Extending Legacy Metadata with Linked Open Data

Jacob Jett
University of Illinois at Urbana-Champaign, USA
jjett2@illinois.edu

Timothy W. Cole
University of Illinois at Urbana-Champaign, USA
t-cole3@illinois.edu

Alex Kinnaman
University of Illinois at Urbana-Champaign, USA
kinnama2@illinois.edu

Deren Kudeki
University of Illinois at Urbana-Champaign, USA
dkudeki@illinois.edu

Myung-Ja (MJ) K. Han
University of Illinois at Urbana-Champaign, USA
mhan3@illinois.edu

Caroline Szylowicz
University of Illinois at Urbana-Champaign, USA
szylowic@illinois.edu

Abstract

Library special collections are valued by scholars and relied on to support both research and teaching. In recent years, libraries have invested heavily in digitizing many of these collections. Unfortunately, less effort and fewer resources have been expended post-digitization and many digitized library special collections today exist on the Web only in isolated information silos, difficult to find and disconnected from other resources that could provide users with valuable context. This begs the question: Can Linked Open Data (LOD) approaches be leveraged to help contextualize & enrich item-level descriptions of such collections and provide links to related information resources? This project report describes preliminary results from *Exploring the Benefits for Users of LOD for Digitized Special Collections*, a project still in progress which is examining this and related questions. Among the findings reported here: while special collections metadata are typically rich and ripe with LOD potential, the idiosyncratic nature of the collections and the metadata schemes used pose unique mapping and transformation challenges; the opportunity for adding links to item-level metadata is great, but finding links still requires significant cataloger involvement; at the scale of most library special collections, information from LOD sources can be retrieved in real time to enhance the presentation of items to end-users, providing context and links to related information. These findings suggest that the transformation of metadata into LOD and the inclusion in item descriptions of links can improve the connectedness of digitized special collections and enhance user interactions with these resources.

Keywords: metadata enhancement; linked open data; metadata mapping; schema.org

1. Introduction

After more than twenty years of digitizing special collections materials, libraries and other cultural heritage institutions have amassed a wealth of unique digital objects that span a broad spectrum of disciplines. As surrogates for primary sources and the products of scholarly research, the items in digitized special collections have significant value for further scholarship and pedagogy. However, studies suggest that, "Libraries are spending far more to create new [digital] resources than they are on maintaining and enhancing the ones they have already created" (Maron and Pickle 2013, p. 2). While the products of special collection digitization are often freely available on the Web, this lack of post-digitization investment and an overreliance on legacy descriptive metadata schemas and single collection finding aids designed originally for describing print resources has relegated many digitized special collections to information silos largely disconnected from the broader Web. This makes such resources difficult to discover and isolates them from potentially useful context. The need in the abstract for a more connected, shared cultural

heritage information space and to develop new methods of collection management and curation to overcome such 'siloization' tendencies has been recognized for several years, cf. the coordinated special issues of *Museum Management and Curation*, *Archival Science*, and *Library Quarterly* addressing this and related topics published at the end of the last decade (Marty 2008). Clearly there is a need to better connect digitized special collections to the broader Web and move away from information silos, but practical approaches to do this are still emerging and still being tested and evaluated. The outstanding question with regards to digitized special collections resources is what can be done after digitization to make these resources more discoverable by and more useful to users? Linking them to additional, related resources through Linked Open Data (LOD) (Burners-Lee 2006, Auer et al. 2007) is one potential value-added step that could be taken. However, our understanding of how LOD benefits digital libraries and their users is still limited (Corbet 2016).

Funded by the Andrew W. Mellon Foundation, the *Exploring the Benefits for Users of LOD for Digitized Special Collections*[1] project at the University of Illinois at Urbana-Champaign has been investigating LOD's potential to benefit both the stewards and users of digitized special collections. Two digitized theater collections (The Motley Collection of Theatre and Costume Design[2] and the Portraits of Actors, 1720-1920[3]) and a collection of digitized researcher notes encoded using TEI (the Kolb-Proust Archive for Research[4]) provide the foundation for the research project. This Project Report describes some of the unique challenges presented by special collections metadata and illustrates ways that LOD resources can be used to expand information presented to users and connect users to additional context external to the collection in order to make the resources contained in digitized special collections more useful for research and instruction.

2. Idiosyncratic data

One barrier to optimizing the utility of digitized special collections is the often imprecise and idiosyncratic nature of metadata records describing a collection's digitized objects or intellectual content. For instance, though the metadata schemes developed for the Motley and the Portraits of Actors collections were both based on simple Dublin Core, several extensions, re-interpretations of properties and classes, and other local scheme modifications were made when the objects were digitized and metadata initially transformed. Jett et al. (2016, 2017) found that besides conflating the objects being digitized with their digital representations (Woodley et al. 2005, Park & Childress 2009, Urban 2012), special collections metadata records often conflate descriptions of multiple individual entities and entities of different classes.

An import first-step towards ameliorating these problems is to choose a good linked-data-compliant vocabulary. At the time the project began Bibframe 2.0 was still under development and despite the existence of FRBR-oriented linked-data vocabularies—specifically $FRBR_{OO}$[5] and the SPAR family of ontologies[6]--schema.org[7] was selected for the project's mapping vocabulary of choice. The reason for this selection was entirely pragmatic: at the time the University library was experimenting with workflows for transforming much-richer MARC records into schema.org compliant metadata records, as was OCLC through their WorldCat Linked Data Vocabulary.[8]

2.1 Mapping Motley Image Metadata

For instance, in the Motley Collection exemplar case (Jett et al. 2016, p 3), metadata records describing costume sketches for a particular production of a play (e.g., as shown in Figure 1) were

[1] http://publish.illinois.edu/linkedspcollections/
[2] http://imagesearch-test1.library.illinois.edu/cdm/landingpage/collection/motley-new
[3] http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/actors
[4] http://www.library.illinois.edu/kolbp/
[5] https://www.ifla.org/free-tags/object-oriented-frbr
[6] http://www.sparontologies.net/ontologies
[7] http://schema.org
[8] https://www.oclc.org/developer/develop/linked-data/worldcat-vocabulary.en.html

found to also contain metadata that described: a specific performance of the production (usually its opening night performance), the performance venue (theater), and the play itself (as a textual object). Persons associated with the production (e.g., producer, director, actor) were concatenated in a single field, individually identifiable by name and role only through punctuation and string-embedded labeling. Persons associated with the play as a written work (e.g., author, composer) were similarly concatenated in another metadata field.

Carefully mapping the Motley collection's legacy metadata model into the schema.org vocabulary (and any other linked-data-compliant vocabulary) begins by identifying these 1-to-1 violations in the existing metadata records. The resulting metadata records can then be divided into multi-part accounts that provide information particular to the different individual entities. Distinguishing the entities described is an essential first step in mapping legacy metadata into schema.org semantics and useful LOD-compatible descriptions. However, simply isolating and mapping the various entities described in the legacy metadata is not sufficient to make the metadata account optimal when displaying descriptions to users of these collections. The snippets of legacy metadata describing these additional entities must be enriched through the addition of links to external LOD services and other external resources (as described in Section 3 below).

FIG 1. Digitized costume sketch from Motley Collection with original metadata.

Since both the Motley collection and the Portraits of Actors collection are fairly traditional image special collections, the initial mapping of metadata records into Schema.org's vocabulary was relatively straightforward and should be adaptable for similar image collections. The primary class chosen to represent the collection's objects was that of schema:VisualArtwork. These objects (i.e., each sketch or photograph) were then disambiguated from the digitized image that represents them in online environments through use of the schema:associatedMedia relationship; in this way the schema:MediaObject (typically a schema:ImageObject) is distinguished in the LOD account

from the original hard copy Artwork (sketch or photograph). The play, production, venue and person entities are then isolated and given their own accounts in the LOD description through the use of the schema:isPartOf, schema:exampleOfWork, schema:locationCreated, schema:creator, and schema:contributor, relationships. This allows the LOD account to provide a series of true assertions, e.g., that the Visual Artwork is part of the Stage Work (the play's production), which in turn, is an example of a Book (the play) or Creative Work (when the play has not been formally published). A full account of the current mapping appears in Jett et al. (2016), though as the project is still ongoing, some further refinement of the mapping is anticipated.

2.2 Mapping Kolb-Proust Archive Data

In contrast to the two theater-focused collections, the digital objects in the Kolb-Proust Archive for Research (KPA) required a different approach. The KPA contains TEI-encoded transcriptions of two of Proust scholar Philip Kolb's research collections—*Bibliographie* and *Chronologie*. Each digital transcript provides the text that appears on a corresponding hardcopy notecard in these collections (see Figure 2). Each notecard in turn represents a single cohesive note made by Kolb in regard to an event or publication mentioned or alluded to in Proust's correspondence or of interest in regards to Proust's literary career or personal history. As the original editor of Proust's correspondence (spanning 21 volumes published between 1970 and 1993), Kolb's notes enabled him to sequence Proust's correspondence (most of which was not dated) and recognize the identity of people and events mentioned in his letters. These notecards provide potentially useful linkages between people, events and contemporaneous accounts of their interactions (e.g., in newspapers and other publications). In this sense, the cards are a sort of precursor to linked data. However, at first glance these transcriptions appear less well-aligned with the Schema.org ontology. Additionally, to express a wide range of complex relationships, succinct natural language was used on these cards; the transcription into TEI makes explicit (in a computational sense) only some of these relationships. To undertake a mapping of these archival collections, it was necessary to begin by carefully considering exactly what the transcriptions represent and contain vis-à-vis the classes of Schema.org. Person entities are prominently represented. Many notecards contain citations, which lend themselves to traditional metadata descriptive practices. However, capturing these names and citation in isolation loses the valuable context of the card on which they appear.

FIG. 2. A card from the KPA and its TEI transcription.

In this case the recognition of how to proceed finally arrived when we realized that each of Prof. Kolb's transcribed notecards could be represented using the schema:DataSet class which is a sub-class of schema:CreativeWork. This approach allowed us to accomplish three important things:

1. Preserve the overall context provided by each transcribed card (as just explained);
2. Use schema:citation to relate bibliographic entities to the card on which each appears;
3. Use schema:mentions to relate names and titles appearing elsewhere on cards.

Bibliographic works cited on cards were mapped to schema:CreativeWork or a subclass. We tried to use the appropriate level of granularity (i.e., schema:Article, schema:PublicationIssue, schema:Book, schema:PublicationVolume) whenever possible. But as in our earlier mappings (Jett et al. 2016, Han et al. 2017), it became clear that Schema could benefit from extensions that name other fine-grained classes for such bibliographic entities as "essays", "poems", and "short stories." In addition, it became clear that our earlier proposed extension for Schema, StageWork (Jett et al. 2016, p 3) would clearly benefit from its own set of finer-grained sub-classes that can better distinguish when a "Stage Work" is a "Play" versus when it is a "Variety Show" versus when it is an "Opera" and so forth.

3. Enriching Metadata with Links

A necessary step in order to migrate descriptions of items contained in digitized special collections from text-based metadata record formats to LOD-compatible serializations (of RDF graphs) is to identify and provide canonical URIs for entities described in the metadata being migrated. A common shorthand for this process is 'strings into URIs.' Note, a single entity, whether tangible (e.g., a person) or conceptual (e.g., a work) may be identified by multiple URIs. In the case of our project, the strings into URIs task has been completed for the Motley collection, and is still ongoing for the Portraits of Actors and KPA collections. For Motley the strings into URIs reconciliation process focused on people (author, composer, actor, director and other associated people), venues (theaters), subject headings (LCSH, TGMI and AAT), and the play / performance. As a precursor to searching for URIs some limited semi-automated metadata remediation was done, primarily de-conflating names and subject headings based on punctuation and formatting rules followed. As described below and once remediation was complete, we relied primarily on manual searching for links. Altogether 240 person-hours were spent on remediation and finding links for entities referenced in Motley metadata.

3.1 Named Entities (Persons) Identification & Reconciliation

URIs for persons mentioned in Motley metadata were sought using both automated and manual processes. This was done in order to ascertain which process worked better, to make an initial assessment of the labor required for the manual process and to help determine the degree to which the two approaches were complementary. Currently two of the most important LOD-compatible sources for person entity URIs are the Virtual International Authority File (VIAF)[9] and DBPedia[10]. Wikidata[11] is also becoming an important part of the Web's Linked Data landscape but at the time the project began we decided to focus our limited resources on VIAF which links to OCLC and other authority resources and DBpedia which is closely associated with Wikipedia. Both are related to other resources. DBPedia extracts structured content from the information found in Wikipedia[12]; Wikipedia URLs can be deduced from examination of DBPedia URIs and vice versa. VIAF draws on (and includes links to) many national catalogs and other library authorities, including the Library of Congress Name Authority File (LCNAF)[13], with which WorldCat Identities[14] also is closely synchronized. VIAF provides an interface for manually searching names, but we had better success (see Table 1) when manually searching using LCNAF. Manually searching for DBPedia URIs is best done by searching Wikipedia. Note that VIAF and Wikipedia have now been largely (but not

[9] http://viaf.org/
[10] https://dbpedia.org/
[11] https://www.wikidata.org/wiki/Wikidata:Main_Page
[12] https://www.wikipedia.org/
[13] http://id.loc.gov/authorities/names.html
[14] http://worldcat.org/identities/

fully) reconciled. Typically VIAF descriptions of persons include Wikipedia links when available and vice versa. For the names that we searched, we found that 10% of the Wikipedia and VIAF descriptions did not link to one another.

As shown in Table 1, manual searching for VIAF links yielded URIs for 218 of the 984 persons mentioned in the Motley metadata. Only 87 were found by searching VIAF directly, the rest were found by manually searching LCNAF and WorldCat Identities. By comparison, using the VIAF Auto Suggest API yielded URIs for 476 names, more than twice as many as found manually, but importantly, automatic searching in this way missed 106 URIs found by manual searching. Not surprisingly given that many individuals mentioned in Motley metadata were performers and directors rather than authors, we found Wikipedia / DBPedia URIs for 311 out of the same 984 persons; only some of which overlapped with VIAF URIs. We did limited experimentation with automating the searching of Motley names in DBPedia / Wikipedia, but in the absence of an API equivalent to the VIAF Auto Suggest API, we did not see much improvement over manual searching which was done first. For the Portraits of Actors Collection, automated searching, including of Wikipedia, was done first, which meant searching a smaller number of names manually. We expect to report quantitative results of this work soon. Our preliminary recommendation for similar collections is VIAF Auto Suggest API first, manual searching of LCNAF (limited to names not found using VIAF Auto Suggest), and lastly manual search of Wikipedia / DBPedia for names not found in the first two steps.

TABLE 1. Count of person URIs found through manual searching

Total persons identified in Motley metadata = 984 *Links have been found for 624 names*	**Count of URIs Found**
having Wikipedia / DBPedia links	311. (32%)
having VIAF links	218.* (22%)
found by searching viaf.org directly	87.
found by searching LC Name Authority File	196.
found by searching WorldCat Identities	93.
having Theatricalia links	475. (48%)
having IMDb links	353. (36%)
having IBDb links	42. (4%)
having more than 1 link	446. (45%)

* VIAF links for 476 persons (364 not found by manual search) were found using VIAF Auto Suggest

While both VIAF and DBPedia provide RDF descriptions of person entities, their scope is limited and their information about individuals is not complete. The scope of VIAF, for example, is biased toward authors of book-length publications. Since many of the individuals mentioned in the Motley metadata were actors, directors, producers, set designers, etc., we also manually searched several theater-related, non-LOD Web resources, specifically Theatricalia[15], IMDb[16] (the Internet Movie Database), and IBDb[17] (the Internet Broadway Database). While links (URLs) to these Web resources do not return RDF when de-referenced, they are a way for users to find more context about an individual. Because these resources focus on persons involved in theater and movie work, we found higher percentage of the names mentioned in Motley metadata in Theatricalia (58%) than in any other source, i.e., more than in either VIAF or Wikipedia. In addition to the ability to link with and acquire (albeit manually) metadata from these domain-pertinent resources provided additional, contextually important information. Theatricalia links allowed us, for example, to capture information regarding more specific roles that "Associated People" played with regards to a play's performance, production, or publication.

[15] https://theatricalia.com/
[16] http://www.imdb.com/
[17] https://www.ibdb.com/

3.2 Finding Links for Other Entities

VIAF and DBPedia / Wikipedia also proved useful sources for URIs for theaters (venue entities), while Theatricalia and DBPedia / Wikipedia were the best sources for URIs and URLs for performances and plays. Many theaters also have Web home pages that were easy to find through Google Search. Tables 2 and 3 give the quantitative results of manual searching to find URIs for theaters, performances and plays. The total number of entities involved was much smaller, i.e., although the Motley collection includes more than 4,700 set and costume design sketches, they are drawn from only 127 plays. Note that for this initial pass at adding URIs to the Motley metadata we conflated plays and performances. Based on feedback we may revisit this decision in the future.

TABLE 2. Count of theater URIs found through manual searching

Total theaters identified in Motley metadata = 59 *Links were found for 52 theaters*	**Count of URIs Found**
having Wikipedia / DBPedia links	49 (83%)
having VIAF links	45 (76%)
having home page links	36 (61%)
having other links	16 (27%)
having more than 1 link	47 (80%)

TABLE 3. Count of play / performance URIs found through manual searching

Total plays / performances identified in Motley metadata = 127 *Links were found for 105 plays / performances*	**Count of URIs Found**
having Wikipedia / DBPedia links	95 (75%)
having Theatricalia links	45 (35%)
having other links	10 (8%)
having more than 1 link	44 (35%)

4. Using LOD to Enhance Discovery and User Interaction with Collections

Transforming the structure of item-level metadata to better align with RDF and LOD best practices makes it easier for LOD-based applications to consume these resource descriptions; in theory, this should facilitate discovery. Similarly, including links to related Web resources and to LOD-based repositories containing additional information provides an opportunity to enhance user interaction with digitized special collections. While it is too early on this project to assess whether discovery of Motley resources (LOD-compatible metadata was only added to resource pages in late spring), work with the Google Structured Data Testing Tool (Figure 3) is encouraging.

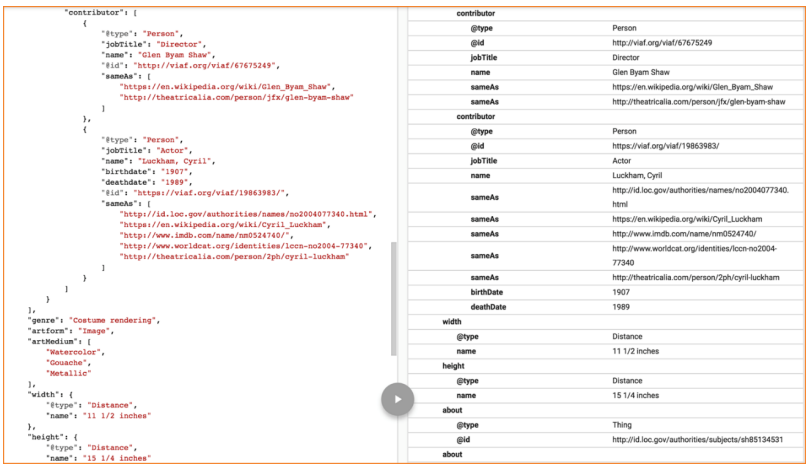

FIG. 3. Snippet of item-level RDF serialized in JSON-LD and viewed in Google Structured Data Testing Tool

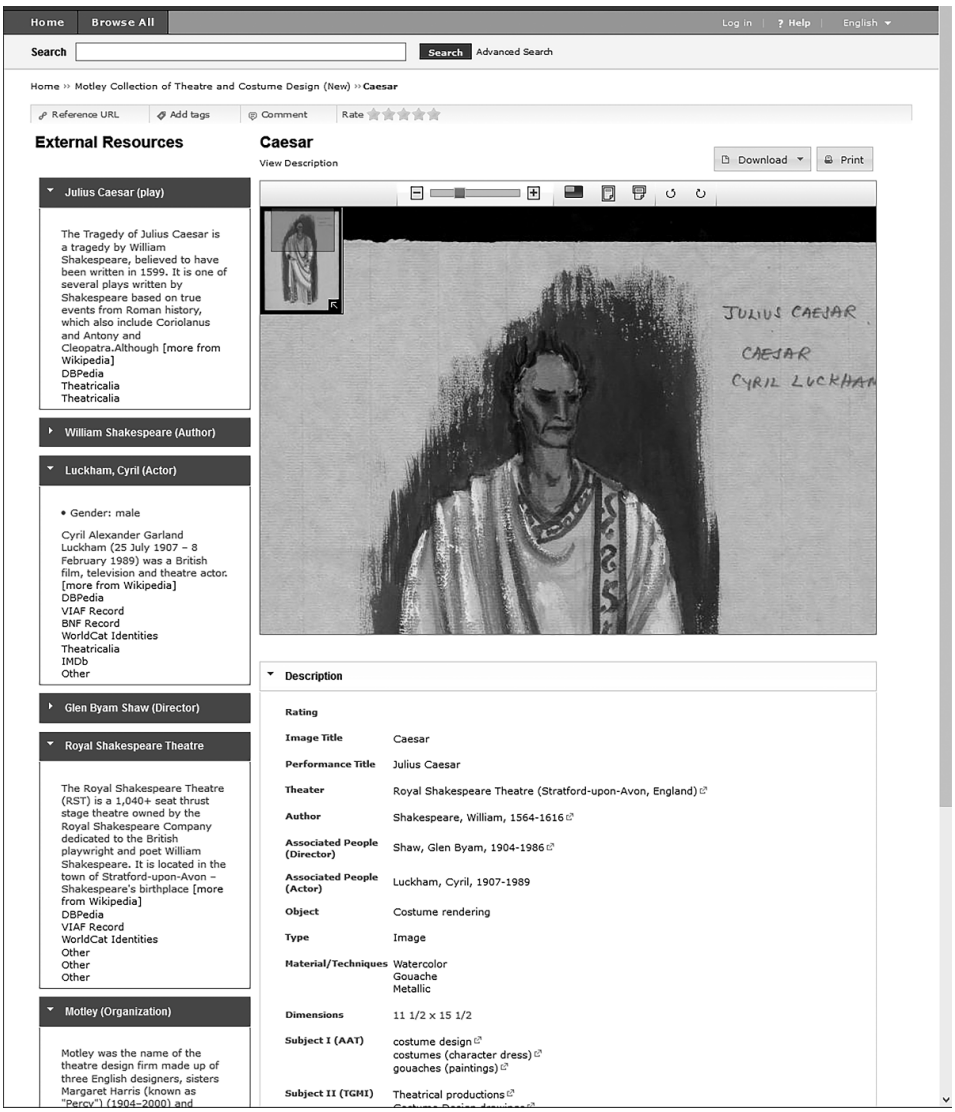

FIG. 4. Dynamically generated side-bar displaying links and context retrieved real-time from LOD services

To assess the viability and potential utility of the links added to Motley item metadata we took as inspiration the knowledge card approach increasingly being adopted by search engines like Google. Figure 4 showcases this approach as implemented. Here descriptions and links pertaining to the various entities mentioned in item-level RDF metadata (e.g., plays / productions, theaters, authors, directors, actors, etc.) are retrieved from LOD repositories (VIAF, DBpedia) and presented in boxes (1 per entity) along the left side of the browser window. The sidebar is produced using client-side JavaScript. The script, which runs on completion of the page load, parses RDF (serialized as JSON-LD and embedded in the Web page in a script element with a type attribute value of 'application/ld+json') to find URIs. These are then used to harvest links and selected metadata from LOD repositories (e.g., retrieving from DBpedia's page for the Royal Shakespeare Theatre[18] the value of the abstract[19]). URIs can be de-referenced directly if the LOD service implements the appropriate CORS (Cross-Origin Resource Sharing) headers or through proxies on our own server. Proxies offer the opportunity to implement short-term (e.g., 24 hour) caching to improve performance. The JavaScript also updates the default CONTENTdm metadata

[18] http://dbpedia.org/page/Royal_Shakespeare_Theatre
[19] http://dbpedia.org/ontology/:abstract

presentation, adding external links for entities described in the sidebar as well as for Getty[20] and Library of Congress[21] subjects.

The JavaScript to do this is relatively brief and uncomplicated, so to this point we have not needed to deploy a full-fledged Model-View-Controller (MVC) JavaScript framework (e.g., Angular.js, Backbone.js, Spine.js, to mention just a few), though our scripting does have JQuery.js dependencies. Nonetheless we have taken a model-view scripting approach in anticipation that we may wish to adopt a MVC framework library in the future. JavaScript object prototypes (illustrated in Figure 5) were defined for each class of entity in our RDF (person, theater, play / production) as our data model layer. Mustache.js templating was then used as our nascent view layer to manage the translation of these JSON data models into HTML.

```
function PersonEntity(id, identifier, homePage, longAbstract, shortAbstract, name, enWikiUrl, ViafUrl,
jobTitle, gender, birthDate, deathDate){
        this.id = id;
        this.identifier = identifier;
        this.homePage = homePage;
        this.longAbstract = longAbstract;
        this.shortAbstract = shortAbstract;
        this.name = name;
        this.enWikiUrl = enWikiUrl;
        this.ViafUrl = ViafUrl;
        this.links = new AutomapperConfig();
        this.type = 'Person';
        this.jobTitle = ( jobTitle ? jobTitle : 'Actor' );
        this.gender = gender;
        this.birthDate = birthDate;
        this.deathDate = deathDate;
    return;
    }
```

FIG. 5. The JavaScript prototype (data model) for person entities

A key benefit of the LOD-based architecture and implementation described above is that user interaction with item descriptions can be enriched without the need to redundantly store metadata values not required by the local system for search and discovery. This reduces risks of metadata staleness. A question remains as to what kinds of links and just-in-time contextual information at point of use would be most useful to users of the Motley collection. While we do not have scope within the current project to answer this question definitively, a small sample user test of the new interface illustrated in Figure 4 is currently underway. Results will be compared with those from a similar test conducted last year with the former user interface design illustrated in Figure 1. Note that user interface improvements between these tests was limited to changes stemming from the inclusion of LOD.

5. Conclusions & Future Work

This project report showcases one manner in which resources included in digitized special collections can be made potentially more useful through the addition of contextual information and links. As browser-based and mobile search engines become better at seamlessly integrating RDF-based, LO-compatible item descriptions LOD resources into their search indexes and search result displays through tabular facts, knowledge cards, and similar presentational arrangements, we can anticipate that the expectations of the users of traditional digital library systems will evolve. The identification of LOD resources through named entity reconciliation and subsequent harvesting of additional metadata and data for use by digitized special collections' users helps to meet some of

[20] http://vocab.getty.edu
[21] http://id.loc.gov

these evolving expectations while also better optimizing digitized collections materials for the more nuanced and data-rich environment that the Semantic Web is beginning to provide.

Linking out to general-purpose resources like Wikipedia or topically pertinent online repositories like the Internet Broadway DataBase (IBDb) or Theatricalia seem likely strategies for further optimizing digital library collections for use, though more in-depth and longitudinal assessments are required to confirm this. Such linking does carry with it risks that the linked to Web resource might go away at some point in the future. But these are typical risks for digital infrastructures and since no data is being fetched from these resources and rendered on-screen to the digital library's users, they are relatively minimal. Given the investment already made in digitizing the special collections with which we are working on this project, the incremental cost to enrich existing metadata with links and transform it into RDF seems reasonable; however, ultimately this determination will wait on further assessment of benefits. Finally in this project we are not attempting to assess in any systematic way the potential benefit that transforming legacy metadata into RDF may bring for machine, semantic-based reasoning. Obviously we have this in mind, especially in transforming the Kolb-Proust Archive for Research data from TEI to RDF using Schema.org semantics. Long term this remains a key motivation for exploring LOD for digital libraries. But for now, simply improving the connectedness of these collections will help us move away from the isolating information silos that have become all too ubiquitous.

Acknowledgements

We gratefully acknowledge the Andrew W. Mellon Foundation for generously funding the LOD for Digital Special Collections project. Though this research was supported in part by the Mellon Foundation, the opinions, conclusions and findings expressed are those of the authors and do not necessarily reflect the views of the Foundation.

References

Auer, S. R., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. *The Semantic Web. Lecture Notes in Computer Science 4825*, p 722. doi:10.1007/978-3-540-76298-0_52.

Berners-Lee, T. (2006). Linked Data. Design Issues. [revised 2009]. Retrieved from: https://www.w3.org/DesignIssues/LinkedData.html.

Corbett, L. E. (2016). Linked data advice anyone? (Who uses Google?) *Technicalities 36*(1), pp 1-7.

Jett, J., Han, M.-J., and Cole, T. W. (2016). Disambiguating descriptions: Mapping digital special collections metadata to Linked Open Data formats. Poster presented at the *79th ASIS&T Annual Meeting* (14-18 October 2016, Copenhagen, Denmark).

Jett, J., Cole, T. W., Han, M.-J., and Szylowicz, C. (2017). Linked Open Data (LOD) for library special collections. Poster presented at *17th Annual ACM/IEEE-CS Joint Conference on Digital Libraries* (19-23 June 2017, Toronto, Canada).

Maron, Nancy L. and Pickle, Sarah. (2013). *Appraising our Digital Investment: Sustainability of Digitized Special Collections in ARL Libraries*, Association of Research Libraries and Ithaka S+R, 2013. Retrieved from http://sr.ithaka.org/research-publications/appraising-our-digital-investment.

Marty, P.F. (2008). An introduction to digital convergence: libraries, archives, and museums in the information age. *Archival Science* 8: 247-250. doi:10.1007/s10502-009-9094-1. Concurrently published: (2009) *Museum Management and Curatorship* 24 (4): 295-298, doi: 10.1080/09647770903314688; and (2010) *Library Quarterly* 80 (1): 1-5, doi: 10.1086/648549.

Park, J. and Childress, E. (2009). Dublin Core metadata semantics: An analysis of perspectives of information professionals. *Journal of Information Science 35*(6), pp 1-13.

Urban, R. J. (2012). Principle paradigms: Revisiting the Dublin Core 1:1 Principle. Doctoral Dissertation. Retrieved from: http://hdl.handle.net/2142/31109

Woodley, M. S., Clement, G., and Winn, P. (2005). DCMI Glossary. Dublin Core Metadata Initiative. Retrieved from: http://dublincore.org/documents/usageguide/glossary.shtml