This study describes a questionnaire that was distributed to managers and staff of institutional repositories (IRs) at ARL-member institutions in order to investigate name authority control in IRs and its relationships to metadata quality. Authority control is important to resource discovery because it provides collocation and search precision. However, a review of the literature indicates numerous challenges to establishing name authority control in IRs. Respondents were asked about their IR’s metadata creation and management, metadata quality, metadata schemas, and usage of authority data and controlled vocabularies. The results of this exploratory study indicate that half of respondents have established name authority control in their repositories. The majority of respondents believe that there is a relationship between name authority control and metadata quality in IRs.

Headings:

Authority control

Metadata -- Quality control

Surveys -- Metadata

Institutional repositories -- United States
NAME AUTHORITY CONTROL IN INSTITUTIONAL REPOSITORIES AND ITS RELATIONSHIPS TO METADATA QUALITY

by

Anna S. Goslen

A Master’s paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina
April 2011

Approved by

Deborah Barreau
# Table of Contents

- **Introduction** 3
- **Literature Review**
  - Institutional repositories - challenges to metadata quality 4
  - Metadata creation in institutional repositories 6
  - Value of name authority control and controlled vocabularies 7
  - Current state of name authority control in institutional repositories 11
  - Relationship of name authority control to metadata quality 12
- **Methodology**
  - General description 14
  - Participants 15
  - Data collection instrument 15
  - Data analysis 16
  - Limitations 16
- **Results**
  - Institutional repository holdings and software 18
  - General metadata practices 20
  - Name authority control 24
- **Discussion** 28
- **Conclusion** 32
- **References** 34
- **Appendix A: Survey** 37
- **Appendix B: Recruitment Emails** 41
Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Holdings of respondents’ institutional repositories</td>
<td>19</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Software platform used in IR</td>
<td>20</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Persons or systems responsible for metadata creation</td>
<td>21</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Metadata schema used</td>
<td>22</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Controlled vocabularies used</td>
<td>23</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Obstacles to establishing name authority control</td>
<td>25</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Importance of name authority control to metadata quality</td>
<td>26</td>
</tr>
</tbody>
</table>
Introduction

As institutional repositories (IRs) continue to grow and develop at universities across the country, common problems have arisen which research is needed to address. One of these problems is the issue of name authority control. Institutional repositories have varied levels of institutional commitment among universities, which can limit the amount of resources devoted to metadata quality (Lynch, 2003). In addition, software and services designed to manage institutional repositories were not designed with authority control in mind (Salo, 2009). Finally, metadata in institutional repositories is generated from a variety of sources, including content creators, metadata professionals, and information systems. This practice introduces inconsistencies in the way metadata for author names is created and further complicates attempts to make this metadata accurate and consistent (Chapman, Reynolds, & Shreeves, 2009). This study uses a questionnaire to gather data from managers and staff of institutional repositories at ARL-member institutions. The relationships between name authority control in IRs and metadata quality are investigated.
Literature Review

Institutional repositories - challenges to metadata quality

In his 2003 article for *ARL Bimonthly Report*, Lynch defines an institutional repository as a “set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members” (p. 2). Institutional repositories have faced a number of problems over the years; among these have been the failure of the “build it and they will come” model for populating an IR and lack of resources and institutional support for repositories (Salo, 2008). Consequently, it is not surprising that issues of name authority control in IRs have been neglected in favor of addressing more pressing challenges.

Several of the problems that institutional repositories have confronted in the recent past have the potential to negatively affect the quality of their metadata. These problems include a variety of deposit methods, a lack of resources, and little involvement from other librarians including catalogers. The cross-disciplinary nature of many institutional repositories can also present unique challenges to metadata quality.

Institutional repositories in the United States were initially founded on the assumption that faculty would eagerly fill them: a process referred to as self-archiving. However, thus far this process has not proven to be a successful method for populating institutional repositories. Barton, Currier, and Hey (2003) explain that because many repositories are focused on encouraging participation, anything that is perceived as a barrier between academics and their parent institutions tends to be downplayed; this
includes intervention and quality control measures in the metadata creation process. As a result of a lack of self-archiving, many institutional repositories now employ a variety of deposit methods to encourage deposit, including librarian-mediated deposit. Salo (2009) explains, “Librarians handle a considerable proportion of institutional repository deposits, either item-by-item or in large batches” (p. 253). In describing the mixed metadata environment of institutional repositories, Chapman et al. (2009) explain that metadata may be “mapped and converted from existing systems, elicited from the document creator or manager, or created by library or repository staff” (p. 310). This variety of deposit methods, including self-archiving, librarian-mediated deposit, and batch ingests, complicates the ability to control the associated metadata.

Institutional repositories have also confronted a lack of resources, institutional commitment, and limited staffing in recent years. As Salo (2008) explains, “I have seen several library administrators, the first planning push past, leave repository managers starved of resources, refusing to undertake any outreach or education work themselves” (p. 105). A common staffing model for institutional repositories is what Salo (2008) calls the “Maverick Manager” model. This manager is often somewhat isolated, has few resources, and has little involvement with other library staff. In addition, there may be little involvement from technical services departments in addressing the metadata in institutional repositories (Salo, 2009). In an environment where repository managers have few resources and are expected to do it all, metadata quality has the potential to suffer.

Many institutional repositories are cross-disciplinary, and this creates difficulties in providing both subject access and name access to materials. As Salo (2009) says, “The cross-disciplinary nature of institutional repositories significantly increases the danger of
name clashes turning up material irrelevant to a user’s information need during an author browse or search” (p. 251). This problem becomes even greater when one considers name searches in federated collections such as OAIster. Materials in cross-disciplinary institutional repositories cover an incredible variety of subjects, which can make choosing a controlled vocabulary for subject terms difficult. In explaining their experience with a cross-institutional repository, Costanza, Knight, and Liu-Spencer (2009) found difficulty in dealing with the scope of materials for providing subject access. They describe the example of a student paper about a local initiative, which necessitates both general and very specific subject terms. Thus, the cross-disciplinary nature of institutional repositories presents yet another challenge to the metadata quality picture.

**Metadata creation in institutional repositories**

In their article detailing the experience of the National Science Digital Library, Lagoze et al. (2006) write that successful provision of metadata involves three distinct skill sets: domain expertise, metadata expertise, and technical expertise. Metadata generated by content creators can have both a positive and negative effect on the quality of metadata in institutional repositories. Content creators often have the domain knowledge to most accurately describe their materials. For example, Chapman et al. (2009) describe the debate at Johns Hopkins University about who should create the metadata for objects in the repository, saying, “Members of the research community who are creating the materials that will populate JScholarship have the domain knowledge necessary to create the most accurate subject descriptors” (p. 316). Allowing authors and content creators to perform this activity has the potential to provide the most accurate
subject terms and improve information retrieval by future users. However, content creators may not have an understanding of the importance of controlled vocabularies, indexing, and cataloging concerns (Park, 2009). Currier, Barton, O’Beirne, and Ryan (2004) found that a collaborative approach to metadata creation involving both authors and metadata specialists may be necessary. Individual institutional repositories must weigh these options and determine the metadata creation process that works best for their situation.

Metadata created by machines and information systems also poses problems for metadata quality. Park (2009) describes the continued need for cataloging professionals and human indexers in the metadata creation process due to the inability of machines to understand the complexities of natural language. Currier et al. (2004) agree, saying, “Those tasks that require human intelligence and creativity can include such areas as subject classification, educational attributes and determining contributors to a resource” (p. 7). Other processes common to institutional repositories present metadata quality concerns. As Salo (2009) explains, “Batch ingests in particular pose metadata issues, especially when the imported material and metadata come from screen scraping a faculty-authored website possessing no particular metadata control” (p. 253). At present, machines are unable to perform all of the tasks necessary in metadata creation.

**Value of name authority control and controlled vocabularies**

It is valuable to begin a discussion of authority control with a review of the functions of the library catalog, as described both by Cutter (1876) over a century ago, and of the user tasks developed by the International Federation of Library Associations and Institutions (IFLA) Study Group on the Functional Requirements for Bibliographic
Records (FRBR) in 1998. Cutter says that the objectives of the library catalog are “to enable a person to find a book of which either the author, title, or subject is known, to show what the library has by a given author, on a given subject, in a given kind of literature, and to assist in the choice of a book as to its edition (bibliographically), as to its character (literary or topical)” (1876, p. 10). These objectives are often referred to as the finding objective, collocating objective, and the identifying objective respectively.

FRBR lists four user tasks: to find entities corresponding to search criteria using an attribute or relationship of the entity, to identify an entity (confirm that the entity described corresponds to the entity sought, or to distinguish between two entities), to select an entity meeting the user’s requirements, and to obtain the entity (acquire or access it) (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998).

In the age of keyword searching, some may question the importance of authority work and controlled vocabularies. Tillett (2004) explains the value of authority control in the Web environment. She says, “When we apply authority control in today’s Web environment, we are reminded how authority control brings precision to searches, how the syndetic structure of references enables navigation and provides the end user with explanations for variations and inconsistencies, and how the controlled forms of names and titles and subjects help collocate (group together) works in displays” (p. 24).

Essentially, authority control provides collocation and search precision. It also helps users navigate the catalog and explains inconsistencies. Gorman (2004) explains the role of authority control in the access point of a catalog record. He says, “Cataloguing cannot exist without standardized access points, and authority control is the mechanism by
which we achieve the necessary degree of standardization” (p. 12). All access points, whether for names or subjects, enable users to find the record and group together records that share that access point as a characteristic. Although users may be able to find the item for which they are searching by taking advantage of full-text searching capabilities, as Gorman says, “the system with authority control is clearly superior to the system without” (2004, p. 17).

Although clearly valuable, authority control is very expensive. Tillett (2004) explains that people have claimed that authority work is the most expensive part of cataloging since the 1970s, and are continuing to look for ways to automate and simplify the process. The Name Authority Cooperative project (NACO) has allowed catalogers to share the work and to use a shared authority file. Tillett envisions taking shared authority work further through the principles of Universal Bibliographic Control (UBC) described by IFLA and the use of virtual international authority file. Gorman (2004) explains that the ideal behind UBC is for each document to only be cataloged once in the country from which it originates and for the results of that cataloging to be available internationally. However, before that ideal can be achieved, he explains that we must have a universally accepted cataloging code and an international authority file. There is much work to be done on the development of an international authority file, but if created, it would further reduce the costs associated with authority work.

The metadata environment has presented several challenges which make authority control less successful than in the library environment. Vellucci (2000) describes four reasons for the success of authority control. First and most importantly, it operates within a “well-defined and bounded universe,” the library catalog, which provides access to a
selective, quality-controlled collection. Additional factors in the success of authority control include the fact that the creation of access points is based on principles and standardized practices, it is aided by reference to authoritative lists, and it is performed by highly trained individuals. These factors may or may not be present in the variety of metadata environments that exist today. In the digital library context, Borbinha (2004) explains that metadata is intended to deal with a greater heterogeneity of genres, be interoperable, and be able to be used and reused as technology advances. The web environment contains many documents that would benefit from bibliographic control. However, the number of documents that are marginal, only somewhat useful, or that are not useful at all suggests that we cannot and should not attempt to extend bibliographic control to all of these objects. Gorman (2004) believes that the solution to this problem lies in either giving up on metadata altogether as something that will be useful in very large databases or in investing in metadata schemes with the attributes of traditional bibliographic records. He believes there must also be a thorough examination of the nature of electronic documents and resources so that we can determine how to identify electronic documents of lasting worth and preserve them once they are cataloged. Barton et al. (2003) explain that with the growing focus on interoperability and federated searches, metadata that is sufficient for discovery on the local level may not be as effective in an aggregated system. They agree with Gorman’s assertions about the problems of metadata in large databases, saying, “the larger the dataset, the greater the likelihood that a problem will manifest itself” (2003, p. 3). If institutional repositories wish to grow and become interoperable with other repositories, metadata quality issues such as name authority control will need to be addressed.
Current state of name authority control in institutional repositories

Name authority control is an important aspect of metadata in institutional repositories. Use of the Library of Congress Name Authority File (LCNAF) is often not a viable option in institutional repositories because authors in IRs are frequently authors of journal articles and/or conference papers and not books and therefore most do not have an entry in the NAF (Chapman et al., 2009). The issues of name authority control in institutional repositories are due in part to highly disparate sources of metadata in institutional repositories and the inability of existing IR software to support this function (Salo, 2009). Currently, IR managers who wish to implement name authority control in their repositories must develop their own methods for doing so, which may involve locally-devised systems or software. As Chapman et al. (2009) explain, “It is essential that an IR be able to accurately identify the researchers who deposit their materials for preservation, access, and rights reasons” (p. 324).

In explaining the findings of their study of Higher Education Institutions (HEIs) in the United Kingdom with an IR, Birrell, Dunsire and Menzies (2010) say, “The inconsistent application of authority control (or indeed, a complete lack of authority control) in IRs in prevalent, leading to difficulties when trying to find all publications by a specific author or outputs from a particular department or research group by searching across OPACs and IRs” (p. 388). Lubas (2009) discovered similar issues when researching metadata practices for deposit of electronic theses and dissertations in institutional repositories. She found that many institutional repositories use drop-down menus for fields such as department names, faculty names, and degree names in order to provide greater consistency in author-supplied metadata. Guy, Powell, and Day (2004)
lend further support to the importance of consistent application of controlled vocabularies, writing that inconsistent application will negatively affect the functionality of the system, and using drop-down boxes or links to authority lists can support consistent use. There is significant room for improvement in the use and consistent application of controlled vocabularies in institutional repositories.

**Relationship of name authority control to metadata quality**

In her 2009 article, Salo explains, “name authority control is only one piece of the metadata-quality puzzle, but a piece worth examining” (p. 250). There have been several attempts to define metadata quality. One of the most comprehensive is the chapter by Bruce and Hillmann (2004) in *Metadata in Practice*, in which the authors list and define seven of the most commonly recognized characteristics of metadata quality: completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness, and accessibility. Metadata quality can also be examined from a functional perspective; in other words, to what degree does the metadata perform the core bibliographic functions of discovery, use, provenance, currency, authentication, and administration? Guy et al. (2004) state that high quality metadata supports the functional requirements of the system it is designed to support. Metadata quality is a growing problem not only in the institutional repository environment but in the digital resource environment in general. In the past, libraries have largely relied on bibliographic utilities such as OCLC for quality control and have thus been slow to develop their own batch-update and quality control procedures (Hillmann, 2008).

The lack of name authority control in institutional repositories negatively affects the overall quality of IR metadata. As Salo (2009) explains, “The naïve user of an
institutional repository will swiftly find that the absence of name authority control inhibits retrieval of items by a singer author” (p. 250). She found eight variants on a single name in one institutional repository. Because institutional repositories are often cross-disciplinary, users seeking to limit results by authors with the same name to the author relevant to their information need will be unable to do so (Salo, 2009). Dragon (2009) describes the relationships between authority control, metadata quality, and the usability of repositories. She says, “Authority control is a large part of what makes the difference between low- and high-quality metadata, and high-quality metadata improve the usability of digital repositories” (p. 194). Therefore, it is important to address the issues of name authority control and metadata quality in institutional repositories because they have a significant influence on the usability of these repositories.
Methodology

General description

In order to explore the relationships between metadata quality and name authority control in institutional repositories, a 16-question survey was used (see Appendix A). This survey was designed and administered using Qualtrics software, available through the Odum Institute at the University of North Carolina at Chapel Hill. A questionnaire was chosen as the most appropriate method for this research because it would allow data to be gathered from a large sample in order to form a picture about the institutional repository community at large and how it is addressing issues of name authority control and metadata quality. As Babbie (2010) explains, “Survey research is probably the best method available to the social researcher who is interested in collecting original data for describing a population too large to observe directly” (p. 254). Babbie also explains that surveys are effective ways of measuring attitudes in large populations. It was anticipated that statements about the trends and attitudes in this community toward name authority control in IRs would be able to be made and a basis for further research in this area would be established. An electronic questionnaire was used because it allowed targeting of potential respondents through the use of relevant online discussion lists. Electronic surveys were also believed to be an easy and convenient method for individuals wishing to participate. Survey questions were adapted from past survey research on institutional repositories (Bailey et al., 2006), general metadata practices (Ma, 2007), and metadata creation practices in digital repositories (Park & Tosaka, 2010).
Participants

This study used non-probability convenience sampling of managers or staff of institutional repositories at institutions belonging to the Association of Research Libraries (ARL). ARL is a nonprofit organization whose membership consists of 126 research libraries in North America. The chosen population of managers or staff of institutional repositories at ARL-member institutions provided a well-defined population from which to sample.

Participants were recruited through emails sent to three online discussion lists: REPOMAN-L, Archives & Archivists, and Digital Curation. The REPOMAN-L listserv is a discussion list hosted by Indiana University for managers of institutional repositories. The Archives & Archivists list is hosted by the Society of American Archivists (SAA) for discussion of all issues related to archives. The Digital Curation list is a Google Group used to discuss issues primarily relating to digital repositories and digital preservation. The recruitment emails (see Appendix B) explained that the study sought participants from ARL-member institutions who were either managers or staff of institutional repositories, that participation was anonymous and voluntary, and that the survey would take approximately 25 minutes to complete. Reminder emails were sent one week after the initial recruitment email and one day before the survey closed. No incentive was provided for participating in the survey beyond the professional’s desire to contribute to research in his/her field of practice.

Data collection instrument

The electronic questionnaire used in this study consisted of 16 questions inquiring about the holdings, software, metadata creation and management, metadata quality,
metadata schemas, and usage of authority data and controlled vocabularies in respondents’ IRs. The survey contained both closed- and open-ended questions. Participants were able to skip questions which they did not wish to answer. Many questions provided the option of answering “unsure” or choosing “other” and providing more information in a free-text field. The last question on the survey was open-ended and asked participants if there was anything on this topic not addressed in the survey that they would like to add.

Data analysis

Descriptive statistical analysis was used to quantify responses regarding current practices and opinions about name authority control in institutional repositories. Babbie (2010) defines descriptive statistics as “statistical computations describing either the characteristics of a sample or the relationship among variables in a sample” (p. 467). Qualtrics and Microsoft Excel were used to convert qualitative data into tables and figures for reporting results in a graphical format. Both Qualtrics and Excel were used to analyze closed- and open-end questions.

Limitations

Due to the fact that this study used a questionnaire in an attempt to be able to make statements and observations about a large population, questions that were believed to be answerable and applicable to all members of the population were used. This practice resulted in the inability to capture nuances about situations at individual repositories, which could have been discovered by methods such as interviews or case studies. Using a majority of closed- rather than open-ended questions also contributed to this result. Self-selection by participants introduced the potential for sampling bias. The
small response rate and sampling method mean that findings cannot be generalized to the entire population. Data analysis was completed solely by the Principal Investigator.
Results

There were 52 responses to the survey; however, a number of surveys were incomplete. Qualtrics indicated that 30 surveys were completed. Incomplete surveys were included in data analysis because they included some useful information. All questions were optional. The results presented here are based on the response rates to individual questions.

Institutional repository holdings and software

Respondents were first asked to describe the holdings of their institutional repository. There were 31 responses to this question, and respondents were able to choose all applicable answers. Responses are presented in Figure 1. The greatest percentage of repositories hold conference publications (71%), followed by multimedia files (68%), digitized archival materials and university records (65%), electronic theses and dissertations (65%), and working papers and technical reports (65%).
Figure 1. Holdings of respondents’ institutional repositories.

The next question addressed the type of software platform used in respondents’ IRs. There were 28 responses to this question. Data from this question is shown in Figure 2. The greatest number of respondents selected DSpace (32%). The responses indicate that most repositories use only one software platform.
<table>
<thead>
<tr>
<th>Software Platform</th>
<th>Numbers of Responses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTENTdm</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>Custom-made IR</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>Digital Commons</td>
<td>7</td>
<td>25%</td>
</tr>
<tr>
<td>DSpace</td>
<td>9</td>
<td>32%</td>
</tr>
<tr>
<td>EPrints</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>ExLibris</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Fedora</td>
<td>3</td>
<td>11%</td>
</tr>
<tr>
<td>Other (please specify)</td>
<td>5</td>
<td>18%</td>
</tr>
</tbody>
</table>

*Figure 2.* Software platform used in IR. Responses to the “Other” option included Past Perfect, Archivist’s Toolkit, Equella, and LUNA and Luna Insight.

**General metadata practices**

The third question on the survey asked respondents how or by whom the metadata for objects in their repositories is created. There were 28 responses to this question, presented in Figure 3. The majority of respondents chose IR staff (79%) and/or content creators (68%). There were several responses to the “Other” option of this question, including academic department administrative assistants and graduate assistants, librarians in departmental libraries, visual resource librarians, curators, student assistants, and original catalogers. These results indicate that, though the majority of metadata is created by content creators and/or IR staff, the range of possible metadata creators is large. Twenty-nine percent of respondents indicated that technical services staff are responsible for metadata creation, and one free-text response cited the involvement of a professional cataloger. Therefore, there is some level of involvement of catalogers and technical services staff in metadata creation in IRs, which has the potential to improve the quality of the metadata in these repositories.
<table>
<thead>
<tr>
<th>Responsible for Metadata Creation</th>
<th>Number of Responses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content creators</td>
<td>19</td>
<td>68%</td>
</tr>
<tr>
<td>IR staff</td>
<td>22</td>
<td>79%</td>
</tr>
<tr>
<td>Information systems</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>Technical Services staff</td>
<td>8</td>
<td>29%</td>
</tr>
<tr>
<td>Other (please specify)</td>
<td>6</td>
<td>21%</td>
</tr>
</tbody>
</table>

*Figure 3.* Persons or systems responsible for metadata creation.

Respondents were then asked to choose the metadata schema used in their repositories. The responses to this question (see Figure 4) indicate that the majority of repositories (70%) use Dublin Core. Twenty-seven respondents answered this question. Respondents were allowed to select multiple answers, and the responses show that some IRs use more than one schema or a combination of schemas.
<table>
<thead>
<tr>
<th>Metadata Schema</th>
<th>Number of Responses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dublin Core</td>
<td>19</td>
<td>70%</td>
</tr>
<tr>
<td>EAD (Encoded Archival Description)</td>
<td>5</td>
<td>19%</td>
</tr>
<tr>
<td>Locally-defined custom schema</td>
<td>6</td>
<td>22%</td>
</tr>
<tr>
<td>MARC (Machine Readable Cataloging)</td>
<td>5</td>
<td>19%</td>
</tr>
<tr>
<td>METS (Metadata Encoding and Transmission Standard)</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>MODS (Metadata Object Description Schema)</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>PREMIS</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>TEI (Text Encoding Initiative)</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>Other (please specify)</td>
<td>4</td>
<td>15%</td>
</tr>
</tbody>
</table>

*Figure 4.* Metadata schema used. Responses to the “Other” option included ETD-MS, Qualified Dublin Core with elements from ETD-MS, and VRA Core.

The next question in this section addressed the use of controlled vocabularies in respondents’ IRs. Data from this question is displayed in Figure 5. There were 21 responses to this question. The responses show that the majority of IRs use Library of Congress Subject Headings (LCSH) (76%) and the Library of Congress Name Authority File (LCNAF) (52%). Nine respondents (43%) indicated using the Library of Congress Thesaurus of Graphic Materials (TGM). The Art and Architecture Thesaurus (AAT), Getty Thesaurus of Geographic Names (TGN), and the Getty Union List of Artist Names (ULAN) are each used by 29% of repositories. These responses indicate that the majority of repositories use one or more controlled vocabularies. The results also indicate that some repositories do not use any controlled vocabularies.
Figure 5. Controlled vocabularies used. Responses to the “Other” option included Bepress discipline taxonomy, Digital Commons list of academic disciplines, NAL thesaurus, and local controlled vocabularies.

Respondents were also asked if their repositories had any involvement from cataloging or technical services staff to address metadata quality. There were 28 responses to this question. Cataloging or technical services staff are involved at 18 institutions (64%). Eight institutions (29%) cited no involvement from these individuals, and two respondents (7%) were unsure about whether these individuals have involvement in addressing metadata quality in their IR.

In addition, the survey asked respondents whether the metadata in their IR is exposed to Open Archives Initiative (OAI) harvesters. Twenty-eight respondents answered this question. Twenty-one repositories (75%) expose their metadata and two (7%) repositories do not. Five respondents (18%) were unsure about whether their metadata is exposed to OAI harvesters. OAI requires Dublin Core metadata for
harvesting. This requirement may be a reason for the number of repositories using Dublin Core, as indicated in Figure 4.

**Name authority control**

The next section of the survey addressed name authority control in the respondents’ IRs. Respondents were first asked whether they had any method for establishing name authority control in their repositories. Twenty-eight respondents answered this question. Fourteen respondents (50%) answered yes, and 14 respondents answered no (50%). Thus, exactly half of the IRs represented by the respondents have some method of establishing name authority control. The respondents who answered yes to this question were then asked if they used an external source of name authority information. Six respondents answered yes to this question, with the majority listing the LCNAF. Eight respondents answered no. These same respondents were also asked if they believe that name authority control results in higher quality metadata in their repositories. Nine respondents answered yes and five respondents were unsure.

Respondents who indicated that they have no method of establishing name authority control in their IR were asked what they considered to be the primary obstacle(s) to establishing name authority control in their repository. The greatest number of respondents who answered “no” (nine, or 69%) chose lack of sufficient resources, with most respondents choosing at least two answers. The responses to this question are displayed in Figure 6.
These respondents were then asked whether they believed that implementing name authority control in their IR would result in higher quality metadata. There were 13 responses to this question. Eleven respondents (85%) answered yes, one respondent answered no, and one respondent was unsure. Therefore, the majority of respondents with no method of establishing authority control in their IR believe that, if implemented, name authority control would result in higher quality metadata.

Next, all respondents were asked how important they feel name authority control is to the overall quality of the metadata in their IR. The results of this question are displayed in Figure 7. The most frequently chosen response was somewhat important, selected by 12 respondents (44%), followed by very important, chosen by 10 respondents (37%).
<table>
<thead>
<tr>
<th>Degree of Importance</th>
<th>Number of Responses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very important</td>
<td>10</td>
<td>37%</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>12</td>
<td>44%</td>
</tr>
<tr>
<td>Neutral</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>Less important</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>Not important</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>Total Responses</td>
<td>27</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Figure 7. Importance of name authority control to metadata quality.*

The next question asked respondents whether they believe there is a relationship between name authority control and metadata quality in IRs. Twenty-two respondents (81%) selected yes, three respondents were unsure, and two respondents selected no.

Question 15 addressed the potential for improved software and services for addressing name authority control in IRs. Respondents were asked if they would like to see standards organizations and software developers address the issue of name authority control in IRs through improved services. Twenty-one respondents (78%) chose yes, four respondents were unsure, and two respondents chose no.

The final question of the survey asked if there was anything on the topic that respondents would like to add which was not addressed. One respondent mentioned that name authority control would help us develop better tools for faculty, so that they could more easily collocate their works. As mentioned in the literature review, the difficulty of using the LCNAF due to the fact that many authors in IRs are students and/or have never published a monograph and will not have an established authority record was mentioned by one respondent. This respondent said that he/she would rather see increased activity around projects like the Open Researcher Contributor ID Initiative (ORCID).
ORCID was also mentioned by another respondent who said that if it takes off, it will significantly help with authority control. One respondent believes that name authority control in IRs should be handled differently from name authority control in Online Public Access Catalogs (OPACs), saying that there should be room for folksonomies, user tagging, uncontrolled keywords, and key phrases as well as controlled vocabularies. The issues of whether there are too many standards, what the best way is to associate digital surrogates with originals, and whether metadata in a global context implies multiple languages were raised by one respondent. The point was made that at some smaller institutions, cataloging is not done in a separate department, and the archivist is the cataloger. Finally, another respondent believes that it is unclear whether name authority control results in better resource discovery, which is its ultimate goal.
Discussion

The results of this study indicate that half of these respondents have a method for establishing name authority control in their IR, and the majority of respondents believe there is a relationship between name authority control and metadata quality in IRs. The responses to the sections of the survey addressing repository holdings, software, and metadata practices are informative and help form a picture about how name authority control and metadata quality may be affected by general repository practices.

Repository holdings are varied, with most repositories holding several types of digital objects. This finding most likely reflects the cross-disciplinary nature of institutional repositories. This cross-disciplinary nature is also reflected in the usage of controlled vocabularies. Of the repositories that are using controlled vocabularies, most are using more than one. Thus, due to the scope of materials and range of subjects covered by objects in IRs, repositories use several controlled vocabularies.

The software platform used most commonly by respondents is DSpace. DSpace (www.dspace.org) is a free, open-source software platform for repositories developed by HP and MIT. The majority of repositories use Dublin Core as the metadata schema for describing their digital objects. Dublin Core is commonly believed to be a “low-barrier” standard that is simple and inexpensive to use (Lagoze et al, 2006). The usage of a free, open-source software platform such as DSpace and a simple metadata standard such as Dublin Core reflect the limited resources available to institutional repositories as described in the literature. However, both DSpace and Dublin Core have characteristics
with the potential to negatively affect the ability to establish authority control. Salo (2009) writes that DSpace does not link author names by default, does not provide an autocompletion feature, and does not provide batch metadata editing facilities. Gorman (2004) describes Dublin Core as limited in the number and nature of categories and lacking the concepts of controlled vocabularies and authority work. Therefore, the common use of DSpace and Dublin Core in the repositories surveyed may affect the ability of these repositories to establish name authority control.

The variety of methods for creating metadata in institutional repositories, as described in the literature, are reflected in the responses to the corresponding question on the survey. The results indicate that most of the respondents are relying on both IR staff and content creators for metadata creation. Surprisingly, only one respondent indicated that information systems have a role in creating metadata. A number of respondents indicated that technical services staff create metadata for objects in their repository, suggesting the technical services staff and/or catalogers may be more involved in this process than was suggested by the literature. Additionally, this role may be filled by IR staff themselves at smaller institutions without specialized technical services or cataloging departments.

In addition to having some responsibility in metadata creation, respondents indicated that technical services staff are involved in addressing metadata quality in their IRs. This finding is promising for the improvement of metadata quality in IRs in the future. It is possible that institutional repositories and their managers are becoming less isolated. As repositories continue to grow, the metadata quality in these repositories is
being addressed, which will improve their usability and interoperability. A collaborative approach to metadata creation and quality evaluation may be the best solution.

The issue of interoperability was also addressed by the question asking respondents whether the metadata in their IR is exposed to OAI harvesters. The majority of respondents are exposing their metadata, which may be related to the common use of the Dublin Core metadata schema, required for OAI harvesting. Any further attempts to improve metadata quality in respondents’ repositories will improve resource discovery and interoperability in large, federated collections.

The results of this study indicate that half of respondents have a method for establishing name authority control in their IR. Despite the challenges to name authority control in the institutional repository environment, many managers and staff of IRs feel it to be important and have chosen to implement it in some manner. About half of the respondents are using the Library of Congress Name Authority File in some capacity, and several respondents cited using it as their external source of name authority information for establishing name authority control in their repository. Many repositories are using local controlled vocabularies, such as a local list of department names or faculty names, for name authority information. Although most of the respondents who have implemented name authority control feel that it results in higher quality metadata in their repositories, some respondents are unsure.

The respondents whose repositories have not implemented name authority control cited a number of obstacles as reasons for this choice. Most respondents chose more than one answer to this question. The most commonly chosen obstacle was a lack of sufficient resources. This finding is also in agreement with the limited resources of IRs as described
in the literature. Despite the lack of a method for establishing name authority control in their repositories, the majority of these respondents feel that, if implemented, name authority control would result in higher quality metadata.

Of all the respondents surveyed, most feel that name authority control is either very important or somewhat important to the overall quality of the metadata in their IR. The majority of all respondents believe that there is a relationship between name authority control and metadata quality in IRs. Responses indicate that managers and staff of IRs would like to see standards organizations and software developers address the issue of name authority control through improved services.

Qualitative responses to the last question on the survey raised important issues, such as the effect of name authority control on usability and resource discovery, the development of systems using open contributor identifiers, and the role of metadata in a global context. One respondent feels that name authority control would help repositories develop better tools for faculty, who would be able to more easily locate and collocate their works in a repository. Another respondent believes that it is still unclear whether name authority control results in better discoverability. Respondents are optimistic about the ORCID initiative (http://www.orcid.org/) for helping with name authority control issues. This initiative seeks to establish an open registry of unique identifiers for researchers and a mechanism to link ORCID to other author ID schemes. The role of metadata in a global context may be clarified somewhat if work continues on the development of an international authority file.
**Conclusion**

Name authority control has been described in the literature as a significant problem for the metadata in institutional repositories, affecting usability, resource discovery, and overall metadata quality. Thus far, there has been little research on how IRs are dealing with this problem. While only exploratory in nature, this study provides an idea of how some institutional repositories are addressing the issue.

A review of the literature provided background information on institutional repositories and their metadata practices. A variety of factors were identified which had the potential to affect metadata quality and the ability of repositories to establish authority control. The value of authority control and controlled vocabularies was discussed, and current practices regarding name authority control in repositories were summarized.

The results of this study indicate that respondents believe there is a relationship between name authority control and metadata quality in IRs. Half of the respondents have developed a method for establishing name authority control in their repositories.

Institutional repositories continue to face limited resources, but there is some involvement of technical services staff and/or catalogers in metadata creation and quality evaluation. The majority of repositories surveyed use one or more controlled vocabularies. Despite the challenges, the results seem to suggest that IRs are addressing the issues of name authority control and metadata quality.
As repositories continue to grow, it is likely that this trend will persist. As mentioned in the literature, metadata quality issues such as a lack of name authority control become more prominent as the volume of digital objects in repositories increases. In addition, as repositories seek to become interoperable and expose their metadata for inclusion in federated collections, it is important that the issue of name authority control be addressed.

Further research may be beneficial in examining different methods for establishing name authority control. This study surveyed a large group of IR managers and staff, and individual methods for establishing name authority control were not investigated. Qualitative responses indicated the belief that open identifier systems will improve authority control issues.
References


Appendix A: Survey

1. Please describe the holdings of your institutional repository
   Check all that apply
   - Conference publications
   - Course content
   - Data sets
   - Digitized archival materials and university records
   - Electronic theses and dissertations
   - E-Resources
   - E-Prints
   - Images
   - Learning objects
   - Multimedia files (audio/video)
   - Non-scholarly institutional publications
   - Undergraduate student work
   - Graduate student work
   - Working papers and technical reports

2. What software platform does your repository use?
   Check all that apply
   - CONTENTdm
   - Custom-made IR
   - Digital Commons
   - DSpace
   - EPrints
   - ExLibris
   - Fedora
   - Other (please specify)

3. How or by whom is the metadata for objects in your repository created?
   Check all that apply
   - Content creators
   - IR staff
   - Information systems
   - Technical Services staff
   - Other (please specify)

4. What metadata schema does your repository use?
   Check all that apply
   - Dublin Core
o EAD (Encoded Archival Description)
o Locally-defined custom schema
o MARC (Machine Readable Cataloging)
o METS (Metadata Encoding and Transmission Standard)
o MODS (Metadata Object Description Schema)
o PREMIS
o TEI (Text Encoding Initiative)
o Other (please specify)

5. What controlled vocabularies does your repository use?
Check all that apply

o Art and Architecture Thesaurus (AAT)
o Getty Thesaurus of Geographic Names (TGN)
o Getty Union List of Artist Names (ULAN)
o Library of Congress Subject Headings (LCSH)
o Library of Congress Name Authority File (LCNAF)
o Thesaurus of Graphic Materials (TGM)
o Other (please specify)

6. Do you have any involvement from technical services or cataloging staff to address metadata quality in your repository?

o Yes
o No
o Unsure

7. Does your IR expose your metadata to OAI (Open Archives Initiative) harvesters?

o Yes
o No
o Unsure

8. Do you currently have any method for establishing name authority control in your IR?

o Yes
o No

9. Do you use an external source of name authority information?

o Yes (Please list all that apply)
o No
10. Do you believe that name authority control results in higher quality metadata in your IR?
   o Yes
   o No
   o Unsure

11. What do you consider to be the primary obstacle to establishing name authority control in your IR?
    Check all that apply
    o Amateur metadata creators
    o Software
    o Sources of name authority information
    o Lack of sufficient resources
    o Do not wish to establish name authority control
    o Other (please specify)

12. Do you believe implementing name authority control in your IR would result in higher quality metadata?
    o Yes
    o No
    o Unsure

13. How important do you consider name authority control to be to the overall quality of the metadata in your IR?
    o Very important
    o Somewhat important
    o Neutral
    o Less important
    o Not important

14. Do you believe there is a relationship between name authority control and metadata quality in institutional repositories?
    o Yes
    o No
    o Unsure

15. Would you like to see standards organizations and software developers address the issue of name authority control in IRs through improved services?
    o Yes
    o No
16. Is there anything I did not ask about this topic that you would like to add?

- Unsure
Appendix B: Recruitment Emails

Initial Recruitment Email

***Please excuse cross-posting***

<Date>

Dear Colleagues,

We would like to invite you to participate in an online survey entitled “Name Authority Control in Institutional Repositories and Its Relationships to Metadata Quality.” The purpose of this study is to explore current practices and attitudes about name authority control and its relationships to metadata quality at institutional repositories (IRs) in the United States.

This study seeks participants from ARL-member institutions who are either managers or staff of institutional repositories. We are interested in surveying participants who can describe their IR’s metadata creation and management, metadata quality, metadata schemas, and usage of authority data and controlled vocabularies. If you decide to be in this study, you will be one of 126 institutions invited to participate in this research.

Participation in the survey is anonymous and voluntary. If you have any questions about the survey, please contact the Principal Investigator of this study at goslen@email.unc.edu. The survey will take approximately 25 minutes to complete.

The survey, at <URL> is now open, and will remain open until <date>.

We will be sending a reminder email about the survey in one week and another the day before the survey closes. Thank you for your assistance in providing information on this topic.

Sincerely,
Anna Goslen

Reminder Email at One Week

<Date>

Dear Colleagues,

This is a reminder that the survey entitled "Name Authority Control in Institutional Repositories and Its Relationships to Metadata Quality" will close in one week. The purpose of this study is to explore current practices and attitudes about name authority control and its relationships to metadata quality at institutional repositories (IRs) in the United States.
This study seeks participants from ARL-member institutions who are either managers or staff of institutional repositories. We are interested in surveying participants who can describe their IR’s metadata creation and management, metadata quality, metadata schemas, and usage of authority data and controlled vocabularies. If you decide to be in this study, you will be one of 126 institutions invited to participate in this research.

Participation in the survey is anonymous and voluntary. If you have any questions about the survey, please contact the Principal Investigator of this study at goslen@email.unc.edu. The survey will take approximately 25 minutes to complete.

The survey, at <URL>, is now open, and will remain open until <date>. We will be sending a final reminder email the day before the survey closes. Thank you for your assistance in providing information on this topic.

Sincerely,
Anna Goslen

Reminder Email Last Day

<Date>

Dear Colleagues,

This is a reminder that the survey entitled "Name Authority Control in Institutional Repositories and Its Relationships to Metadata Quality" will be closing in one day. The purpose of this study is to explore current practices and attitudes about name authority control and its relationships to metadata quality at institutional repositories (IRs) in the United States.

This study seeks participants from ARL-member institutions who are either managers or staff of institutional repositories. We are interested in surveying participants who can describe their IR’s metadata creation and management, metadata quality, metadata schemas, and usage of authority data and controlled vocabularies. If you decide to be in this study, you will be one of 126 institutions invited to participate in this research.

Participation in the survey is anonymous and voluntary. If you have any questions about the survey, please contact the Principal Investigator of this study at goslen@email.unc.edu. The survey will take approximately 25 minutes to complete.

The survey, at <URL>, is now open. The survey will close at 5pm EST on <date>.

Thank you for your assistance in providing information on this topic.
Sincerely,
Anna Goslen