

Introducing Name Authority into an ETD Collection

Abstract

Purpose of this paper

The purpose of this study is to develop and evaluate a workflow for establishing name authority in uncontrolled collections.

Design/methodology/approach

We developed a workflow incorporating command-line tools and tested it in our electronic theses and dissertations (ETDs) collection. We narrowed the scope of the study to born-digital ETDs in the collection and to contributor names, including chairs and committee members.

Findings

This workflow can save staff time and allows for flexible implementation depending on staff numbers and skills as well as institutional needs.

Originality/value

This workflow could be used by other institutions with little or no modification, as it does not rely on specialized software or extensive expertise.

Keywords

ETDs, electronic theses and dissertations, institutional repositories, name authority, metadata quality, authority control, digital library collections

Article classification

Case Study

1. Introduction

Establishing name authority control represents a key component for usability in digital collections. As the number of digital objects increases, there is a corresponding likelihood of greater inconsistency and problematic search results. Inconsistencies can stem from the changing of a person's name, pseudonyms, entry errors, differences in formatting, similarities in names, or even multiple people with the same name. All of these prove problematic when establishing a name authority control mechanism.

Although a number of organizations provide bodies of name authority records online, these authority files have limited scopes depending on the goals of the organizations. This presents a particular problem for universities that collect and curate works created by members of their communities, many of whom do not have representation within the established authority records. Local bibliographic catalog records, the Library of Congress Authorities (LOC), the Virtual International Authority File (VIAF), and even Wikipedia offer various authorized versions of names; however, many university community members do not have authority records. Other local sources may not prove sufficient, as campus-wide directories and faculty profiles generally rely on an author-submitted version of a name.

One large category of university-produced materials consists of electronic theses and dissertations (ETDs). The University of North Texas (UNT) was among the first in the United States to require electronic submissions of theses and dissertations, beginning in 1999. A partnership between the UNT Toulouse Graduate School and the UNT Libraries led to the creation of the ETD collection, housed within the UNT Digital Library. Because of the heavy use and impact of this collection, UNT also began a retroactive digitization of theses and dissertations from the 1930s onward. Both of these aspects have contributed to the large number of records with uncontrolled names already in the system.

In a random sampling of 200 names in the UNT ETD collection, 61% are in the local bibliographic authority file, 28% have an authority file in the VIAF database, 26% have an authority file in the LOC Name Authority File, and only .5% have a Wikipedia page (see Table 1).

Table 1: Frequency of UNT faculty names in various authority files

Name Authority System	Number of Matches	% of Test Sample
Local Bibliographic Authority File	122	61%
Virtual International Authority File (VIAF)	56	28%
Library of Congress (LOC) Authority File	53	27%
UNT Name App	43	22%
Wikipedia	1	2%

Aside from the lack of established authorities, name variation is a particular concern within the ETD collection. The current, accepted practice involves copying the names of authors, committee members, and chairs as they appear on the signature page of the thesis or dissertation rather than checking for authorized name authority records. Although this practice has arisen primarily from the difficulty of establishing name authority in these collections, it creates authority problems because authors sometimes use different version of their advisors' and committee members' names based on their personal history. Thus these differences are transferred into the collection as new ETDs are added.

In the spring of 2013, the UNT Libraries Digital Projects Unit introduced the UNT Name App, a tool built in-house for organizing and creating consistent names within the digital collections (Tarver et al. 2013). The UNT Name App allows staff to create authorized records for names of people, organizations, events, and software. Initially, to address the concern regarding UNT community names not controlled elsewhere, metadata editors started adding names of faculty and staff represented in the UNT Scholarly Works collection, the UNT institutional repository. As noted in Table 1, only 21% of the sample names in the UNT ETD collection have a corresponding name authority record in the UNT Name App. The workflow proposed in this paper intends to address this issue by increasing the level of name authority within the collection.

2. Related Research

Name authority has always challenged library catalogers, though digital collections present new issues. The process of creating authorized name records began with the card catalog and later transitioned to electronic files in MARC records (Xia 2006). In doing so, name authority moved from locally-controlled mechanisms in individual libraries to shared authority records that can be utilized nationally and internationally across libraries and organizations (Niu 2013).

In particular, the move from print to electronic resources has compounded the issue of name authority control in theses and dissertations (TDs). Print versions of TDs were traditionally considered first-order scholarly output and were given little to no attention for cataloging purposes (Wolverton et al. 2008). The development of ETDs has increased the dissemination and use of these works. Many libraries are now realizing that these ETDs merit greater cataloging attention and name authority control mechanisms, as their worldwide use continues to increase (Wolverton et al. 2008).

As a collection grows, there is a correspondingly greater likelihood for inconsistencies, similarities, or same names. Name authority is increasingly important in large digital collections, such as ETDs: "It is a well-known fact that personal name variants can reduce the precision level of online searches for digital materials" (Xia 2006 pp. 1). Name authority control structures are important not only for maintaining consistency and providing successful search and retrieval, but also for linking objects

and researchers, managing copyright and licensing, and allowing citation analysis (Xia 2006 pp. 1).

Name authority control mechanisms and cataloging procedures for ETDs vary widely from institution to institution (Wolverton et al. 2008). The benefits of providing name authority control mechanisms in ETD collections are gaining more attention. Additionally, the necessity to include unique identifiers to disambiguate similarities and same names of authors and to provide ways of linking these records is an area that merits more attention (Niu 2013). In 2006, Tim Berners-Lee introduced the five stars of open linked data, which can serve as a framework for more modern name authority control models (Berners-Lee 2006). The UNT Libraries developed the UNT Name App, a tool for creating consistency, disambiguation, and reference for names in the UNT digital collections (Tarver et al. 2013). It serves as a step in the right direction, but does not solve all of the challenges associated with name authority control, including the process of exerting authority control on uncontrolled digital collections, discussed in this pilot project.

3. UNT Name App

To address name authority in digital collections, the UNT Libraries developed the UNT Name App, which is a Django Web application. Each authority record represents a single person, organization, event, or software, and contains several fields that administrators can fill in to add pertinent details (see Table 2).

Table 2: Editable fields used in the UNT Name App

Field	Qualifiers	Description
Name		Authorized form of a name for a single entity.
Name type	Personal Organization Event Software	Controlled list for identifying the kind of entity to which the name belongs.
Biography		Free-text box for information associated with the name such as important dates, affiliations for persons, locations for organizations, or fields of study. Allows for the use of semantic markup of biographical information using microdata.
Begin		Starting date, e.g., birth date for persons, establishment date for organizations, etc.
End		Ending date, e.g., death date for persons,

		defunct date for organizations, etc.
Disambiguation		String used to distinguish between similar names such as acronyms or dates.
Variants	Acronym Abbreviation Translation Expansion Other	Alternate forms of the name.
Identifiers	Academia Facebook Google Scholar Homepage Linked In LOC ORCID ResearchGate Scopus Author ID Tumblr Twitter UNT Faculty Profile VIAF Wikipedia	Unique identifiers and URLs that represent the same person or entity in another system.
Notes	Biographical/Historical Deletion Information Nonpublic Source Other	Citation, administrative, and other information.
Record status	Active Deleted Suppressed Merged with	Current status of the record that can be changed to suppress names or subsume a duplicate record under another authority record.

The Biography and Note fields allow for free-text information with semantic markup, while other fields parse specific kinds of information (e.g., dates, links, etc.). The Identifier field is particularly important, as it creates links between authority records in the UNT Name App and other systems. Tim Berners-Lee has proposed “five stars” of open data as a way of measuring the level of accessibility for data (Berners-Lee 2006). Five-star data must be available online, in a machine-readable and non-proprietary format using open standards from the World Wide Web Consortium (W3C), and must link to other data for context. By adding links to other authorities and unique identifiers, the records in the UNT Name App meet all requirements for five-star open data. Additionally, the Name App automatically generates a unique

identifier for each authority record that is also incorporated into a Uniform Resource Locator (URL); this helps to create infrastructure for the implementation of additional linked data in the future. Unique identifiers also allow disparate records for the same person or organization to be merged into the single, authorized version; permanent links ensure that the original name(s) always redirect appropriately to the correctly established version.

To facilitate metadata creation using authorized names, the UNT Name App is connected to the form-based metadata application -- also written in Django -- used to create and edit metadata at the UNT Libraries. As an editor types a name into the Creator, Contributor, or Publisher fields, Asynchronous JavaScript and XML (AJAX) requests search within the Name App for matches; potential matches appear in a list below the field (see Figure 1). A metadata editor can choose an appropriate name from the list to insert the value into the field. This process helps to promote consistency and prevent data-entry errors as more names are added to the authority records.

The screenshot shows a web form titled "CREATOR". Below the title is a subtitle: "The person, agency, or organization primarily responsible for creating the intellectual content of the resource". There is a "More..." link to the right. The form has three main sections: "name:", "role:", and "info:".

name:
The name of a person, organization, or event (conference, meeting, etc.) associated in some way with the resource

Input field: Joe Sm | Authorized Names

Smith, Joe T.		Dean of Students, PhD		B. 1951	
Smith, Joe		B. 1876	D. 1933		

Dropdown menu: Personal

role:
The role that the person or organization played in the creation of the item

Dropdown menu: Author

info:
Additional information about the creator related to the specific item

Empty text input field

Figure 1. Example of the type-ahead name tool in the metadata Web form.

The Name App is also publicly accessible so that others can search records or use the authorities to create other kinds of metadata. Users who search the authorities can view any information added to a record except notes marked "Nonpublic" (see Figure 2). Some field labels are different in the public interface for clarity.

UNT Name App About Personal Submit

American Society for Information Science and Technology

Authorized:	American Society for Information Science and Technology
Name Type:	Organization
URI:	http://digital2.library.unt.edu/name/nm0000025
Disambiguation:	ASIS&T
Founded Date:	1937-03-13

History:

Important Dates

March 13, 1937: Established as American Documentation Institute (ADI)
 January 1, 1968: Name changed to American Society for Information Science (ASIS)
 2000: Name changed to American Society for Information Science and Technology (ASIS&T)
 2013: Name changed to Association for Information Science and Technology (ASIS&T)

Location

- Silver Spring, Maryland

Links:

- Internal: UNT Libraries Digital Collections
- Homepage: <https://www.asis.org/>
- LOC: <http://id.loc.gov/authorities/names/n00043240>
- Wikipedia: http://en.wikipedia.org/wiki/American_society_for_information_science_and_technology
- VIAF: <http://viaf.org/viaf/135628777>

Variant Name:

- American Society for Information & Technology
- ASIS&T
- ASIST
- The American Society for Information Science and Technology
- Association for Information Science and Technology
- American Society for Information Science

Alternate Formats



Figure 2. Example of an authority record in the UNT Name App.

The UNT Name App establishes an important piece of infrastructure to assist with the introduction of name authority into the uncontrolled collections within the UNT holdings.

4. Workflow for Introducing Name Authority

Introducing name authority is a daunting proposition for a collection of any size; within the UNT digital collections, the theses and dissertations alone encompass more than 7,000 records, each containing names taken from the publications in varying formats. As an initial step toward improving name quality, the authors proposed a workflow for implementing name authority incrementally within the digital collections and decided to pilot the process on a subset of the names and records within the thesis and dissertation collection. Those published after 1999 seemed most feasible, as they have a greater possibility of overlap with authority records in the UNT Name App, which contains names from the UNT Scholarly Works collection, UNT’s institutional repository. This approach would let staff members expand the name authority files in a logical progression based on the name authority work already completed.

Additionally, the authors decided to focus only on names for “contributors” within the collection. The UNT Libraries metadata model distinguishes between the creator of a work (e.g., the author of a thesis or dissertation) and other individuals who had a

secondary or supportive role in the creation of the document (e.g., committee members). In this collection, the names of authors are likely to be unique, while many faculty members have served on multiple committees, allowing for greater comparison of name variations.

The general workflow includes compiling the list of names currently in use for the chosen records, applying basic command-line functions to identify names that have inappropriate formatting to narrow the original list, using an automated lookup function to determine which names already have authority records in the UNT Name App, and manually creating new authority files for the remaining contributor names.

4.1. Step 1: Harvest Metadata Records

The first step involves harvesting the existing set of ETD metadata records from the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) endpoint in the UNT Digital Library. A Python OAI-PMH harvester requests the records in the Dublin Core (oai_dc) metadata format and saves the resulting records harvested from the OAI-PMH repository into a file named untetd.dc.xml. This method of acquiring records was selected because it is transferable to other repositories and can be used for other collections in the UNT Digital Library.

4.2. Step 2: Extract Contributor Names

To determine the level of consistency across the harvested records, contributor names and associated identifiers are extracted using *dc_breaker.py*, a custom-scripted command-line tool for working with metadata records in Unix (Phillips, 2013). One basic command calls the metadata-breaker program and returns the contributor names, ordered by record identifier (see Figure 3).

```
$ python dc_breaker.py -i -e contributor untetd.dc.xml

ark:/67531/metadc2168   Ross, John Robert
ark:/67531/metadc2168   Hummell, Austin
ark:/67531/metadc2168   Wright, Eugene
ark:/67531/metadc2184   Bratton, Sue C.
ark:/67531/metadc2184   Altekruse, Michael K.
ark:/67531/metadc2184   Baker, David
ark:/67531/metadc2181   Clower, Terry L.
ark:/67531/metadc2181   Weinstein, Bernard
ark:/67531/metadc2183   Norton, Scott J.
ark:/67531/metadc2183   Massarachia, Ruth Anne
```

Figure 3. Sample output from *dc_breaker.py*

This output is directed to a file called names.txt, used for the remaining steps of the workflow.

4.3. Step 3: Identify Possible Duplicates

At this point, an important step in any name authority process is to establish a baseline and identify obviously erroneous names from the dataset. There are a number of reasons that errors occur in the names added to ETDs, from variations in the versions of names on the signature pages which metadata creators use for data entry, to basic typographical errors.

The authors have found a simple, effective way to identify a large number of erroneous names from the names in this workflow by using standard UNIX tools such as *sort* and *uniq*. These tools take the sorted list of names (names.txt), provide the number of instances for each name variant, and list the variants in succession to allow for comparison (see Figure 4).

```
$ sort names.txt | uniq -c

  2 Yaffe, Martin
  5 Yaffe, Martin D.
 13 Yancey, George
  5 Yancey, George A.
  3 Yang, Kiseol
  6 Yang, Philip
  2 Yang, Philip Q.
  1 Yang, Phillip
  3 Yeatts, Dale
 29 Yeatts, Dale E.
  2 Yeatts, Dale Elgert
```

Figure 4. Example command and output providing a sorted list of name variants with frequency occurrence.

It is easy to infer through manual evaluation that these eleven lines (in Figure 4) should be normalized down to five distinct names. In all instances of this merging it is possible to identify the version of the name that is the most predominantly used in the existing dataset, this is often a good indicator of the correct version of the name.

After metadata editors merge names that occur sequentially in a sorted list, a new round of processing can take place. Next, a series of weak hashing functions generate new values for each name, appended to the start of the name string in tab-delimited columns. These weak hashing functions help to group similar names which may not have the same formatting.

The first hash function returns an anagram string comprising the unique letters in each name, reordered alphabetically. For example the name “Mark Phillips” and “Phillips, Mark” have an anagram hash value of “ahiklmprs.” By processing the entire list with this function and then sorting the output with frequency counts, non-inverted names become co-located with names that have correctly-inverted formatting (see Figure 5).

```

$ cut -f 2 names.txt | python anagram_hash.py | sort | uniq -c

1 abcdhinors      Brian Richardson
16 abcdhinors     Richardson, Brian
3 abcdhirsuvwz   Schwarz, David B.
20 abcdijlort    Callicott, J. Baird
7 abcdiknprt     Brandt, Patrick
52 abcdklmnouy   Bullock, Lyndal M.

```

Figure 5. Example output of names after applying the anagram hash function.

In the example above, the name “Brian Richardson” and “Richardson, Brian” have the same anagram hash value of “abcdhinors.” These two versions of the name would not have been identified as duplicates through a standard sort process (described above), because they will not sort near each other in a large list of names.

A second hash function helps to identify punctuation and capitalization variations in the dataset. This function utilizes the Name Authority Cooperative (NACO) normalization rules (Library of Congress n.d.) developed to provide a consistent way of sorting and matching name authority records in traditional library catalogs by neutralizing differences in punctuation and capitalization. The implementation used in this project was developed by the Online Computer Library Catalog (OCLC) (OCLC NACO Normalization n.d.) and employs a simplified version of the rules (Hickey Toves & O’Neill 2006) that researchers at OCLC found most helpful for use in sorting and matching names (see Figure 6).

```

$ cut -f 2 names.txt | python naco_hash.py | sort | uniq -c

1 al shareef husam      Al-Shareef, Husam
66 alai rosales shahla  Ala'i-Rosales, Shahla
6 alai rosales shahla  Alai-Rosales, Shahla
3 alai rosales shahla  Ala'i-Rosales, Shahla
1 alai rosales shahla  A'lai-Rosales , Shahla
4 alai rosales shahla s Ala'i-Rosales, Shahla S.
1 albarran alan        Albarran, Alan

```

Figure 6. Example of output of the NACO hash function on a set of names.

This technique helps to catch general typographical errors and extra spaces, as well as names that contain punctuation. These problems may not surface as easily through other methods since disparate functions can interpret punctuation differently.

Identifying name values that have multiple spaces can also present a challenge, depending on the font used for name analysis. Because of this, the authors utilized a simple hashing function they referred to as the *spaceless_hash*, which replaces a space character with a plus sign. While simple, this function helps to identify double spaces more quickly and efficiently than other methods (see Figure 7).

```

$ cut -f 2 names.txt | python spaceless_hash.py | sort | uniq -c

10 Baxter,+Denise+Amy      Baxter, Denise Amy
 1 Beals,++Roberta+J.     Beals, Roberta J.
 1 Beaujean,+Alexander   Beaujean, Alexander
 1 Beck,+Debrah+A        Beck, Debrah A

```

Figure 7. Example of output of the *spaceless_hash* function.

As illustrated, this function allows for the quick and easy identification of double spaces, such as “Beals, Roberta J.” in Figure 7 that may not be as obvious in other function outputs.

Finally, a simple *grep* command can quickly identify names that do not have the correct, inverted format by searching for instances of names that do not contain a comma. These typically indicate records that do not have authorized formatting (see Figure 8).

```

$ sort names.txt | uniq -c | grep -v ","

 1 Lawrence Williams
 1 May. Andrew
 1 Molina. David J.
 1 Nann Goplerud
 1 Noreen Goggin
 1 Sue Bratton
 1 Urs Kreuter
 1 William. D. Deering

```

Figure 8. Example output of values that do not contain commas.

In Figure 8, the *grep* function has provided a list of names that need to be converted to the proper inverted form, as well as several instances where a period was used incorrectly instead of a comma; once corrected by metadata editors, those names will match the proper inverted form.

By working through these lists, metadata editors are able to identify many obvious, easily-identifiable duplicate names and eliminate them by editing the metadata records containing these values.

4.4. Step 4: Look Up Non-Authoritative Names

The next step is to determine whether those names have an authority record in the UNT Name App, and if not, to create a record for the name. Checking for existing records can be automated, while adding new authority files is a manual task. This process can be broken down into three distinct tasks:

- 1) Create an inverted list of names.
- 2) Lookup each name in the UNT Name App.
- 3) Add records for names not present in the authority file.

4.4.1. Create an Inverted List

Up to this point in the process, the workflow has used name files listing the identifier in the first column, followed by a name in that record, forming a second column (see Figure 9).

```
ark:/67531/metadc2168 Hummell, Austin
ark:/67531/metadc2168 Ross, John Robert
ark:/67531/metadc2168 Wright, Eugene
ark:/67531/metadc2181 Clower, Terry L.
ark:/67531/metadc2181 Weinstein, Bernard
ark:/67531/metadc2183 Norton, Scott J.
ark:/67531/metadc2184 Altekruise, Michael K.
ark:/67531/metadc2184 Baker, David
ark:/67531/metadc2184 Bratton, Sue C.
```

Figure 9. Example format for name files used in the workflow.

The format in Figure 9 provides a metadata editor with the record identifier for a given name instance but does not allow for ease of comparing similar (correctly-formatted) name versions, or the number of times that any name occurs across the collection. The authors created a Python script to take this format as input and emit a list with a single name followed by all record identifiers containing that name. In this output, each unique name is represented as a single name entry in the file (see Figure 10).

```
$ cat names.txt | python inverted_list.py

Blow, David ark:/67531/metadc4221 ark:/67531/metadc5829
Blum, Joanne ark:/67531/metadc2565 ark:/67531/metadc3240
Boals, Adriel ark:/67531/metadc103313 ark:/67531/metadc103375
Bocanegra, Cheryl D. ark:/67531/metadc2820
Bodenhamer-Davis, Eugenia ark:/67531/metadc103294 ark:/67531/metadc2185
Bodner, Sarah L. ark:/67531/metadc5306
Boetcher, Sandra ark:/67531/metadc103319 ark:/67531/metadc103393
```

Figure 10. Example output of inverting the list by unique names instead of identifiers.

4.4.2. Look Up Names

A Python script, *name_lookup.py*, uses the inverted list through UNT Name App's "Label API" to search for names in the authority files (About the UNT Name App 2013). The "Label API" for the UNT Name App mirrors the functionality of a similar API in the Library of Congress' ID service, and facilitates automated lookups of names. The UNT Name App "Label API" receives a name string as a request for lookup, applies NACO normalization rules to the input name string, and attempts to find the name string in the database. If the string is present, it redirects the user to the appropriate name record in the system; if it does not find the record, then it returns a "404 Not Found" response code. The Python script prepends the status code to the beginning of each line, using "200" to represent a located name and "404" to represent names not found in the UNT Name App (see Figure 11).

```

$ cat names.txt | python inverted_list.py | python name_lookup.py

200 Albarran, Alan B.      ark:/67531/metadc103415 ark:/67531/metadc115052
200 Albers-Miller, Nancy ark:/67531/metadc4877
404 Albright, Jim        ark:/67531/metadc3064
404 Albright, Lettie K.  ark:/67531/metadc9111

```

Figure 11. Example output of automated searching for authority files using the Label API.

In the example above, each line represents a name string in the dataset, allowing metadata editors to easily identify names that do not have authority records by filtering the list for lines beginning with the “404” code.

4.4.3. Add New Authority Records

As metadata editors consult the output of the *name_lookup.py* script, they can identify name records that need to be created in the UNT Name App and add authority files, along with any relevant alternate forms, disambiguation strings, and links. Additionally, editors may identify name variations and quickly access any metadata record by using the record identifiers in the file to make corrections. Ultimately, all names from the list ought to be added to the UNT Name App or corrected to match an authorized version.

5. Pilot Project

Thus far, the authors have discovered several benefits of using this workflow to apply name authority. First and foremost, automating any part of the process saves valuable staff time. Although the process does not eliminate correctly-formatted variations of a name (e.g., Phillips, Mark, Phillips, Mark E., and Phillips, Mark Edward), it makes many of the variations easier to identify by normalizing incorrect values. Additionally, automated lookups within the UNT Name App quickly narrow the list to names that require manual intervention.

Another important aspect of this workflow is its scalability across staff. Automation allows one person to do many of these steps in a feasible manner. However, if more staff members can be allocated for parts of a name authority project, any part of the workflow can be broken up and performed in parallel. During the pilot project, two staff members were designated to enter new authority records while others edited metadata records to reflect authorized forms of each name. The ability to divide labor also means that improving name authority will not be dependent on specific staff members. A streamlined workflow can reduce errors while improving speed and quality.

The authors also discovered some aspects that may need improvement. For example, the “Label API” used for automatic lookups of names applies NACO normalization to find “fuzzy” matches, rather than exact matches. This means that both a correctly capitalized version (e.g., Phillips, Mark Edward) and an incorrect version (e.g., Phillips, mark Edward) would both return a “200” code to represent a

name in the UNT Name App. Some changes to the API could eliminate the need to find these matches manually by only performing exact matches against the query string.

6. Discussion

Several points have come to the forefront as a result of the pilot project. Despite the relative simplicity of the tools and processes described in the workflow, they have the potential to make significant inroads on name authority and other metadata quality issues, regardless of collection size. However, these flexible pieces do not comprise a complete solution for any institution.

Similarly, although the UNT Name App provides a useful infrastructure for local name authority in the UNT digital collections, it serves as an initial step in the overall process. Currently, UNT metadata records store names as strings rather than links to the unique identifiers in the UNT Name App. Changing to storing links will require a significant shift in the current digital library infrastructure, but it will allow metadata editors to specify when a name in a metadata record refers to an entity in the UNT Name App (or when the connection is unverified). Ultimately, it also offers the potential for creating more engaging interfaces for end users to make information in online digital collections meaningful and useful (Tarver et al. 2013).

Also, the authors made use of an API in the UNT Name App which allowed for the programmatic lookup of name records and resolution to their authoritative format in an easy straightforward way. This ability mirrors functionality present in the Library of Congress ID service and has proved to be quite useful. As more institutions seek to integrate locally-developed and controlled name authority data, it is suggested that a similar set of APIs be made available by these new services.

Finally, the authors feel that the current practice in the UNT ETD metadata creation workflow will need adjustment; copying advisor and committee member names exactly as printed on the signature sheet cannot continue once full name authority control has been implemented in the ETD collection. As new metadata records are created, the majority of names will have authorized forms in the UNT Name App from previously-submitted documents, simplifying the process and substantially raising the quality of the new records over previous records.

However, when metadata creators encounter names that are not in the UNT Name App, the revised workflow will need to include establishing an authoritative version of the names during metadata creation. This issue may also affect other collections as metadata editors exert name authority across other portions of the digital holdings.

7. Conclusion

While introducing name authority in large, established collections - or any digital library environment - may seem daunting, basic tools can assist with automating

initial steps in the process and inform staff members about the best ways to proceed based on name frequency in their metadata. The project described in this paper provides one possible method of approaching name authority. This process allows for flexibility in the numbers of staff members assisting with the problem and does not rely on having name authority software in place at the start of the project. Although all name authority has challenges, taking first steps will improve metadata quality and set the groundwork for improved infrastructure and processes.

References

About the UNT Name App 2013. Available from <http://digital2.library.unt.edu/name/about/>. [19 August 2013]

Berners-Lee, T 2006, *Linked data*. Available from: <http://www.w3.org/DesignIssues/LinkedData.html>. [8 August 2013].

Hickey, T, Toves, J, O'Neill, E 2006, 'NACO Normalization: A detailed examination of the authority file comparison rules', *Library Resources & Technical Services*, vol. 50, no. 2, pp. 166-172.

Library of Congress Linked Data Service Technical Center, n.d., Available from: <http://id.loc.gov/techcenter/searching.html>. [19 August 2013].

Library of Congress Cataloging and Acquisitions NACO - Name Authority Cooperative Program of the PCC, n.d., Available from <http://www.loc.gov/aba/pcc/naco/>. [19 August 2013].

Niu, J 2013, 'Evolving landscape in name authority control', *Cataloging & Classification Quarterly*, vol. 51, pp. 404-419.

OCLC NACO Normalization Services, n.d., Available from <http://www.oclc.org/research/activities/naco.html>. [19 August 2013]

Phillips, M, 2013. Metadata Analysis at the Command-Line. *Code4Lib Journal*, 19.

Tarver, H, Waugh, L, Phillips, ME, Hicks, W 2013, 'Implementing name authority control into institutional repositories: A staged approach', *Open Repositories Conference* <http://digital.library.unt.edu/ark:/67531/metadc172365/>.

Wolverton, G, Hoover, L, Hall, SL, Fowler, R 2008, 'Issues of name authority control in the cataloging of theses and dissertations', *Technical Services Quarterly*, vol. 25, sup. 1, pp. 75-94.

Xia, Jingfeng 2006, 'Personal name identification in the practice of digital repositories', *Program: electronic library and information science*, vol. 40, no. 3, pp. 256-267.