Rethinking the University of Maryland Authority File for the Linked Data Environment

[Shortened title: Rethinking the Maryland Authority File]

Bria Parker, Head, Discovery and Metadata Services (blparker@umd.edu)

Adam Gray, Graduate Assistant, Discovery and Metadata Services (adgray987@gmail.com)

University of Maryland Libraries

Abstract

The development and implementation of a new digital collections system built on the Linked Data Platform has provided University of Maryland Libraries with an ideal opportunity to prototype and test ways to model local corporate name authorities in RDF. This includes assessing the local corporate names metadata, reconciling these names against existing authorities, and devising and executing an RDF model for unreconciled names in support of the new linked data environment.

Keywords: name authorities, linked data, authority control, digital collections, data reconciliation

Introduction

The University of Maryland Libraries began digitizing and depositing digitized archival material into a digital collections system built on Fedora (https://wiki.duraspace.org/display/FF/Fedora+Repository+Home), an open source repository system, in 2005. To date, the Libraries have digitized and made available 45,000 items, ranging from letters and photographs to newspapers, films, and radio programs, with varying degrees of

access. While this repository has been operational for over a decade, our efforts toward

controlling agents and names have not been consistent or entirely successful. This has resulted in

having multiple entries for some individuals or corporations, with alternate name forms as well

as misspellings making their way into our system.

While digital collections of archival materials are stored in the Fedora-based system, the

Libraries maintain a separate repository, built on DSpace, for the scholarly output of the

University. This second repository, the Digital Repository at the University of Maryland

(DRUM) (https://drum.lib.umd.edu/), serves as the catalog of record for theses and dissertations

produced at the University, in addition to preserving and providing access to a wide variety of

other types of research products.

With new systems in development, it was time for us to evaluate our approaches thus far and

begin planning and testing how it can be improved moving forward.

Background

As mentioned above, University of Maryland Libraries (hereafter UMD) uses Fedora for its

underlying digital collections architecture. While the bulk of our digital collections of archival

materials are maintained in a system built on Fedora version 2, the Libraries have recently

launched some notable digital collections in Fedora version 4 (hereafter Fedora 4). A significant

difference between our legacy system (Fedora 2) and Fedora 4 is the support Fedora 4 offers for

the Linked Data Platform (LDP) specification. In the case of Fedora 4, LDP:

[d]escribes a set of best practices and simple approach for a read-write Linked Data

architecture, based on HTTP access to web resources that describe their state using the

RDF data model. Fedora 4 implements the LDP specification for create, read, update and

delete (CRUD) operations, allowing HTTP, REST, and linked data clients to make

requests to Fedora 4

(https://wiki.duraspace.org/display/FEDORA475/Linked+Data+Platform).

In short, Fedora 4 supports metadata modeled using the resource description framework (RDF)

model, whereas our Fedora 2 system stores metadata in XML. This shift in modeling, from XML

to RDF, has provided UMD the opportunity to rethink and revise our metadata creation and

implementation workflows to take advantage of the opportunities provided by RDF.

Much has already been written about the challenges of authority control in digital collections and

the shortfalls of digital collections systems and workflows to address those challenges (Dragon,

2009; Myntti & Cothran, 2013; Myntti & Neatrour, 2015; see also Lopatin, 2010). University of

Maryland Libraries is no different from other institutions in that respect. The Fedora 2 repository

at UMD has no formal or systematic control for names, relying on the consistency of the library

staff entering the metadata. Our data entry and quality control documentation advises looking up

names in Library of Congress Name Authority File (LCNAF) for the correct name form.

Additionally, our administrative interface has a vocabulary list that can be added to for nearly

every field, including names, and also includes a blacklist. The metadata entry form uses this for

typeahead autocomplete suggestions for items in these vocabularies. However, this approach has

not always yielded consistent data, as the blacklist does not prevent the blacklisted terms from being entered. As we have been in development with Fedora 4 since 2014, we deemed it unnecessary to go back and do additional development work to remedy this faulty blacklist in Fedora 2, a system that we would eventually migrate away from. At this time, there is no concrete plan to develop a similar blacklist tool in Fedora 4. The result of these limitations to authority control might be considered a form of technical debt that complicates discovery and management of digital resources (Clair, 2016; Joyce et al., 2018).

One opportunity we have pursued in Fedora 4 is name disambiguation and authority control, both personal and corporate. For collections in Fedora 4, we have been able to create agent objects to which digital collections objects can be linked. Thus far the focus has been mostly on individuals, but the further we get into implementation, the more need we see to not only disambiguate names, but to also define relationships where appropriate. This is especially true with corporate names, particularly the names of schools, departments, and offices at the University of Maryland. The nature of these local corporate names - how frequently they are changed, how they fit into larger hierarchies of a school or department - makes them an ideal case study (Yoshimura, et al., 2016, p. 14-15).

Concurrent with our development of Fedora 4 has been our migration to ArchivesSpace, a content management system for archival metadata. While ArchivesSpace is not an implementation of the Linked Data Platform, some aspects of its architecture for named entities and subjects lend themselves to the type of hierarchical relationships that RDF makes easier to demonstrate, as well as the changes made to the authoritative version of named entities over

time. For example, ArchivesSpace enables us to indicate superordinate/subordinate relationships between offices and departments within a large labor organization for which we have significant archival holdings, creating a fuller picture of the relationships between the agents, in addition to having controlled forms of the names.

UMD is not the first institution to grapple with workflows around local name authorities, or with managing authorities in RDF. There are several projects that are seeking to support authoritative data. One is Vitro, which is a "general-purpose web-based ontology and instance editor with customizable public browsing. Vitro is a Java web application that runs in a Tomcat servlet container" (http://vitro.mannlib.cornell.edu/). Vitro supports the creation of ontologies in Web Ontology Language (OWL) format as well relationships. Vitro is most widely known for being one of the technologies that support VIVO, an ontology and open-source software for representing scholarship (https://wiki.duraspace.org/display/VIVO/VIVO).

Vitro and VIVO both originate from Cornell University Libraries, which has spearheaded the IMLS-funded "Shareable Authorities" project. This project highlighted the possibilities that are afforded by the use of linked data in creating authorities, including encouraging reuse by and interoperability with other systems (Casalini et. al., 2018).

Also of note is the grant-funded Western Name Authority File (WNAF) Project, which aims to provide open, scalable, reusable infrastructure and workflows around authority control. "Currently in the Mountain West Digital Library (MWDL) and at local institutions, name variants provide users with unnecessary additional search options. A central name authority file

like the WNAF can provide an essential reference tool for catalogers and metadata librarians"

(https://sites.google.com/site/westernnameauthorityfile/).

Reconciliation and Enhancement with OpenRefine

As the state's flagship university, the University of Maryland is home to many graduate programs

whose students produce a large number of theses and dissertations, which were originally

cataloged in the local ILS (prior to Institutional Repositories becoming the catalog of record for

these materials). The cataloging of these materials necessitated developing control over the

names of the University departments, schools, research centers, etc., related to the creation of

these materials. Beginning in at least 1998, catalogers in the Authority Control Unit maintained

the Maryland Authority File, a physical card catalog authority file that eventually transitioned to

being represented as a list on the Libraries' website. As of today this list of University of

Maryland names established for use on theses or university publications is on the Libraries'

intranet site, and has not been updated since approximately 2012. Around the same time as this

last update, theses and dissertations were no longer routinely cataloged in the ILS and were

instead deposited into our Institutional Repository, DRUM. While DRUM serves as the catalog

of record for theses and dissertations, it does not make use of the Maryland Authority File for

authority control of local department names. Our repositories for digital archival collections,

Fedora 2 and Fedora 4, also do not make use of the Maryland Authority File.

The Maryland Authority File website contains a list of approximately 240 established names of

UMD schools, academic departments, and research centers. On this site, names were categorized

as either subordinate or independent (i.e., names that are preceded by *University of Maryland*, as opposed to names that stand alone) and listed in alphabetical order within each of the two categories. Several names contained notes on usage dates, earlier and later forms, and "see" notes. However, the relationships among research centers, academic departments, and schools were not described. Additionally, 46% of names in the original list did not have LCNAF authority records. Some of these pointed to other names in the list with "see" notes; most others are names of programs, departments, centers, and offices.

Before we could reconcile and convert the Maryland Authority File into RDF, we needed to finalize the model we wanted to use for these local authority objects. We examined the data we already had, the data available in the LCNAF records, and also consulted prior work done by Europeana and the Digital Public Library of America (DPLA) to inform our initial data model (Europeana, 2017; DPLA, 2017). We considered what we believed would be beneficial down the road, such as earlier established names, alternate name forms, dates associated with the name, related names (including both subordinate and superordinate), and associations to Real-World Objects (RWO). We also used a notes predicate (skos:note), which allows us to collect any notes from an established authority record, but will also allow us to store notes about a locally created entity that can be used for reference when a review or update is needed. This data model will facilitate future changes in our institutions by allowing us to indicate prior names, dates, and location within the institutional hierarchy. We ultimately used a mix of properties from SKOS and MADS/RDF, mostly informed by ones already in use by Library of Congress Linked Data Service (http://id.loc.gov). SKOS is widely used for creating and classifying thesauri, while MADS/RDF has specific affordances for names (https://www.w3.org/2004/02/skos/; http://www.loc.gov/standards/mads/rdf/) (see Figure 1).

Class/Property used	Relationship	Related class/property
skos:Concept*	reg:hasSubClass	madsrdf:Authority*
madsrdf:Authority*	rdfs:subClassOf	skos:Concept*
skos:prefLabel*	reg:hasSubproperty	madsrdf:authoritativeLabel*
skos:altLabel*	reg:hasSubproperty	madsrdf:hasVariant*
skos:inScheme*	reg:hasSubproperty	madrdfs:isMemberOfMADSScheme*
skos:note	owl:equivalentProperty	madsrdf:note
skos:broader	reg:hasSubproperty	madsrdf:hasBroaderAuthority
skos:narrower	reg:hasSubproperty	madrdfs:hasNarrowerAuthority
skos:related	reg:hasSubproperty	madsrdf:hasReciprocalAuthority
skos:closeMatch	reg:hasSubproperty	madrdfs:hasCloseExternalAuthority
madsrdf:authoritativeLabel*	rdfs:subPropertyOf	skos:prefLabel*
madsrdf:hasEarlierEstablishedForm*	sub-subPropertyOf	rdfs:seeAlso*
madsrdf:hasLaterEstablishedForm*	sub-subPropertyOf	rdfs:seeAlso*
madsrdf:identifiesRWO*	rdfs:subPropertyOf	foaf:focus
dcterms:modified		

Namespaces for selected Classes and Properties

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
@prefix reg: <http://metadataregistry.org/uri/profile/RegAp/>
@prefix owl: <http://www.w3.org/2002/07/owl#>
@prefix skos: <http://www.w3.org/2004/02/skos/core#>
@prefix dcterms: <http://purl.org/dc/terms/>
@prefix madsrdf: <http://www.loc.gov/mads/rdf/v1#>

* indicates class or property from Library of Congress Linked Data Service Model

Figure 1: List of Classes and Properties Used, with namespaces

With the data model set, we transferred the Maryland Authority File into an Excel spreadsheet and prepared it for reconciliation using OpenRefine (http://openrefine.org), hereafter referred to as Refine. The organization of the list by name type - subordinate and independent - was maintained at this stage by using two separate sheets and importing these into two separate Refine projects. The list of "independent" names was the smaller of the two, and was therefore used to test the procedures for reconciling headings with those in the LCNAF, fetching URIs for those headings, and extracting earlier and later established forms of the names, alternate headings, and subordinate headings, from the authority files, as per our data model. Once this smaller list had been reconciled and enhanced with the values specified by the data model, the same procedure could be applied to the larger list. The two lists could then be combined and structured to meet the requirements of the RDF schema alignment tool, which exports Refine projects as RDF in Turtle and RDF/XML serializations.

The nature of the source metadata demanded some initial manipulation using basic Refine functions. For example, some "see" notes were unintentionally transferred into the spreadsheet column used for names; this was easily fixed using the filter and add column functions. Similarly, a number of entries were not headings, but administrative notes indicating usage dates for earlier or later headings for a single entity. In such cases, the information was atomized and retained in separate columns to enable accurate matching during reconciliation.

The procedures we used to reconcile headings followed a course similar to those developed and described by others (Carlson & Seely, 2017; Myntti & Neatrour, 2015; Southwick, 2015; Wright & Carruthers, 2015). In our case, several reconciliation services are available. We tested a

reconciliation process written by Matt Carruthers, LCNAF Named Entity Recognition (https://github.com/mcarruthers/LCNAF-Named-Entity-Reconciliation). One of the potential benefits of using this script is that it creates a column containing the LCNAF names that match the original names, and another column containing the URLs of the corresponding authority records. This script matches names automatically, and manual confirmation showed the majority of the matches it produced for our names were accurate.

We also experimented with another reconciliation service, Conciliator, created by Jeff Chiu (https://github.com/codeforkjeff/conciliator), and found it to be the most useful for our purposes. This service produces a ranked list of candidate matches for each name that, like the LCNAF Named Entity Recognition script, provides an efficient means of updating name headings to match their current, authorized forms. Just as the majority of the matches returned by LCNAF Named Entity Recognition were accurate, so were the majority of "best candidate" matches returned through Conciliator. The ability to view and choose from the list of candidate matches offers the convenience of evaluating potential matches before making a judgment. The ranked list feature also provides the ability to facet names by judgment, which allows for more granular manipulation and progress assessment. These additional benefits made the Conciliator service the most useful for our purposes. Of course, this was not helpful for the 46% of names lacking authorized headings in LCNAF.

Once we identified and selected the appropriate matches returned by the reconciliation service, we used General Regular Expression Language (GREL) expressions to access the reconciled object (*cell.recon*) to retrieve and store both the literals of the skos:prefLabel for matched cells

(e.g., A. James Clark School of Engineering) as well the URIs from which those labels derive

(e.g., http://id.loc.gov/authorities/names/no96025592). The content of the URIs could then be

fetched and parsed to extract the additional metadata specified by the data model

(https://github.com/OpenRefine/OpenRefine/wiki/General-Refine-Expression-Language).

Through trial-and-error, we developed a reliable, if cumbersome, procedure to extract the values

specified by our preliminary data model and prepare them for RDF serialization. The LC Linked

Data Service makes LCNAF records available in several formats, such as XML, JSON, and n-

triple serialization (http://id.loc.gov). We experimented with fetching JSON RDF, but found the

records challenging to parse. As this project was an initial foray, and we had previous experience

parsing HTML from prior work, we chose to fetch and parse the HTML source. After fetching

the URIs from the LC Linked Data Service for each reconciled authority, we parsed the HTML

source to isolate the container elements that held values we wanted to extract. For example, the

literal values for alternate labels are held in an HTML list element which we parsed and moved

into a new column (figure 2a). This list, in turn, was further parsed to yield the alternate labels

themselves, which we accessed by specifying their index numbers in the GREL expression

(figure 2b).

Add column based on column Variants

New column name   skos:altLabel

On error   ● set to blank   ○ store error   ○ copy value from original column

Expression   Language   Google Refine Expression Language (GREL) ▼

```
value.parseHtml().select("span")[0].innerHtml()
```
No syntax error.

Preview   History   Starred   Help

row	value	value.parseHtml().select("span")[0].innerHtml()
1.	` <h3>Variants</h3> <ul class="std"> <li rel="madsrdf:hasVariant skosxl:altLabel"> <div about="[_:bn8]" typeof="madsrdf:Variant skosxl:Label madsrdf:CorporateName"> Glenn L. Martin Institute of Technology. A. James Clark School of Engineering </div> <li rel="madsrdf:hasVariant skosxl:altLabel"> <div about="[_:bn20]"`	Glenn L. Martin Institute of Technology. A. James Clark School of Engineering

OK   Cancel

Add column based on column Fetch 1

New column name   Variants

On error   ● set to blank   ○ store error   ○ copy value from original column

Expression   Language   Google Refine Expression Language (GREL) ▼

```
value.parseHtml().select("li").slice(16,21).join("|")
```
No syntax error.

Preview   History   Starred   Help

row	value	value.parseHtml().select("li").slice(16,21).join("|")
1.	`<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd"> <html version="XHTML+RDFa 1.0" xmlns="http://www.w3.org/1999/xhtml" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v xmlns:ri="http://id.loc.gov/ontologies/RecordIn xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"`	` <h3>Variants</h3> <ul class="std"> <li rel="madsrdf:hasVariant skosxl:altLabel"> <div about="[_:bn8]" typeof="madsrdf:Variant skosxl:Label madsrdf:CorporateName"> Glenn L. Martin Institute of Technology. A. James Clark School of Engineering </div> <li rel="madsrdf:hasVariant skosxl:altLabel"> <div about="[_:bn20]" typeof="madsrdf:Variant skosxl:Label`

OK   Cancel

Figures 2a and 2b. Parsing HTML from id.loc.gov to extract alternate labels.

As Refine's RDF schema alignment tool draws the values for each assertion of a property from a single column, we used repeating fields to accommodate multiple values: as there were as many as twelve alternate labels for a given name, for instance, there were twelve columns for the skos:altLabel property. We used the same procedure to extract URIs and labels for the other SKOS properties. However, our reliance on repeating fields would create complications for the export of matchless names, as will be described in the next section.

Predictably, the 110 names lacking LC authority files demanded significantly more time to enrich. Furthermore, while inspecting these names we discovered that several headings for colleges and departments were left out of the original data. Some of the omissions were due to organizational changes in the last 10 years (e.g., in 2010 the College of Computer, Mathematical and Natural Sciences formed from the merger of two colleges: the College of Chemical and Life Sciences, and the College of Computer, Mathematical and Physical Sciences. There were headings for the two, pre-merger Colleges, but not for the newer College). Fortunately, the Maryland State Archives maintains current information on the University's organizational structure, including its fourteen colleges and schools, as well as their many departments and research centers, via the *Maryland State Manual* (2018). Since Refine is not designed for creating new records, we exported the project and continued the work in Microsoft Excel. Using the *Maryland State Manual* as a guide, we updated the unreconciled headings and entered headings for entities omitted by the original list, and in turn reused these new headings to represent their relationships, mapping them to the appropriate properties. By transposing columns to rows, a carefully structured group of related headings could be copied from one

group of columns and used to populate the repeating columns holding the skos:related property. To avoid making reflexive statements (e.g., <Maryland Institute for Applied Environmental Health> skos:related <Maryland Institute for Applied Environmental Health>), we removed those related headings that were identical to the subjects of the RDF triples in which they occurred. Once the unmatched headings were updated and enriched, we moved the project back to Refine to configure the data using the RDF schema alignment tool and exported a prototype RDF serialization.

Challenges

While the process was a step toward an efficient means of reusing LCNAF data to enrich our local authorities and structure them as RDF, it presented several issues that we will have to address before we attempt to carry it out on a larger scale. First, the majority of the HTML fetched from the LC Linked Data Service was irrelevant to our project. The values we *did* want to extract were nested under several layers of HTML elements and required a combination of manual searches in the HTML document and experiments with GREL expressions to isolate the appropriate elements. Second, our HTML parsing generated a significant amount of noise. This was especially true when we extracted alternate headings (skos:altLabel). As we looped through the index numbers corresponding to each alternate heading, we began to extract extraneous HTML content for names that have few alternate headings. For future iterations, we plan to improve our ability to parse different languages and fetch one of the available RDF serializations rather than the HTML to make extracting values simpler.

A third challenge arose during RDF export and impacted those unmatched headings that were updated and enriched in Excel. When we initially removed reflexive-related entities, we left null cells between cells with values, creating gaps between the repeating fields. Our first attempt at serialization revealed that Refine's RDF export stops when a null cell is reached. Consequently, columns of repeating fields must be contiguous for the complete RDF triples to be exported. This necessitated a workaround using Excel to shift values in repeating fields to make them contiguous. While there is likely also a way to do this in Refine, our familiarity with Excel made it the most expedient option.

The RDF schema alignment tool in Refine was a straightforward way to convert our data into RDF, with very little learning curve for us. Now that we have a proof of concept in place, we hope to develop a more scalable and reusable script and workflow using a language such as Python to convert datasets in .csv files to RDF. This would require us to create a .csv template, as well as the set of RDF predicates to which the data would map.

Future Work

The goal for this first project was to prototype a data model that worked for the metadata at hand to format names with no formally established authorities, using the model established by Library of Congress. Now that this prototype exists we can evaluate it against other sets of uncontrolled names as well as additional metadata schema to see where more focused properties might be required, where additional or alternate properties and classes would be necessary, as well as where broader predicates might be required.

As our data model grows and evolves, one property we hope to add to this work is mads:hasHiddenLabel. This could do much to keep the names in the repository consistent, as it has the potential to gently direct personnel entering data to use the correct name form and avoid common spelling errors; any variant entered, whether it is stored in mads:hasVariant or mads:hasHiddenLabel, would prompt the suggestion of the preferred label. Use of the mads:hasHiddenLabel predicate could be a way to create a successful blacklist feature in Fedora 4, and enable discovery without displaying a particular name form.

A number of the uncontrolled names across our systems are attached to archival materials, so we foresee including classes and properties from the EAC-CPF ontology. EAC-CPF may be particularly useful modeling in family names and relationships. Wherever possible, our intention is to note equivalent properties and classes as they are defined in the MADS/RDF schema, as well as subproperties and subclasses for what is selected, in order to be transparent about how the model relates to other schema used for other purposes (http://www.loc.gov/standards/mads/rdf/v1.n3). Documenting this mapping will be important as this work scales and necessitates developing the indexes used by our systems.

Final Thoughts

Modeling UMD's corporate name authorities as linked data has made relationships among authorities more visible and brings context to these related entities. Making authority data available in this structure will lend clarity to University collections by making historical name

forms and relationships more explicit. It also integrates well with agent modeling in ArchivesSpace, which makes use of relationship types (e.g., parent/child) that can be expressed using RDF. Similarly, this method of structuring information about authority relationships follows conventions already in place in UMD Libraries' Fedora 4 implementation. The model, as well as the processes of reconciling, structuring, and exporting metadata can be modified and repurposed in the service of other sets of corporate and personal name authorities.

Another system where we can see this type of work having great benefit is in the Digital Repository at the University Maryland (DRUM), the University's digital repository that is built on Dspace, mentioned above. As the central repository for dissertations and theses going back at least 90 years, structured name authority data for University departments, schools, and colleges can provide added value by linking the research record of the past to departments today, as well as provide a way to maintain that link into the future, as the names of colleges and departments are certain to evolve.

While development for Fedora 4 continues, UMD is investigating the implementation of a triplestore, separate from the main Fedora 4 triplestore, for storing this structured authority metadata in RDF, as well as providing a way to keep the metadata up-to-date. We do not yet have a specific plan in place for sharing our local authorities externally, however, we planned our work with this end goal in mind by establishing a data model that enables sharing. The implementation of this proposed triplestore for the local authorities established in this project would enable us to open up a SPARQL endpoint of our local names. Participation in aggregated digital collections such as DPLA could also benefit from local name authorities established by

UMD, especially once work is expanded to include personal names associated with our premier archival collections on labor, broadcasting, and state history. The focus for this initial project was on corporate name authorities related to the University of Maryland but there is a great deal of potential for expanding this work to other local names, as well as for locally storing names and identities already established in places like LCNAF or VIAF.

References

Casalini, M., Chiat Naun, C., Cluff, C., Durocher, M., Folsom, S., Frank, P. (2018, March) *National Strategy for Shareable Local Name Authorities National Forum* (White Paper). http://hdl.handle.net/1813/56343

Carlson, S. & Seely, A. (2017). Using OpenRefine's Reconciliation to Validate Local Authority Headings. *Cataloging & Classification Quarterly, 55*(1), 1-11. DOI: 10.1080/01639374.2016.1245693

Clair, K. (2016). Technical Debt as an Indicator of Library Metadata Quality. *D-Lib Magazine*, *22*(11-12). http://www.dlib.org/dlib/november16/clair/11clair.html

DPLA (2017). Metadata Application Profile version 5.0. Retrieved from DPLA website: http://dp.la/info/map

Dragon, P. (2009). Name Authority Control in Local Digitization Projects and the Eastern North Carolina Postcard Collection. *Library Resources & Technical Services*, *53*(3), 185-196.

Europeana (2017). Europeana Data Model – Mapping Guidelines v2.4. Retrieved from Europeana website: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf.

Joyce, D., Corrin, J., Hirsch, R., Johnston, R., Kinhart, E., McPhee, L. (2018 August). *True Confessions: Paying Off the Technical Debt of Early Digital Projects*. Session presented at Archives*Records 2018: Joint Annual Meeting of Council of State Archivists, National Association of Government Archives and Records Administrators, and Society of American Archivists, Washington, D.C.

Lopatin, L. (2010). Metadata Practices in Academic and Non-Academic Libraries for Digital Projects: A Survey. *Cataloging & Classification Quarterly, 48*(8), 716-742. DOI: 10.1080/01639374.2010.509029

Maryland State Archives (2018, September 18). University of Maryland, College Park. Retrieved from https://msa.maryland.gov/msa/mdmanual/25univ/umcp/html/umcp.html

Myntti J. & Cothran N. (2013). Authority Control in a Digital Repository: Preparing for Linked Data. *Journal of Library Metadata, 13*(2-3), 95(113). DOI: 10.1080/19386389.2013.826061

Myntti, J. & Neatrour, A. (2015). Use Existing Data First: Reconcile Metadata before Creating New Controlled Vocabularies. *Journal of Library Metadata, 15*(3-4), 191-207. DOI: 10.1080/19386389.2015.1099989

Salo, D. (2009). Name Authority Control in Institutional Repositories. *Cataloging & Classification Quarterly, 47*(3-4), 249-261, DOI: 10.1080/01639370902737232

Smith-Yoshimura, K., Gatenby, J., Agnew, G., Brown, C., Byrne, K., Carruthers, M., . . . OCLC Research. (2016). Addressing the challenges with organizational identifiers and isni. Dublin, Ohio: OCLC Research. (2016). http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-organizational-identifiers -and-isni-2016.pdf.

Southwick, S. (2015). A Guide for Transforming Digital Collections Metadata into Linked Data Using Open Source Technologies. *Journal of Library Metadata, 15*(1), 1-35. DOI: 10.1080/19386389.2015.1007009

Wright, J. & Carruthers, M. (2015). Breaking the Bottleneck: Automating the Reconciliation of Named Entities to the Library of Congress Name Authority File [Presentation]. Retrieved from http://hdl.handle.net/2027.42/138107