Improving Discovery of and Access to Digital Repository Contents Using Semantic Web Standards: Columbia University's Academic Commons[1]

Robert J. Hilliker[2]
Hotchkiss School

Melanie Wacker[3]
Columbia University

Amy L. Nurnberger[4]
Columbia University

ABSTRACT

This article describes the progress made towards developing Academic Commons (AC), Columbia University's digital repository, as an interoperable repository through the use of RDF and non-RDF Semantic Web technologies. Approaches taken include the implementation of microdata to add semantic markup to HTML content; a collaboration with Oregon State University's (OSU) digital repository, ScholarsArchive@OSU (SA@OSU), to implement an application that indexes RDF data from OSU for use in AC; as well as an exploration of the recently released MODS RDF.

Keywords: Microdata, MODS, RDF, Semantic web

INTRODUCTION

Academic Commons (AC) is the digital research repository of Columbia University (CU) and its affiliated institutions: Teacher's College, Barnard College, Jewish Theological Seminary, and Union Theological Seminary. Over the past eight years the underlying technology has evolved, the service offerings around it have matured, and the collection has grown. Today it contains over 10,000 resources, such as working papers, conference papers, theses, articles, book chapters, presentations, performances, musical scores and data sets. This article presents the Academic Commons Team's approach to and experiences with transforming AC from a stand-alone resource into an interoperable repository using a variety of Resource Description Framework (RDF) and non-RDF Semantic Web technologies.

BACKGROUND

The Academic Commons Team is not alone in its efforts in using RDF and Semantic Web approaches to improve repository functionality, as illustrated by the number and variety of projects that have emerged in recent years. An "inference-based knowledge discovery, retrieval and navigation" (Koutsomitropoulos, Solomou, Alexopoulos & Papatheodorou, 2009a, p. 1) service, sitting on top of a traditional DSpace repository, was developed at the University of Padras digital repository. The outcome of this project has been made available to all DSpace users and its developers see potential for its application in other repository platforms as well (Koutsomitropoulos, Solomou, Alexopoulos & Papatheodorou, 2009b). Other initiatives, such as COnnecting REpositories (CORE) project (http://core.kmi.open.ac.uk/search), focus on making semantically related resources discoverable across repositories. The RKBExplorer (http://www.rkbexplorer.com/explore/) is a semantic browser enabling researchers to access content from a range of data providers (Glaser, Millard & Jaffi, 2008). A number of institutional repositories along with other relevant Linked Data sources such as DBpedia are currently listed as RKBExplorer content sources.

Specialist repository communities are also working towards enabling their repositories for the Semantic Web. Subiratis et al. (2012) describe the efforts made by the agricultural information management community. The Agricultural Information Management Standards Team of the Food and Agriculture Organization of the United Nations issued recommendations to repositories in their domain suggesting ways to improve functionality and interoperability through Semantic Web standards. One main recommendation is the use of controlled vocabularies expressed in Simple Knowledge Organization System (SKOS), particularly the domain specific AGROVOC thesaurus (http://aims.fao.org/standards/agrovoc/).

Oregon State University's institutional repository ScholarsArchive@OSU has approached authority control slightly differently, utilizing RDF. This feature is of particular utility for institutional repositories, since they contain a great deal of content by authors whose names are not registered in standard name authority systems (Johnson & Boock, 2012). These successful examples inspired the

---

[1] This is an electronic version of an article published in *Journal of Library Metadata*, 13(2-3), 80–94. *Journal of Library Metadata* is available online at: http://www.tandfonline.com/doi/full/10.1080/19386389.2013.826036.
Recommended citation:
Hilliker, R. J., Wacker, M., & Nurnberger, A. L. (2013). Improving Discovery of and Access to Digital Repository Contents Using Semantic Web Standards: Columbia University's Academic Commons. Journal of Library Metadata, 13(2-3), 80–94.
doi:10.1080/19386389.2013.826036
[2] ORCID: 0000-0002-1729-0222
[3] ORCID: 0000-0002-7005-5627
[4] Corresponding author:
420 West 118th St., New York, NY 10027
anurnberger@columbia.edu
ORCID: 0000-0002-5931-072X

Academic Commons Team to invest efforts in improving AC repository functionality and accessibility by increasing its capacity for interoperability.

APPROACHES

Setting the Stage with Blacklight

Academic Commons is not only a place where the work of CU affiliated researchers, faculty, and students is being collected and preserved: it also makes CU's research output openly accessible to the world. Enhanced discoverability of AC contents through its portal and the major search engines is therefore of great importance. The Academic Commons Team at Columbia's Center for Digital Research and Scholarship (CDRS) has worked to continuously improve AC's portal, functionality, and underlying metadata to increase accessibility and discoverability. To this end, a new portal with a host of additional features designed to support these goals was built on Blacklight (http://projectblacklight.org/), an open source Ruby on Rails application, and launched in April 2011 (Bufanio, 2011).

The decision to use Blacklight was based on a recognition that its technical architecture allows for superior indexing, search results ranking and display of metadata, particularly for heterogeneous collections like AC (Moore & Greene, 2012). A key component of that technical architecture is the Open Source search engine, Apache Solr (http://lucene.apache.org/solr/), which is used by major commercial enterprises, including Zappos and Netflix, because it provides rich full-text indexing for documents in a wide variety of formats and can handle a wide range of complicated queries, including Boolean operators, double-quotes around phrases, and wildcard operators (Alhabashneh, Iqbal, Shah, Amin & James, 2011). It also handles faceted search natively (Sadler, 2009). In fact, search limiting by facets and better browsing by departments and subjects were, from an end-user perspective, the most noticeable improvements that Blacklight enabled in AC.

In order to exploit these improved search and browsing facilities AC staff use information from the institutional directory to improve item metadata, verifying the correct name and department affiliation for Columbia faculty and staff with content in the repository, and assigning ProQuest subject categories to all resources, which allows for controlled access by topic and clean faceting. Indeed, one of the major challenges of using Blacklight is that it exposes metadata so well that any misspellings or other entry errors are immediately apparent to users. Thus AC staff spent a good deal of the spring and summer of 2011 remediating repository metadata to ensure that all values were normalized and that the individual records were compliant with version 3.4 of the Metadata Object Description Schema (MODS: http://www.loc.gov/standards/mods), an XML-based metadata schema developed by the Library of Congress, and Columbia University Libraries / Information Services' (CUL/IS) preferred schema for descriptive metadata.

After these improvements, AC enjoyed radically increased visibility in major search engines, with traffic from search (as opposed to direct traffic and referrals from other sites, such as the CUL/IS homepage) leaping from 13% to 61% as a proportion of the overall traffic, and a 128% jump in visitors for May, June and July 2011 as compared to the previous quarter, despite June and July historically being lower traffic times of year.

In the wake of this success, the Academic Commons Team focused on leveraging the existing, high quality metadata created and curated by repository staff for further search engine optimization (SEO), but also to enable greater interoperability with other information systems. On a basic level, this has meant serializing descriptive metadata in a wider variety of formats. With so many possible formats and the limited availability of developer time to implement those serializations, the Academic Commons Team has attempted to identify formats that will have the greatest impact, so that these limited resources can be marshalled effectively.

A False Start with OAI-PMH

Throughout its history the AC repository has striven to make its contents discoverable to the world. To this end, AC has been a registered Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) data provider since 2008. OAI-PMH, designed to provide a "low-barrier mechanism for repository interoperability," (Open Archives Initiative, n.d.) was introduced in the early 2000s. It requires data providers to expose their metadata in the simple Dublin Core schema supported by OAI-PMH at a minimum, though optionally metadata may be exposed through richer schemas as well.

However, service providers looking to harvest and aggregate the metadata made available to them have been faced with a number of problems: metadata exposed through Dublin Core only is often not rich enough to be meaningful, and the quality of the harvested metadata can vary considerably due to different content standards, encoding practices, and levels of granularity applied (Tennant, 2004). In addition, OAI-PMH is not the most efficient way to increase discoverability of repository contents through the major search engines. Google, for example, discontinued its support of OAI-PMH for sitemaps in 2008 (Mueller, 2008).

The experience of the Academic Commons Team is instructive in this regard. After migrating to Blacklight, it was discovered that the AC OAI-PMH feed was no longer updating to provide access to new content in the repository. Because AC was the first IR to use Blacklight as its primary search-and-discovery layer, the only example on which the CDRS developers could draw was a limited OAI-PMH implementation for WGBH's Open Vault (http://openvault.wgbh.org/), a multi-media archive for Boston's Public Broadcasting Service. It took several

```
45 <meta name="citation_title" content="Reagents and Strategies for the Total Synthesis of Halogenated Natural Products">
46 <meta name="citation_author" content="Daniel Scott Treitler">
47 <meta name="citation_publication_date" content="2012">
48 <meta name="citation_handle_id" content="http://hdl.handle.net/10022/AC:P:20607">
49 <meta name="citation_keywords" content="Organic chemistry">
50
51 <meta name="citation_abstract_html_url" content="http://academiccommons.columbia.edu/catalog/ac:161947">
```

Figure 1. An example of <meta/> tags embedded in the HTML header for a representative item in Academic Commons.

months of on-and-off work to modify the code from Open Vault to work with AC and, in the end, there was little noticeable impact once it was re-implemented in late 2012: the Academic Commons Team was able to update its record in the Registry of Open Access Repositories (ROAR: http://roar.eprints.org/6350/), but the change brought very little additional traffic to the repository. Indeed, the only major consumer of OAI-PMH data from AC, Scientific Commons, appears to have disappeared from the Web around the time that these upgrades were completed.

However, while the reimplementation of OAI-PMH was a bust on its own terms, the Academic Commons Team managed to deploy the knowledge gained while working on this project to create a more complete representation of its MODS metadata in Blacklight. This allowed them to explore Semantic Web technologies now available make discoverability and interoperability easier to achieve than they have been in the past. As these technologies continue to develop, AC has maintained its experimental stance, engaging in investigations to determine which serve the goals of increasing repository content discoverability and repository interoperability.

Using Microdata to Enhance Search Results

In September 2012, CDRS developers implemented a major enhancement to AC, ensuring that key metadata from the repository's index appear in <meta/> tags in the HTML header for each item in the collection (Figure 1).

These <meta/> tags are used by Google Scholar (GS) to identify descriptions of scholarly publications on the Web and to pair them with downloadable versions of the described content. On the first of October, GS staff reindexed AC. Overnight the repository went from having 1,586 items visible in GS to 4,760, bringing it to an index rate of 63%. Additionally, the visible items in GS linked to the AC item page, rather than directly to the downloadable object, as had previously been the case.

The impact on traffic to the AC site was immediately noticeable. The increase in visits in October as compared to the previous 31 days was nearly 63% (18,781 vs. 11,529). This mid-semester spike in traffic was not in keeping with past patterns: in the previous year the September to October gain was 16% and the year before there was actually an 12% drop in visitors during the same time span. Interestingly, the majority of the increased traffic in October 2012 (about 83%) came from search engines, not from GS, which was interpreted as an indication that these <meta/> tags were improving the visibility and ranking of AC content in search engine results.

Seeing the rapid and sizeable impact of exposing the AC metadata in an additional format, the Academic Commons Team investigated additional schemas for delivering metadata on the Web, with a focus of increasing search visibility and quality. Ultimately, the Academic Commons Team decided to begin with schema.org microdata (http://schema.org), to provide nested semantics for its already existing page content. Schema.org microdata enjoys broad support from popular search engines, and operates at a more general level of specificity than RDFa, making it easier to implement (Ronallo, 2012).

Microdata is particularly valuable because it can be readily embedded in HTML to provide Semantic markup: not merely increasing the visibility of Web pages, but providing structured contextual information that allows search engines (and other microdata-aware Web applications) to provide enhanced functionality. In Google search results, these microdata-enhanced features are known as "rich snippets" and are expressly designed to "help users recognize when your site is relevant to their search" (Google, 2013). Yandex, on the other hand, is rolling out a microdata-based feature they call "islands" that leverages structured HTML data to provide more robust Web services, such as booking a flight or making a doctor's appointment, directly from the search results page (Meyer, 2013). Either of these perspectives is productive in considering the use of microdata to aid in achieving AC's overarching goal of increased discoverability.

Working within the schema.org "CreativeWork" vocabulary, AC staff were able to easily map existing MODS fields to their microdata equivalents. Next they identified which fields were displayed on the landing page for individual items in the collection and modified the Blacklight code to insert the appropriate attributes into the HTML (Figure 2). After validating the initial microdata (Figure 3), the Academic Commons Team refined their implementation by matching various child vocabularies (e.g., "ScholarlyArticle") to the genres of content within AC.

```
<div itemscope itemtype="http://schema.org/CreativeWork">
        <div style="clear:both;">
          <dl class="defList clearfix">
                    <dt>
                        Title:
                    </dt>
                    <dd itemprop="name">
                            Reagents and Strategies for the Total Synthesis of Halogenated Natural Products
                        </dd>
                    <dt>
                        Author(s):
                    </dt>
                    <dd itemprop="creator">
                            <a href="/catalog?f[author_facet][]=Treitler, Daniel Scott">Treitler, Daniel Scott</a>
                        </dd>
```

Figure 2. An example of schema.org microdata embedded in the HTML for a representative item in Academic Commons.

Extracted structured data

Item

type:	http://schema.org/creativework
property:	
name:	Reagents and Strategies for the Total Synthesis of Halogenated Natural Products
creator:	Treitler, Daniel Scott
datepublished:	2012
genre:	Dissertations
url:	http://hdl.handle.net/10022/AC:P:20607
description:	Chapter 1. Introduction Natural product total synthesis has long fulfilled many roles in synthetic organic chemistry, one of the foremost being inspiration of the development of novel methods...
keywords:	Organic chemistry

Figure 3. The same item's microdata as identified by a schema.org validator.

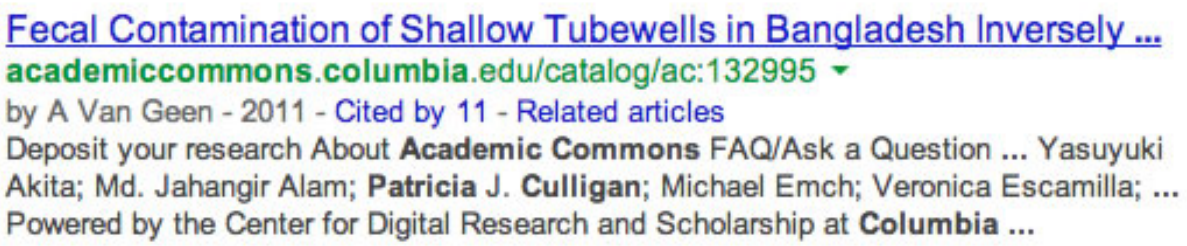

Figure 4. An article from Academic Commons as displayed in Google search. Note that it includes the first author's name, the date of publication, and links to citations and related articles in Google Scholar.

This work was completed in March 2013. Unfortunately, while Google had clearly begun using the GS indexing to enhance results in their main search (Figure 4), providing citation counts and links to related scholarly content, the additional Semantic markup provided by this microdata does not seem to have provided additional end user functionality. That said, there was a 22.6% bump in search traffic between March and April 2013, compared to a 3.7% bump in the previous month-over-month results. This seems consistent with other reports about the positive impact of schema.org microdata that indicate a typical page with this markup will rank three positions higher in search results (Silver Smith, 2013). However, seeing the enriched search results displays that microdata makes possible for content like recipes (thumbnail images, ingredient lists, ratings, etc.) and hotel Web sites (map locators, average nightly rates, ratings, etc.), there are clearly opportunities for librarians to push for better modeling for the kinds of content in our collections. Hopefully the work of the Schema Bib Extend Community Group (http://www.w3.org/community/schemabibex/) will enable these enhancements in the future.

In order to begin exploring possibilities for true cross-repository discoverability, the broader CDRS team undertook an experimental collaboration with Oregon State University's (OSU) digital repository, ScholarsArchive@OSU (SA@OSU), to implement an extension to their Blacklight application that indexed RDF metadata from OSU for use in AC. During the winter of 2013, CDRS staff coordinated with OSU's repository team to identify points of commonality in their data models and to articulate a vision for a 'related content widget': a mechanism whereby users of AC could be referred to relevant content in SA@OSU.

OSU began generating RDF metadata for their electronic theses and dissertations (ETDs) in 2012 using a data model that brings together elements from several widely-adopted schemas as well as the recently-introduced Metadata Authority Description Schema in RDF (MADS/RDF) to provide a local name authority system for authors and their advisors, as well as representing basic bibliographic metadata about the dissertations and theses themselves (Figure 5).

Because OSU made their RDF metadata openly available on the Web, despite the fact that their primary use was internal in nature, it offered CDRS developers a unique opportunity to gain familiarity with Semantic Web data while exploring a possible path to inter-repository discoverability.

The first step was to map key fields from OSU's RDF data into AC's own MODS-based schema. The SA@OSU staff also had to modify their RDF to include key additional fields, particularly the permanent URL, to allow us to link to relevant results. Finally, in order to provide a mechanism to determine the "relatedness" of a given document, the Academic Commons Team worked with the SA@OSU Team to map their departments to rough equivalents at Columbia and its affiliates. This mapping allowed us to leverage a novel feature of Solr, the MoreLikeThisHandler, which "[r]eturn[s] similar documents either based on a single document or based on posted text" (Apache Software Foundation, 2013). In our implementation, an AC user looking at the item page for a Columbia University dissertation would receive recommendations to view related dissertations in the SA@OSU collection.

This initial implementation was deployed to a development server as a proof of concept for evaluation in late January 2013. The design was kept minimal (Figure 6) to allow the focus to fall on the functionality of the related item widget: a sidebar item, providing the title for and links to the first five relevant items.

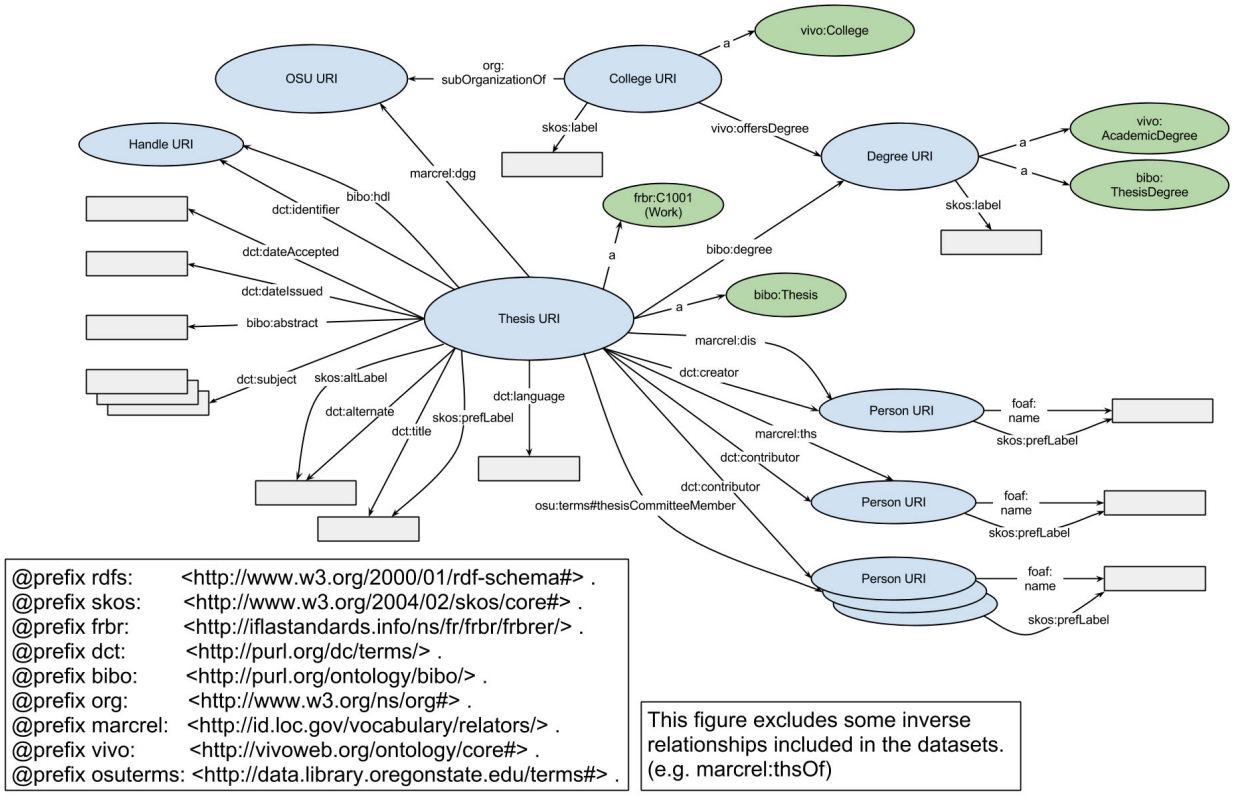

Figure 5. The Electronic Theses and Dissertations data model used by the SA@OSU (Johnson & Boock, 2012)

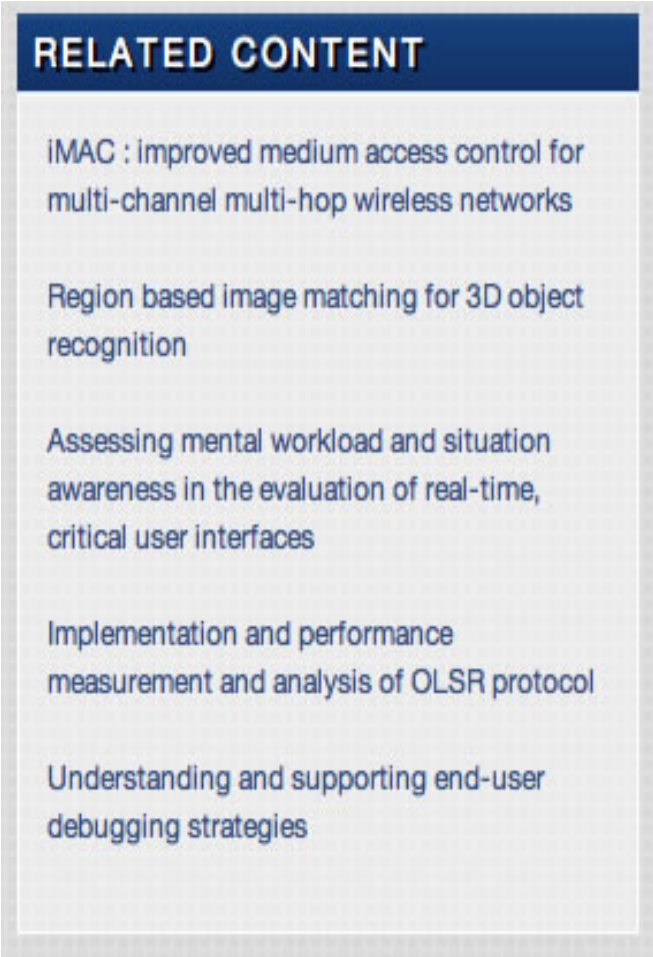

Figure 6. The initial implementation of the "Related Content Widget" in AC.

Based on early feedback, the number of items was subsequently reduced to three, with the author's names provided (with ETDs there is conventionally only a single author), and a small OSU icon next to each link to provide a visual cue distinguishing individual items.

The general reaction from library-based stakeholders on both campuses was grudgingly positive: people supported, at least in the abstract, the goal of repository cross-discoverability, and approved of the deployment of a sidebar widget as a method of providing *in situ* references to external resources, but the challenge of providing truly relevant related content in a scalable manner struck them as an overwhelming obstacle. In other words, while the user interface made sense to them, they were not convinced that the available metadata were sufficiently rich to meet user expectations.

Ultimately, the feedback suggested that one path forward is through developing the widget as a mechanism to point users to related *intra*-repository content that could potentially be integrated into the core code for Blacklight. By iteratively refining algorithmic models of relatedness based on their relative success as measured through site analytics, the Academic Commons Team could certainly develop a valuable service for their users, many of whom land directly on an individual item page after querying a commercial search engine without gaining a sense of the range of potentially useful resources in the AC collections. However, if that model relies too heavily on specialized features of the AC MODS implementation then it would become unworkable as a mechanism for repository interoperability, since it would have to be able to work with varied types of metadata of varying quality and consistency.

The key, it would seem, is to leverage the flexibility and scalability of Solr itself, which can accommodate many millions of records, to create a richer index of the content available in a broader range of institutional repositories. CUL/IS developers from outside CDRS have already tackled aggregated discovery platforms of a similar scale and complexity using Blacklight and Solr: for example, the Human Rights Web Archive (http://hrwa.cul.columbia.edu) contains the full text of more than 50 million Web pages and other documents in its Solr index and it still provides sub-second search response times (Columbia University Libraries, 2013). By working with key repository partners who, like SA@OSU, are taking a proactive approach to enhancing their metadata, AC could position itself as a leader in interoperable repositories and blaze a path that others could follow.

Exploring MODS RDF

After the successful, if limited, collaboration with OSU, the Academic Commons Team was eager to utilize Semantic Web and Linked Data technologies to solve a number of issues within AC, such as authority control of author and department names. At the same time, the team aimed to convert the repository's metadata to RDF to lay the ground for cross-repository interoperability. One solution meeting both of these goals is MODS RDF.

MODS, as mentioned above, is the preferred metadata schema for the majority of CUL/IS digital projects as well as for the institutional repository. Since MODS is derived from the MARC 21 standard, it allows for repurposing of MARC bibliographic data from the main CUL/IS catalog. It is quite compatible with other schemas as well, thus allowing CUL/IS staff to map legacy data from a variety of schemas, many of them locally developed, into MODS for ingest into the repository. The relatedItem element in MODS allows for nesting the descriptions of related items such as host items or series into the main record. The entire set of elements can be applied here as well, thus permitting very detailed descriptions of these related resources. MODS can be extended to include elements from other schemas if necessary. For AC, MODS has proven to be a flexible solution since it is easily adaptable to describe various resource types, particularly as the collection grows to include an increasing number of datasets, videos, and other non-bibliographic materials.

As documented above, the Academic Commons Team has worked hard to ensure that this high quality metadata is

fully represented in the repository's Solr index and its public-facing Web site, and has been rewarded with sizeable increases in Web traffic, particularly from popular search engines. However, because the repository's faceted search-and-discovery interface relies on uniform string values, author and department name control are both a necessity and an ongoing challenge, particularly since so many authors of repository content are not represented in traditional name authority systems.

The release of the draft MODS RDF Ontology (Library of Congress, 2012) in February 2013 presented itself as a welcome opportunity to move forward with the planned conversion of AC MODS records to RDF to help manage some of the issues around name control. MODS RDF is based on MODS XML thereby ensuring its compatibility with existing MODS records. It draws heavily on MADS RDF by expressing a number of MODS RDF properties using the MADS RDF ontology (Library of Congress, 2013).

The Team decided to experiment with the draft MODS RDF ontology thereby contributing to its advancement into a published standard rather than investing into the development of yet another local solution. They began by examining the available documentation. The MODS RDF Ontology Website includes the MODS RDF Namespace document, a MODS RDF Ontology Primer, an additional primer containing information for MODS XML to RDF data conversion, examples, as well as a stylesheet.

As a next step, an AC MODS to MODS RDF mapping table was created. Using this mapping and the stylesheet made available by the MODS/MADS Editorial Committee the Academic Commons Team made an initial, but unsuccessful attempt to convert several sample AC MODS records to MODS RDF. It was necessary to make some adjustments to the existing stylesheet to account for some AC specific metadata characteristics, such as empty elements contained in the data, before the conversion of a test record set was successful. At this point, the Academic Commons Team is continuing to work with the MODS/MADS Editorial committee in testing and developing the MODS RDF ontology, and to ease some of these issues.

OUTLOOK

As noted by Manola & Miller (2004), the ultimate success of semantic technologies employing RDFs "depends on increasing the general use of URIrefs to refer to things instead of using literals". This, along with the ever-growing collection of duplicative standards and namespaces that also limit interoperability, are the great challenges to our community as we continue to push forward to successful implementations of Semantic Web standards in our libraries and repositories. The price of failure is equal to the lost opportunity costs experienced by current and future members of our communities, whether information seekers or providers, who are stuck in non-interoperable silos of semantic metadata.

Currently, the outlook for MODS RDF is tenuous, given its early reliance on literal values which limit interoperability and restrict the potential of semantically linked data. The recent commitment by the MODS/MADS Editorial Committee to enhance the MODS RDF stylesheet so that XLink or authorityURI resources from the MODS XML record are reflected in the RDF (R. Denenberg, personal communication, June 9, 2013) begins to create more confidence that MODS RDF will be capable of supporting cross-repository interoperability. Continuing to build in more facilities for incorporating semantically meaningful data linkages will contribute to making MODS RDF a robust solution for increasing repository interoperability and discoverability.

The promise of MODS RDF is that it provides a standardized ontology that allows easily implementable crosswalks between existing records, preparing repositories to take the leap into broader interoperability. The inherent standardization of MODS and its wide usage also provides a low barrier to linking between MODS elements and other vocabularies. These provisions permit repositories to rapidly enhance descriptions of their collections by allowing consumption of Linked Open Data (LOD) from sources such as Virtual International Authority Files (VIAF, http://viaf.org/) or DBpedia (http://dbpedia.org/About). As the Academic Commons Team's collaboration with SA@OSU demonstrated, this LOD can readily be indexed in Solr, thereby enhancing on-site search.

Just as implementing MODS RDF enables enhanced collection description and increases the rapidity with which collections may be described, having the ability to enrich the MODS metadata in the triple store accelerates the speed with which microdata may be deployed. This deployment in turn impacts the Web applications where these data are displayed and enhances the interoperability of repositories by enabling repository linking through embedding microdata in the Web application interface. Using embedded microdata allows repositories to spend their efforts in refining the algorithms that find, select, and link to non-local repository resources that a user may find useful, rather than employing time and effort in building the ultimate collection of all possible resources.

In order to continue pursuing the opportunities offered by Semantic Web technologies, the CDRS programming staff has recently installed 4Store (http://4store.org/) to serve as a scalable RDF database. The hope of the Academic Commons Team is that this open source platform will provide a foundation for the continued effort to convert the current MODS metadata to MODS RDF triples as well as enabling for a more robust approach to name authority work for authors and departments (Figure 7).

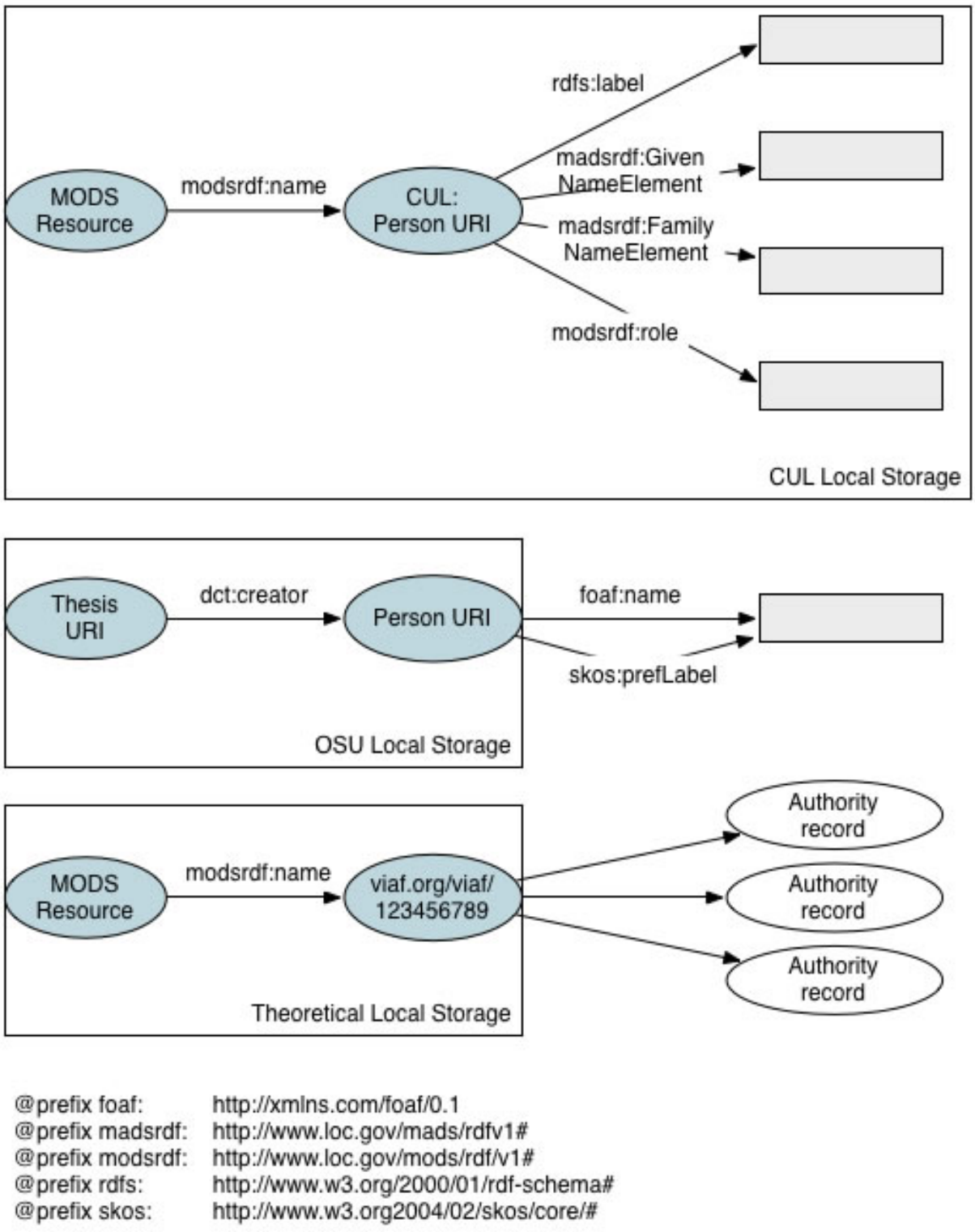

Figure 7. Three ways to encode author name information in RDF. Note that the third example relies on the Virtual International Authority File (VIAF) URI to point to authority data from multiple national libraries, where available

.

It will also equip AC to take advantage of new Semantic Web technologies as they become available, as well as emerging unique identifier platforms, such as the Open Research and Collaborator ID (ORCID: http://orcid.org) and the International Standard Name Identifier (ISNI: http://www.isni.org).

Ultimately, employing these technologies will make it possible for AC to serve not just the visitors it currently enjoys by revealing more resources to them, but also to serve searchers of other repositories by creating the interoperable connections that assist them in breaking out of their current information silos and allow them to find the information they seek. This enhanced discoverability of AC content through interoperability furthers the goals of both Academic Commons and Columbia University in "convey[ing] the products of [Columbia's scholarly] efforts to the world" (Columbia University, 2013).

RECOMMENDED CITATION:

Hilliker, R. J., Wacker, M., & Nurnberger, A. L. (2013). Improving Discovery of and Access to Digital Repository Contents Using Semantic Web Standards: Columbia University's Academic Commons. Journal of Library Metadata, 13(2-3), 80–94. doi:10.1080/19386389.2013.826036

REFERENCES

Alhabashneh, O., Iqbal, R., Shah, N., Amin, S. & James, A. (2011). Towards the development of an integrated framework for enhancing enterprise search using latent semantic indexing. In S. Andrews, S. Polovina, R. Hill & B. Akhgar (Eds.), *Conceptual Structures for Discovering Knowledge: Vol. 6828. Lecture Notes in Computer Science* (pp. 346-352). Berlin: Springer. Available from http://dx.doi.org/10.1007/978-3-642-22688-5_29

Apache Software Foundation. (2013). *MoreLikeThisHandler (Solr 4.3.0 API)*. Retrieved from http://lucene.apache.org/solr/4_3_0/solr-core/org/apache/solr/handler/MoreLikeThisHandler.html

Bufanio, N. (2011, April 18). *Springing ahead with Academic Commons*. [Web log post]. Retrieved from http://cdrs.columbia.edu/cdrsmain/2011/04/springing-ahead-with-academic-commons/

Columbia University. (2013). *Mission statement*. Retrieved from http://www.columbia.edu/content/mission-statement.html

Columbia University Libraries. (2013, February 14). *Columbia University Libraries releases Human Rights Web Archive*. Retrieved from http://library.columbia.edu/news/libraries/2013/2013-2-14_Human_Rights_Web_Archive.html

Glaser, H., Millard I. C. & Jaffi, A. (2008). RKBExplorer: A knowledge driven infrastructure for linked data providers. In S. Bechhofer, M. Hauswirth, J. Hoffmann & M. Koubarakis (Eds.), *The semantic web: Research and applications* (pp. 797-801). Berlin: Springer. Available from doi:10.1007/978-3-540-68234-9_61

Google. (2013, June 5). About rich snippets and structured data: Rich snippets (microdata, microformats, RDFa, and Data Highlighter). *Google Webmaster Tools*. Retrieved from http://support.google.com/webmasters/bin/answer.py?hl=en&answer=99170

Johnson, T. & Boock, M. (2012). Linked data services for thesis and dissertations. In *Proceedings of the 15th International Symposium on Electronic Theses and Dissertations, Lima, Peru, 12-14 September 2012*. Retrieved from http://etd2012.unmsm.edu.pe/pdf/presentation/TJohnsonETD2012FT.pdf

Koutsomitropoulos, D. A., Solomou, G. D., Alexopoulos, A. D. & Papatheodorou, T. S. (2009a). Knowledge management and acquisition in digital repositories--a semantic web perspective. In K. Liu (Ed.), *KMIS 2009: Proceedings of the International Conference on Knowledge Management and Information Sharing, Funchal - Madeira, Portugal, October 6-8, 2009*, (pp. 117-122). Portugal: INSTICC Press. Retrieved from http://www.hpclab.ceid.upatras.gr/viografika/kotsomit/pubs/kmis09.pdf

Koutsomitropoulos, D. A., Solomou, G. D., Alexopoulos, A. D. & Papatheodorou, T. S. (2009b). Semantic Web enabled digital repositories. *International Journal on Digital Libraries, 10*, 179-199. Retrieved from doi:10.1007/s00799-010-0059-z

Library of Congress. (2012, September 28). *MODS RDF Ontology*. Retrieved from http://www.loc.gov/standards/mods/modsrdf/
Library of Congress. (2013, June 7). *MODS RDF Ontology: Primer.* Retrieved from http://www.loc.gov/standards/mods/modsrdf/primer.html

Manola, F. & Miller, E. (Eds.). (2004). *RDF Primer: W3C Recommendation 10 February 2004*. Retrieved from http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

Meyer, D. (2013, June 5). Think Google's rich snippets are useful? Russia's Yandex goes one better. *GigaOM*. Retrieved from http://gigaom.com/2013/06/05/think-googles-rich-snippets-are-useful-russias-yandex-goes-one-better/

Moore, K. B. & Greene, C. (2012). The search for a new OPAC: Selecting an open source discovery layer. *Serials Review, 38,* 24-30. Available from http://dx.doi.org/10.1016/j.serrev.2011.12.005

Mueller, J. (2008, April 18). Retiring support for OAI-PMH in sitemaps. [Web log post]. Retrieved from http://googlewebmastercentral.blogspot.com/2008/04/retiring-support-for-oai-pmh-in.html

Open Archives Initiative (n.d.). *Open Archives Initiative Protocol for Metadata Harvesting*. Retrieved from http://www.openarchives.org/pmh/

Ronallo, J. (2012). HTML5 Microdata and Schema.org. *Code{4}Lib Journal* (16). Retrieved from http://journal.code4lib.org/articles/6400

Sadler, E. (2009). Project Blacklight: a next generation library catalog at a first generation university.

Library Hi Tech, 27.1, 57-67. Available from http://dx.doi.org/10.1108/07378830910942919

Silver Smith, C. (2013, June 18). From Microdata & Schema to rich snippets: Markup for the advanced SEO. [Web log post]. Retrieved from http://searchengineland.com/from-microdata-schema-to-rich-snippets-markup-for-the-advanced-seo-162902

Subirats, I., Malapela, T., Dister, S., Zeng, M., Gooaverts, M., Pesce, V., et al. (2012). Reorienting open repositories to the challenges of the semantic web: Experiences from FAO's contribution to the resource processing and discovery cycle in repositories in the agricultural domain. In J. M. Dodero, M. Palomo-Duarte, P. Karampiperis (Eds.), *Metadata and semantic research: 6th Research Conference, MTSR 2012 Cádiz, Spain, November 28-30, 2012 Proceedings* (pp. 158-167). Heidelberg: Springer. Available from doi:10.1007/978-3-642-35233-1_17

Tennant, R. (2004, May 14). *Bitter harvest: problems & suggested solutions for OAI-PMH data & service providers*. Retrieved from http://roytennant.com/bitter_harvest.html