The Name Game: Creating Organization Name Authority within an ERM System

Kristen Blake and Jacquie Samples
North Carolina State University Libraries

Electronic Resources and Libraries
Atlanta, GA – March 20, 2008

Introductions:
Kristen
Library Fellow in Acquisitions and M&C

Jacquie
Continuing and Electronic Resources Librarian in M&C

Today we are going to talk to you about NC State’s creation of a name authority within its ERM system, a homegrown product called E-Matrix. We’ll discuss the motivations for creating a name authority tool and the process NCSU went through.
The Name Game

- Why build an ERM system with a name authority tool?
- Have other institutions explored ERM and name authority?
- How was NCSU’s authority created?
- What’s next?

Questions that our presentation will attempt to answer.

Kristen: Slides 1-11, 27-30
Jacquie: Slides 12-26
Why build a name authority?

Transition: The first and most obvious question to answer is: why build a name authority into an ERM system?

The answer is that much of the data about organizations found in E-Matrix is without authority control, because of where it comes from. Each day, E-Matrix imports data from our catalog and our link resolver. Name data comes from several areas within those sources, including the 260, SFX fields including publisher, vendor field in order record, human data entry. This data has not been subjected to any kind of authority control, so it is completely inconsistent and duplicative.

Advance slide:

In this screenshot, for example, you can see Elsevier repeated 80 times within E-Matrix.

In traditional library resources like a catalog, this inconsistency isn't much of a problem as org names are primarily being used by humans on an individual basis. In an ERM setting, however, where serials are being analyzed and processed on a large scale basis, inconsistency is not acceptable.
Why build a name authority?
Why build a name authority?

Transition: The need for data consistency can specifically be seen in a few key areas.

Roles: E-Matrix will use roles, which allow the same organization to be related to a title in a number of ways within the system. Roles currently include publisher, vendor, provider, licensor. The roles avoid duplication of data, but that processes becomes pointless if the data is duplicated anyway because of lack of control. We implemented roles before we implemented name authority – a situation that made the lack of authority very apparent.

Reports: In order to produce the kind of sophisticated reports that will take collection evaluation to the next level, E-Matrix must take advantage of roles that are consistent throughout the data. Example: Reports of money spent by publisher, vendor, etc. The name of that organization must be the same in all cases.

Licensing: Same as above – for license to link to other data, the name of the licensor must be the same name that is used on every other occasion.
Transition: Now that you’ve heard the basics on why name authority is important, I want to take a step back and put the idea of ERM name authority into a broader context. While name authorities are certainly a long-established library tool, the idea of using one as a component of an ERM system is a new one.

To find out if and how the idea of ERM name authority had been considered in other institutions, we conducted an informal survey of peer institutions last fall.

*Of 9 libraries surveyed, 8 owned ERMs. Four still considered themselves in beginning implementation stages and had not though about name authority. The other four reported consideration of name authority. Only one had actually implemented any kind of authority procedure.
Organization name control in the field

- Survey conclusions
  - Control of organization names a new concept
  - Importance of control varies by local needs
  - Tends to be an enterprise venture

Transition: From those responses, we drew a few broader conclusions.

New concept

Importance varies: Of the four libraries considering name authority, three libraries had considered authority control and rejected it – not important enough, cost-benefit analysis failed, built-in features could be a substitute. One library -- MIT -- had implemented name authority as an offshoot of a database dating back to the 80s.

Enterprise venture: Both NCSU and MIT have undertaken name authority work of their own accord using local systems. Vendor ERMs do not offer any real authority control. The decision to implement it really relies on strong local need and desire to use ERM for advanced reporting.
Transition: In addition to talking with librarians, we also took a look at some new products at OCLC that use of name authority in new contexts.

Advance slide

The WorldCat registry is a product that OCLC released last year that is designed to let institutions manage their own identities and store data about their institutions. Because the needs of our ERM also demand consistent and unambiguous data, we initially considered harvesting data from this registry. Unfortunately, the data wasn’t quite as unambiguous as our needs demanded. As you can see from the example, NC State has nine different identities.

Advance slide

We also spoke with two researchers at OCLC who are building a publisher authority server. This project is still in the very beginning phases. Like our ERM authority, the creators envision it as a collection intelligence tool. In fact, their project is quite similar to ours with one major difference: their focus is on books, while ours in on serials. So there may be some overlap, but not enough to justify an attempt at collaboration.

Transition: So without a clear precedent for the development of an ERM name authority, NC State libraries faced the challenge of designing the project from the ground up. I’m going to turn the presentation over to Jacquie, who will talk to you about the development of the E-Matrix name authority.
## Organization name control at OCLC

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Carolina A&amp;T State University, Landrum Library</td>
<td>GREENSBORO, North Carolina 27411</td>
</tr>
<tr>
<td>North Carolina State University</td>
<td>RALEIGH, North Carolina 27695-7111</td>
</tr>
<tr>
<td>North Carolina State University at Raleigh, D.H. Hill Library</td>
<td>RALEIGH, North Carolina 27605</td>
</tr>
<tr>
<td>North Carolina State University at Raleigh, Heron Library</td>
<td>RALEIGH, North Carolina 27605</td>
</tr>
<tr>
<td>North Carolina State University at Raleigh, J. B. Taylor Library</td>
<td>RALEIGH, North Carolina 27605</td>
</tr>
<tr>
<td>North Carolina State University at Raleigh, Natural Resources Library</td>
<td>RALEIGH, North Carolina 27605-7111</td>
</tr>
<tr>
<td>North Carolina State University, Veterinary</td>
<td>RALEIGH, North Carolina 27605</td>
</tr>
</tbody>
</table>

- **WorldCat Registry**
- **Publisher Authority Server**
Organization name control at OCLC

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>North Carolina A&amp;T State University, Ferdinand Douglass Black Library</td>
</tr>
<tr>
<td>2</td>
<td>North Carolina State University</td>
</tr>
<tr>
<td>3</td>
<td>NORTH CAROLINA STATE UNIVERSITY AT RALEIGH</td>
</tr>
<tr>
<td>4</td>
<td>North Carolina State University, Veterinary Medical Library</td>
</tr>
<tr>
<td>5</td>
<td>North Carolina State University, D.H. Hill Library</td>
</tr>
<tr>
<td>6</td>
<td>North Carolina State University, Harry R. Byrd Library</td>
</tr>
<tr>
<td>7</td>
<td>North Carolina State University, Estes T. Jones Library</td>
</tr>
<tr>
<td>8</td>
<td>North Carolina State University, Natural Resources Library</td>
</tr>
<tr>
<td>9</td>
<td>North Carolina State University, Veterinary</td>
</tr>
</tbody>
</table>

- WorldCat Registry
- Publisher Authority Server
How was NCSU’s authority created?

Transition: As Kristen said, the lack of precedent forced NC State to begin this project from the ground up.

During initial conceptualizations of e-matrix, the idea for a name authority was discussed for the reasons mentioned earlier, but no solution was found. The idea was put on the back burner as the development of the ERM began.

When new group of librarians were brought together to form an implementation team, we discussed all aspects of e-matrix holistically in terms of specific functionalities and use cases. Many of E-Matrix’s planned features relied on being able to access titles tied to a certain organization with a specific role. Without that ability, reports, licenses, etc. wouldn’t function well. In team discussions, I found myself thinking ‘we could do that if we had authority control’ again and again. After some investigation on what might be available in the field, I realized we would have to create something ourselves since there wasn’t an extant system to leverage.

So at that point, it was decided that building a name authority tool (or ONA tool as we’re calling it) was a top priority for E-Matrix to function as it was supposed to.

**In simple terms, this authority would be a tool that would allow us to choose and assign one authoritative heading for each and every organization name in E-Matrix.**

Unlike like traditional authority control tools, this tool isn’t used to alter names in their native interfaces. It only controls them within E-Matrix.
Designing project workflow

- Started with our biggest packages
- Worked on publishing partnerships to disambiguate

Transition: With so many names to assign, the project needed a plan that would allow us to work through them in the most effective way possible.

Advance slide: We began working with our biggest licensors (Elsevier, Blackwell, Springer, Wiley, Sage, Nature, ACS) (high dollar and restricted access). Because these companies are responsible for providing so many of our resources, this tactic allowed us to hit a large number of titles by assigning only a few names. It also allowed us to control organizations related to those titles that are most important to our collection managers and to the licensing process.

Advance slide: Next, we worked on societies that published in association with our big licensors to avoid confusion.
Designing project workflow

- Started with our biggest packages
- Worked on publishing partnerships to disambiguate
Designing project workflow

- Started with our biggest packages

- Worked on publishing partnerships to disambiguate
Designing project workflow

- Cleaned up entries beginning with punctuation and diacritics
- Now continue through the list alphabetically

Advance slide: Simultaneously, we worked on cleaning up names that started w/ a diacritic or punctuation – these were problematic for alphabetical lists. These no longer come in thanks to programming. The programmers have stripped out all names w/ punctuation at beginning or end.

Advance slide: Now that we’ve taken care of those priorities, we’re basically working through alphabetically and just trying to get as many done as possible. New organizations are coming in to E-Matrix continually. New orgs that match on any form previously evaluated automatically are linked to their authorized heading. Still, there are many that need to be reviewed – screenshot shows over 9,000.
Designing project workflow

- Cleaned up entries beginning with punctuation and diacritics
- Now continue through the list alphabetically
Designing project workflow

- Cleaned up entries beginning with punctuation and diacritics
- Now continue through the list alphabetically

| "Carol Davila" University Press |
| "Journalfranz" Arnulf Liebing DHG |
| "Tung-pei ta hēuēh hēuēh pao" pen chi pu |
| (社)鹤鸣学会 |

<table>
<thead>
<tr>
<th>Reports</th>
<th>View Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>Date</td>
</tr>
</tbody>
</table>

Unverified Organizations
5,333 entries found, displaying 176 to 200.

<table>
<thead>
<tr>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Association of Cost Engineers</td>
</tr>
<tr>
<td>American Association of Critical-Care Nurses</td>
</tr>
<tr>
<td>American Association of Law Libraries</td>
</tr>
</tbody>
</table>
Guidelines for decision-making

- Collocate organization names regardless of source or role
- Take into account collection managers’ expectations
- Use national standards when possible
- Accommodate local needs

Transition: In addition to the big picture, it was also important to determine a workflow that would allow for accurate and consistent choices of authoritative names.

We needed a workflow that would allow us to manually evaluate names and make consistent decisions about authoritative forms. The procedure would be based on local needs, as well as widely available resources and web sites – traditional library sources and some more general ones.

I developed a set of flexible guidelines for decision-making. The guidelines were based on the need to collocate org names regardless of source, to use names based on what our collection managers expect to find, to use national standards whenever possible, and to accommodate other local needs (no parentheses).
Choosing authoritative names

- LCNAF
- Collection Mgmt input
- Translate v. transliterate
- Local uses

Transition: to summarize we:

1. Use LCNAF when we can. Use AACR2 when we have to create new ones. Online tools: Acronym finder, company web sites, even Wikipedia on occasion.
2. Used collection management input to determine how to choose when multiple orgs are listed in one statement, order of institutions when a hierarchy exists, for business relationships CM prefers to use imprint for publisher over current owner of said imprint,
3. Use translations for non-roman scripts.
4. Local use cases. Use IEEE as an example. Stands for Institute of Electrical and Electronics Engineers. Did not follow LC practice in this case.
Transition: In addition to making decisions, we of course needed a way to store them in E-Matrix.

Once I was set to begin the name authority pilot project, I needed to come up with an idea of how E-Matrix would store and organize all the data the project would generate. I came up with the following optimal structure, which I provided to the library’s programmers as a model.

Records for data
Auth record linked by unique IDs to other names
These would also have records with similar info – facilitate searching
Many names linked to an authorized name
Title records would be in database – have publisher, vendor, licensor, etc.
Authorize name will inherit all the titles linked to other names in the hierarchy.
Search for any name and get the authorized name and anything that was related it to by specific roles

Stress: Diagram is of initial concept. What we have currently is basically just the lines without the actual records, however development is ongoing. In future phases, the ONA tool will be expanded to include all of these features.
Transition: The programmers were able to adapt our vision into a very workable model for the first phase of ONA development. Part of that model is a name authority tool within E-Matrix, which allows for the storage and management of authoritative names.

During this part of the development phase, Kristen came on board and was able to give input to improving our interface for the ONA tool.

The name authority tool is currently on its second iteration, which was released with E-Matrix 1.0 in December 2007. The tool allows us to access a list of all organizations within E-Matrix, see whether an authoritative name exists for each, and group names to assign a heading. It also has new capabilities that may sound routine, but are actually a big improvement for us. We can now browse and search, delete mistakes, and create cross-references to names not previously found in the data.

As mentioned before, the E-Matrix name authority still doesn’t make use of records to store more detailed information, so we have been using an MS Access database that has been around since the beginning. Here we note the source of the authoritative name, business relationships, prior and later names, and other necessary information. This data will be transferred to E-Matrix when that area of the ONA has been developed to more closely resemble the record structure model first proposed.
Designing tools

Transition: Just to give you a bit of a clearer picture of exactly how all these pieces come together, I’ll quickly talk through the process of assigning a Name.

1. First, we choose which name to evaluate. In the example above we’ve done a search for Yale Law.
2. Decide which of the names are actually the same entity using research sources when needed.
Assigning an authority in E-Matrix

Select the names using the authority tool.

Advance slide

Here we have a choice between using one of the selections as the authoritative name or entering a new name if none of the existing ones suit our needs. At this point we would also enter any additional details into our database.

Advance slide

Finally, the tool groups the names under the authoritative heading. We hit finish, and it’s done!

Now that I’ve discussed the background and process of Authority control in Ematrix, and a little bit about how this has become the backbone to facilitate user services like reports; Kristen will wrap-up with some discussion of the status of our project and some thoughts about the future.
Assigning an authority in E-Matrix

Summary
The following Organization(s) are currently selected:
1. Yale Law Journal Co (Remove)
2. Yale Human Rights & Development Law Journal (Remove)
3. Yale Law Journal Company, Inc (Remove)
4. Yale University, School of Law (Remove)
5. Yale Law School (Remove)

Organizations > Name Authority
Yale Law School will be assigned as
- Yale Law Journal Co
- Yale Human Rights & Development Law Journal
- Yale Law Journal Company, Inc
- Yale University, School of Law
- Yale Law School

Create Authoritative Organization
Authoritative Organization Name

Previous Step  Finish
Assigning an authority in E-Matrix

Summary

The following Organization(s) are currently selected:
1. Yale Law Journal Co (Remove)
2. Yale Human Rights & Development
3. Yale Law Journal Company, Inc
4. Yale University, School of Law (Remove)
5. Yale Law School (Remove)

Organizations > Name Authoritative Organization

Yale Law School will be assigned as:
- Yale Law Journal Co
- Yale Human Rights & Development
- Yale Law Journal Company, Inc
- Yale University, School of Law
- Yale Law School

Organizations > Name Authority > Select/Create Authoritative Organization

Select Authoritative Organization

Organizations: Yale Human Rights & Development

Create Authoritative Organization

Authoritative Organization Name

Previous Step Finish
Assigning an authority in E-Matrix

Summary

The following organization(s) are currently selected:
1. Yale Law Journal Co (Remove)
2. Yale Human Rights & Development
3. Yale Law Journal Company, Inc (Remove)
4. Yale University, School of Law (Remove)
5. Yale Law School (Remove)

Organizations > Name Authority > Select/Create Authoritative Organization

Select Authoritative Organization

Organizations > Name Authority > Confirm Organizations (Step 3 of 4)
Where are we now?

- Our results
  - Evaluated 1319 organization names in main project
  - Reduced to 532 authorized organization names
  - These 532 authorized names are related to 21,672 titles

Transition: By last summer, the procedures we just talked about were pretty firmly in place. Since then, what kind of results have we seen from our name authority?

Working with our biggest publishers, vendors, and providers first helped contribute to this last number.
Name authority at work

- Reports

Transition: Practically speaking, what does that do for us?

Reports – most are for collection management evaluation and serials review, when necessary. Can be quite complex. Also screen shot of report w/ authority control. Others are to help us keep up to date with data. New organization added report in development.
Name authority at work

- Reports in development
  - Top publisher by subject and amount spent
  - Organization name and all titles by role
  - Custom reports by request

- Licensing

- Roles

Reports in development – search by name and find all titles w/ any role. CM wants to know everything we get from Elsevier. Also gives us items that have business relationships. –find top publishers by subject.

Licenses – Describe how only authoritative names can be chosen when license is being mapped.

Roles – mention how authority feeds roles.
What’s next?

For our local authority?
- Improve structure
- Develop end-user displays

For authority control in the field?

What’s next?

Improve structure to match our original vision. – developing records, storing contact info, tracking business relationships, etc.

Encourage data sharing between intuitions – advocate for usefulness of name authority in commercial ERMs, Keep up-to-date with trends in the field such as NISO organization id group and of other big efforts to do the same kind of work.

Develop end-user displays and searches/reports where the use of Organization names depends on knowledge of the name authority project. Lacking this knowledge, the displays need to make relationships and decisions clear to users. Right now the display of authorized names is limited to the Organization tool within Ematrix and we understand that there will be issues with the way organizations are displayed throughout the tool as people start using it more heavily, and we’ll develop this area based on use cases and input from Libraries staff.
We’re going to stop here and open up the floor for questions.