contact information for technical and sales representatives, details of product trials, and other internal notes as needed. Such detailed records will support the library's goal of creating an ERM system that facilitates storage of the myriad details of transactions related to serials and electronic resources.

Equally important will be the creation of a hierarchical structure to identify and describe business relationships between organizations. These connections may be represented through simple linking relationships similar to the role relationships that link organizations to resources. Among the relationships suggested for this structure include business-centered relationships such as "purchased by," "merged with," or "split from." Alternately, a complex, record-based structure could provide a more sophisticated representation of the series of acquisitions and mergers that characterize the serials industry. The ability to link organizations in a hierarchy would result in a dynamic, family tree-like structure more illustrative than the flat linking structure currently used in the assignment of roles. In either case, the capture of business data about serial publishers remains a top priority.

In addition to enhancing the structure of the E-Matrix authority tool, the

name authority team also intends to evaluate and streamline the process of investigating and assigning authoritative names. With the project under way and procedures in place for the determination of names, several strategies may help the name authority team in its task of evaluating thousands of organizations. In addition to adding staff to the project, algorithmic text analysis of organization name data offers several potential new courses for the name authority project. While the name authority team initially dismissed the prospect of using an automated process to parse organization names and choose the most appropriate heading, immersion in the process has shown that the vast majority of organization names are small publishing companies, self-publishers, and associations—many of whom are responsible for the publication of only one resource in the library's collection. Evaluating each of these types of names one by one has been extremely time consuming and does not make the best use of library staff resources.

One option is to use textual analysis to identify similar names, choose a likely authoritative name, and assign that name as a heading. Another would be to algorithmically group similar names, but then manually choose the authority. In both cases, the machine-based solution would result in rough authority control over a large set of rarely used organizations. In either situation, the E-Matrix name authority tool still would enable any authority to be manually evaluated and changed upon request. Any organization names not included in the textual analysis would be assigned a priority and worked on as time allowed. No decisions have yet been made on the role of textual analysis in the creation of authority headings, and additional consideration will be necessary before the name authority team changes its policy of manual evaluation for all organizations.

Finally, the name authority team plans to turn its efforts to the use

of authoritative headings in E-Matrix end-user displays. E-Matrix has not yet taken advantage of the potential of authoritative headings to improve the interfaces used by staff across the library. In many display functions, E-Matrix continues to use nonauthoritative organization names where authoritative names would produce a clearer, more coherent view of the library's serial holdings. To make the best use of the data being created, the name authority team plans to take a comprehensive look at each of E-Matrix's modules, determine where and how organization names are used, and outline the most effective type of display name to use in each circumstance.

In addition to enhancing the local uses of its name authority work, staff at the NCSU Libraries also aim to explore growing awareness in the library field of the need for authority control in the context of ERM. Involvement beyond NCSU Libraries so far is preliminary but has the potential to take many forms, including data and technology sharing with other libraries, collaboration with vendors in related efforts, and monitoring of and participation in standards groups that examine the issues of organizational identities.

The responses to the survey conducted for this paper, as well as in feedback gathered from a presentation about the E-Matrix name authority tool at the 2008 Electronic Resources and Libraries conference in Atlanta, have demonstrated broad recognition of the need for organization authority control. Feedback indicated that many electronic resources librarians have begun to view development of reporting aspects of ERM tools as an important consideration for the future. These librarians have acknowledged that an organization name authority would be very useful for reporting functions. They have also made known that the data contained in such a name authority would have value outside of ERM systems as a reference

source for all librarians who work with vendors, publishers, and providers. Because many libraries do not yet have the experience or resources necessary to implement organization name authority, the librarians who contributed their opinions to this project expressed a strong interest in sharing organization name authority data and the tools used to manage and create it. As development of the name authority tool at NCSU Libraries continues, the team will work with the E-Matrix administrative group to explore possible methods for sharing the data, the tool, or both.

In addition to librarians, vendors have also emerged as supporters of this enterprise. The publisher name authority and WorldCat Registry products under development at OCLC represent the efforts of a major library service provider to establish authoritative identities for monographic publishers as well as for libraries and consortia. The OpenIdentify Look-Up Service (<http://openidentify.com>), a product recently developed by the journal supply chain support provider Ringgold, marks another vendor-driven effort to establish widespread organization name authority control. OpenIdentify has assigned unique identifiers to more than one hundred thousand subscribers to academic journals and organized them hierarchically. Like the WorldCat Registry, OpenIdentify approaches the problem of organization name authority from the perspective of a service provider with a need to control the identities of its clients: universities, libraries, corporations, and other information institutions at the purchasing end of the serials and electronic resources transaction. The dataset of authoritative publisher, vendor, and provider names would be an ideal complement to vendor-created name authorities controlling serials purchasers. Together, these two authorities would cover both ends of the serials transaction.

Standards bodies will also be a

presence in any organization name authority efforts that span the library and information field. The National Information Standards Organization (NISO) has approved the formation of a working group dedicated to the creation of a standard for institutional identifiers for libraries and publishers. The group will establish the metadata elements required for use of institutional identifiers as well as develop use cases for the standard.²⁰ The NISO work will continue efforts begun by the Journal Supply Chain Efficiency Improvement Pilot (<http://journal.supplychain.com>), a collaborative project dedicated to exploring the benefits of institutional identifiers and preliminary implementation strategies. By involving a standards group like NISO in the process of creating and disseminating authoritative names, the project will gain legitimacy and avoid conflicts of interest that could arise from close association with a single vendor or institution. While this level of development for organization name authority initiatives is still in its infancy, the inclusion of such projects in the agenda of standards groups recognizes the need for and is a step toward a solution that could integrate the work already begun by a variety of groups.

The implementation of an organization name authority to enhance electronic resources collection intelligence has been an innovative and successful venture at NCSU Libraries. Imposing authority control on the organization name data within the ERM system has laid the groundwork for greater precision and comprehensiveness in NCSU Libraries' reporting and collection analysis efforts as well as contributed to the creation of cleaner, more accurate data. Exploration of name authority practices throughout the library and vendor communities have confirmed that a need exists for the control of organization data and demonstrated that opportunities abound for collaboration and enhancement

of the original project. Management of serials and electronic resources is often a complex and difficult endeavor. NCSU Libraries' organization name authority illustrates the power of creating an ERM tool to meet specific local requirements and the potential benefits of expanding it to address the broader needs of the library and information community.

References

1. Timothy D. Jewell et al., *Electronic Resource Management: Report of the DLF ERM Initiative* (Washington, D.C.: Digital Library Federation, 2004), www.diglib.org/pubs/dlf102 (accessed Oct. 24, 2007).
2. Diane Grover and Theodore Fons, "The Innovative Electronic Resource Management System: A Development Partnership," *Serials Review* 30, no. 2 (2004): 110; Tony A. Harvell, "Electronic Resources Management Systems: The Experience of Beta Testing and Implementation," *Serials Librarian* 47, no. 4 (2004): 110–16.
3. Darlene Hawthorne, "Administrative Metadata to Support the Acquisition of Continuing E-Resources," *Serials Review* 29, no. 4 (2003): 276–78.; Jewell et al., *Electronic Resources Management*.
4. Michael Gorman, "Authority Control in the Context of Bibliographic Control in the Electronic Environment," *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 11–22.
5. Diane Grover, phone interview with Kristen Blake, Oct. 26, 2007.
6. *The Standard Usage Statistics Harvesting Initiative (SUSHI) Protocol*, ANSI/NISA Z39.90-2007 (Baltimore: National Information Standards Organization, 2007).
7. Patricia Martin, phone interview with Kristen Blake, Oct. 5, 2007.
8. Barbara Weir, phone interview with Kristen Blake, Oct. 11, 2007.
9. Grover, phone interview.
10. Jill Emery, phone interview with Kristen Blake, Nov. 12, 2007.
11. Patrick Carr, phone interview with Kristen Blake, Oct. 23, 2007.
12. Kim Maxwell, phone interview with Kristen Blake and Jacquie Samples, Oct. 15, 2007.
13. OCLC, "WorldCat Registry Offers Management of Organizational Data," news release, Feb. 26, 2007, www.oclc.org/americalatina/es/news/releases/200652.htm (accessed Nov. 27, 2007).
14. Lynn Silipigni Connaway and Timothy Dickey, phone interview with Kristen Blake and Jacquie Samples, Oct. 1, 2007.
15. Ibid.
16. Timothy D. Jewell et al., *Electronic Resource Management*.
17. Jewell et al., *Electronic Resources Management*.
18. Connaway and Dickey, phone interview.
19. *Anglo-American Cataloguing Rules*, 2nd ed., 2002 rev., 2005 update (Chicago: ALA; Ottawa: Canadian Library Association; London: Chartered Institute of Library and Information Professionals, 2005).
20. Cynthia Hodgson, "Call for Participation: New NISO Working Groups on Institutional Identifiers and Knowledge Bases and Related Tools," online posting, Jan. 18, 2008, SERIALST, <http://list.uvm.edu/cgi-bin/wa?A2=serialst;wLRVmq:20080118120133-0500> (accessed Jan. 18, 2008).