FINAL REPORT |

G-31500650 | University of Illinois at Urbana-Champaign

▶ Table of Contents

▼ Grant Information

Organization Name:	University of Illinois at Urbana-Champaign
Grant Amount:	$248,000.00
Grant Number:	G-31500650
Grant Title:	Linked Open Data for Digitized Special Collections
Grant Description:	to support initiatives in digital services
Program Area:	Scholarly Communications
Grant Start Date:	9/23/2015
Grant End Date:	6/22/2017
Duration in Months:	21
Principal Investigator:	Timothy Cole
Grant Management Contact:	Ryan Dubnicek
Grant Signatory:	Robert Jones
Financial or Other Contact:	Linda Gregory
Additional Contact:	
Type of Support:	Spendable

Report Information

Due Date:	3/31/2018
Period Covered by Report From:	11/1/2015
Period Covered by Report To:	12/31/2017

Report Narrative

Please describe activities during the reporting period, progress made toward milestones, and any other significant accomplishments.

Description of Proposed Work:

Progress Report:

We successfully transformed legacy metadata for all 3 of our target collections into RDF descriptions using schema.org semantics. As we report in the **attached white paper #1 (***Transforming Special Collections Metadata into Linked Open Data***)**, we were able to reconcile with linked data authorities 60% of the persons found in our theater collection legacy metadata, but only about 30% of the persons in the Kolb-Proust Archive for Research legacy metadata. For one of our theater collections we were also able to reconcile about 75% of the theater and play / stage production entities in our legacy metadata. Though reconciliation was less successful than we had targeted, it was sufficient to support the development of several new interface features, offering users enhanced context for these digitized special collection resources and this did allow the gathering of some user feedback. However, as noted in discussion of Setbacks and Challenges, unexpected changes in the content management system used by the University of Illinois Library and more time and effort than we expected required to generate RDF graphs for the resources in our target collections meant that only a few of our experimental interface enhancements have been incorporated into the production interfaces for our collections. In particular, we reached only a very early prototype (alpha) phase in developing an interface for visualizing the social network around Marcel Proust. (This early test interface was created and shared with a limited number of users, garnering some feedback, but less than we had planned.)

We did conduct one-on-one user testing both before and after interfaces were enhanced and the details of this testing are described and analyzed in **white paper #2 (***Analysis of Early User Feedback***)**. We also had good feedback from **our in-person Project Advisory Committee meeting** (minutes from this meeting are being finalized and will be available from our Project Website next week). However, a group expert user (faculty) panel we had planned to hold was converted into additional one-on-one testing given that we did not make as much progress with our experiments making use of RDF descriptions as we had planned. In fact the last few of the interviews we did conduct were completed after the end of the period of performance (since the Library here is committed to further work with linked data for special collections), but in time for inclusion in white paper #2, a near final draft version of which is attached. Though the sample is small and arguably not specific enough in all cases to the collections tested, early user feedback regarding enhanced functionalities made possible through the use of Linked Open Data was positive. Overall, participants are enthusiastic about the enhancements made possible by the addition of linked open data, but with several caveats and qualifications. Considerable iteration over time will be required to really see the beneficial results wanted.

Very preliminary work was carried out **exploring ways to leverage connectedness to linked data resources such as Wikipedia/DBPedia**. Again we did not get to this until late in the project due to delays in transforming our legacy metadata to RDF, so there was not time or resources to assess impact. Of 58 theatres represented in the Motley collection legacy metadata, only one was identified as not having a pre-existing page in Wikipedia--*The Alexandra Theatre in Stoke-Newington* (London). Project staff developed a Wikipedia article for this theater, https://en.wikipedia.org/wiki/Alexandra_Theatre_(London), which was officially adopted by Wikipedia on 16 November 2017, and includes a link back to our collection as well as to several other resources. Our reconciliation and transformation to RDF work also revealed linkages to a number of already existing Wikipedia pages about plays, specific productions, and people. Of 26 such pages analyzed,10 already had links back to Motley collection pages (e.g., https://en.wikipedia.org/wiki/Becket), but for a few other pages that did not we were able to add links to our Motley Collection, e.g., https://en.wikipedia.org/wiki/Alec_Guinness. Only about 24-person hours were spent investigating opportunities to leverage connectedness of items in our collections to Wikipedia. While we established that there are viable opportunities do this, the cost in resources (project staff time) is high with as yet unproven value.

As expected we did produce JavaScript, Python code to remediate, reconcile and transform our legacy metadata into RDF. We also produced or modified JavaScript, XML Style Sheets and CSS stylesheets to add and enhance user interface features that made use of the RDF created. These materials have all been posted to GitHub and made available for use free of charge. **We are attaching to this report a summary of relevant URLs - ProjectWebsiteAndSoftwareRepoURLs.pdf. This file provides GitHub URLs and links to working implementations (some still in early development).**

Setbacks or Challenges:

These were three in number:

1. As detailed in our request for a no-cost extension a year ago, staffing changes and in particular hiring replacements for our primary project developer and our post-doc (converted to increased PhD research assistant hours) slowed us down initially and necessitated an extension to the performance period for the grant.

2. Also as detailed in last year's request, the decision of the Library here to move away from CONTENTdm as our image

content management system reduced (or at least has delayed) the opportunity to transfer the technology we have developed into production.

3. Finally, as alluded to above remediation, we were also challenged by the fact that reconciliation and transformation of legacy metadata into RDF has gone much slower and required more resources than we anticipated, leaving us fewer resources and less time to experiment and assess potential benefits. Our experience suggests that in particular entity reconciliation of special collection legacy metadata requires more manual intervention than we assumed.

Personnel/Contacts:

Have there been changes since the last report to contacts for organizational leadership, management or grant management staff?	No
The individuals listed in the Grant Information section above have been reviewed and changes are required.	No

Lessons Learned:

These are covered in detail in supporting materials, i.e., the attached White Papers covering the legacy metadata remediation, reconciliation and transformation into RDF and our analysis of early feedback from users. At a high, institutional level, we have gained confidence working with RDF, linked data and the semantic web; we have a better sense, albeit still incomplete, of the potential of these technologies; and we have come to realize that, given our still incomplete understanding of the potential benefits and the state of our special collections legacy metadata, more resources and more iterative research with users to identify the benefits of linked data will be required than originally anticipated.

Goals Following the Grant End:

As mentioned, our image content management system changed unexpectedly halfway through our project. This has been impediment to moving the technical solutions we developed into production. We are now working with our production IT staff to accomplish the necessary technology transfer and anticipate that our production services will benefit directly from our experimentation. We also anticipate applying much of what we learned on this project to our ongoing *Emblematica Online* project, as well as to work being done here in connection with *HathiTrust* and the *HathiTrust Research Center*. Finally, we are also currently exploring the use of Amazon Web Services to deliver more of our special collection services (potentially including *Emblematica Online*). Our experience on this project has positioned us well to begin experimentation with Amazon's Neptune triple store.

Evaluations:

Publications and Press:

Press Announcement of Project: *Mellon Foundation supports CIRSS research on linked open data, digitized special collections*, 1 December 2015. Available: http://ischool.illinois.edu/articles/2015/12/mellon-foundation-supports-cirss-research-linked-open-data-digitized-special

Jett, Jacob, Myung-Ja K. Han, and Timothy W. Cole. (2016). "Disambiguating Descriptions: Mapping Digital Special Collections Metadata into Linked Open Data Formats," Proceedings of the Association for Information Science and Technology 2016 Meeting. Available: https://www.asist.org/files/meetings/am16/proceedings/submissions/posters/16poster.pdf

Cole, Timothy W., Myung-Ja K. Han (June 8, 2016). Works in Progress Webinar: Linked Open Data for Digitized Special Collections, presented for OCLC. Available: https://www.oclc.org/research/events/2016/06-08.html

Cole, Timothy W., Deren Kudeki, Alex Olivia Kinnaman. (June 6, 2017). Exploring the Challenges & Uses of Linked Open Data for Digitized Special Collections, webinar presented for ASIS&T Special Interest Group in Digital Libraries. Available: https://www.asist.org/events/webinars/exploring-the-challenges-uses-of-linked-open-data-for-digitized-special-collections/

Jett, Jacob, Timothy W. Cole, and Myung-Ja K. Han (2017). "Linked Open Data (LOD) for Library Special Collections." Joint Conference on Digital Libraries 2017, Toronto, Canada.

Jett, Jacob, Timothy W. Cole, Deren Kudeki, Myung-Ja K. Han, and Caroline Szylowicz. (2017). "Expanding Metadata

with Linked Open Data." Proceedings of the International Conference on Dublin Core and Metadata Applications 2017: pp 29-38.

Han, Myung-Ja K., Timothy W. Cole, Jacob Jett, Deren Kudeki, and Caroline Szylowicz. (December 5, 2017). Integrating LOD into Library's Digitized Special Collections, presented at the Semantic Web in Libraries (SWIB) 2017, Hamburg, Germany.

Han, Myung-Ja K. Timothy W. Cole, Deren Kudeki, and Jacob Jett. (February 10, 2018). What Works and What Doesn't: Name Reconciliation Work and Management of Names in Digitized Special Collections, presented at the Linked Data Interest Group Meeting at the American Library Association Conference, Denver, CO.

Han, Myung-Ja K. (February 10, 2018). Linked Data for Library Services, presented at the OCLC Linked Data Roundtable at the American Library Association Conference, Denver, CO.

Additional Information:

▼ Grant Budget

Financial Reports should be prepared on the same Budget and Financial Report template submitted with the proposal or the approved Revised Budget and Financial Report template (if applicable). Financial Reports should be uploaded by clicking on the "+" sign below and attaching the spreadsheet.

Budget Variances and Investment Income:

Other than the variances previously approved as part of our a no-cost extension (March 2017), there were minimal variations in expenditures versus budget. The largest adjustment, $13,734, was reallocated from the MS grad hourly line to help cover an increase in hours allocated to our academic hourly researcher (who had just received her MS degree) and a small increase in level of effort of the project developer during the final months of the project. Actual expenditures (mostly non-employee travel) for the Advisory Committee meeting last spring in Chicago were $3,063.09 less than anticipated; the surplus helped cover travel expenses to present papers disseminating project research at the 2017 DCMI and SWIB conferences (see above under publications). There were other small adjustments in co-PI level of effort, supplies and hourly benefits categories necessary to cover overages in staff benefits and a slight increase in PhD RA level of effort during the extension period. All investment income was reinvested in the budget, and as detailed in the no-cost extension paperwork, went toward staff time during the extension period.

Supporting Material

Is there supporting material for this report?   Yes

Please upload materials in the Documents section.

Publication and Reproduction Limitations:

1. MellonFdn31500650-C3002-Interim-Financial-28Feb2017.pdf, signed Financial report covering 1 Nov 2015 through 28 Feb 2017 (previously submitted as part of no-cost extension request)

2. MellonFdn31500650-C3002-FinalFinancialReport.pdf, signed Financial report covering 1 March 2017 through 31 Dec 2017

3. WhitePaper1-TransformingSpecialCollMetadata2LOD.pdf, *Transforming Special Collections Metadata into Linked Open Data: Mappings, entity reconciliation, workflows implemented & lessons learned* (a deliverable listed in our project proposal).

4. WhitePaper2-EarlyUserFeedback_draft20180328kf.pdf, *Analysis of Early User Feedback* (a deliverable listed in our project proposal)

5. ProjectWebsiteAndSoftwareRepoURLs.pdf, *Project Websites and Software Repository URLs* (includes links to all remaining deliverables including code and prototype Websites).

Documents

DOCUMENTS
ProjectWebsiteAndSoftwareRepoURLs.pdf **Other Document - Report** Added by Timothy Cole at 6:41 PM on March 31, 2018
WhitePaper2-EarlyUserFeedback_draft20180328kf.pdf **Other Document - Report** Added by Timothy Cole at 6:41 PM on March 31, 2018
WhitePaper1-TransformingSpecialCollMetadata2LOD.pdf **Other Document - Report** Added by Timothy Cole at 6:41 PM on March 31, 2018
MellonFdn31500650-C3002-FinalFinancialReport.pdf **Financial Report** Added by Timothy Cole at 6:41 PM on March 31, 2018
MellonFdn31500650-C3002-Interim-Financial-28Feb2017.pdf **Financial Report** Added by Timothy Cole at 6:41 PM on March 31, 2018

Grantee Authorizations

In the space below, please enter the name and contact details of the person in your organization with institutional responsibility for financial reporting who reviewed the financial report.

Name: Linda Gregory

Title: Interim Director, GCO Post Award

Date: 3/27/2018