UNLV Libraries: Linked Open Data Implementation Report

Inna Gogina

INFO 241-12. Automated Library Systems

San José State University

November 16, 2015
UNLV Libraries: Linked Open Data Implementation Report

In fall 2012, Head of Digital Collections at University of Nevada, Las Vegas (UNLV) Libraries Cory K. Lampert and Digital Collections Metadata Librarian Silvia B. Southwick set out to design an exploratory project aimed at studying “the feasibility of developing a common process that would allow the transformation of metadata records from digital collections into LOD, preserving their original expressivity and richness,” and publishing data from the UNLV Digital Collections “in the LOD Cloud to improve discoverability and connections with other related data sets on the Web” (Southwick, 2015, p. 12). Three years later, as of this writing, the project has approached a final phase of implementation.

Following the overview of the UNLV Libraries’ Digital Collections and their user needs, as well as history of linked open data (LOD), this paper focuses on the LOD implementation at UNLV, by way of transformation of the local digital collection metadata into linked data using open-source technologies. “As a relatively new topic, the library literature lacks best practices for implementing LOD,” Southwick points out (2015, p. 1). Therefore, rather than relying on the LOD implementation case studies at peer institutions and best practices, this analysis is mainly supported by the LOD implementation reports published by UNLV, as well as additional professional and scholarly reports from the field.

Part I. UNLV Libraries: Background

University of Nevada, Las Vegas is a public research-intensive university with strong emphasis on science and technology, business management, and law. It offers 95 undergraduate degree majors and 71 minors. The University maintains several libraries, including four on-the-ground facilities—Lied Library, Architecture Studies Library, Curriculum Materials Library, and Music Library—and two online branches: the Shadow Lane Library, serving Dental School
faculty and students, and the Singapore Campus Library, serving College of Hotel Administration. The Libraries employ 113 staff—including 37 tenure-track librarians—anually serving over 1,775,302 visitors. The operating budget is over $17 million including $6.5 million for collections (University of Nevada, Las Vegas, 2015, *Quick Facts about UNLV Libraries*).

**Digital Collections**

The Libraries maintain rich traditional and digital collections, in order to serve the main user groups—students, faculty, and visiting scholars. The Digital Collections, in particular, “feature a wide range of original source materials from UNLV Libraries Special Collections, the Nevada State Museum, the Historical Society of Las Vegas and the Clark County Heritage Museum” (ibid.). “There are approximately 45,000 items in the collections, including photographs, maps, manuscripts, architectural drawings, multimedia, and oral histories,” Lampert and Southwick report (2013, p. 242).

The main goal in serving the needs of Digital Collections’ users is to provide online access to these unique materials via multiple ways of discovery. The Libraries use CONTENTdm to manage content of, and provide access to, the UNLV Digital Collections (Southwick, 2015, p. 19).

**Mission and Vision**

“In support of the University’s mission and shared values, the Libraries contribute to and support learners as they discover, access, and use information effectively for academic success, research, and life-long learning,” the Libraries’ mission statement maintains (University of Nevada, Las Vegas, 2015, *Mission and Vision Statements*).

The Libraries’ Vision Statement is focused on defining “the new academic research library—bringing people and information together in innovative ways. As UNLV emerges as a
leading urban research institution, the Libraries will pioneer dynamic, user-focused methods of reaching, connecting, and engaging learners” (ibid.).

**User needs**

In the 21-st century information environment, users want to find relevant information, and to find it fast. In the last decade, UNLV Libraries—while pursuing their mission and strategic vision—have experienced challenges facing all academic libraries in the U.S., in meeting the information needs of their patrons. “The biggest challenge facing the library profession in the twenty-first century is staying relevant to its users,” Jennings points out (2013). Of particular challenge to the UNLV Digital Collections has been the “inability to adjust fast enough to keep up with the changing technologies, and users’ needs and rising expectations” (Xu and Yang, 2015, pp. 45-46), at the time when user search behavior has shifted toward a Google-like information search from a “one-stop-shop”.

Encapsulated in the CONTENTdm’s object records, metadata describing the digital objects in UNLV Libraries’ collections have been “only accessible to users when records containing them are retrieved in a search. This approach for managing data, although a common practice that extends far beyond digital collections, creates silos of data,” UNLV digital librarians explain. “Data associated with records is isolated and does not directly link to related data existing in other records. At most, when present, links are at the level of records. These silos most certainly hide valuable relationships among data, leaving to users the task of discovering these hidden collections” (Southwick & Lampert, 2013, p. 1). In addition to the inability of creating “explicit connections between related materials in different digital collections,” other challenges included “limited metadata schema choices, limited data structures, lack of ability to export data in common reusable formats (such as PDF),” “redundant work processes to create
and maintain metadata records, and the lack of tools to transform records into linked data,”
according to Lampert and Southwick (2013, p. 242).

Prior to LOD implementation, a user “looking for information about a particular person
and bases the search on the name of the person […] might get many results of people or
organization names, book titles, or other texts that contain similar names but are not exactly the
person the user was looking for. In such a scenario, users will need to weed out the wrong
information and select the relevant ones”, Southwick observes. “In the context of LOD, the
particular person who is the subject of the search would have a unique identifier that could be
used to find all information about him/her. In this scenario, only data for the specific person will
be retrieved” (2013, p. 3).

In order to overcome these challenges and provide a “more granular (and far more
powerful) approach to managing data” (ibid.), as well as a tool for seamless discovery and
delivery of resources from their own collection, and collections from partner institutions, to the
users, the UNLV Digital Collections department embarked on the Linked Open Data project.

Part II. Linked Open Data: Overview

“The term Linked Data refers to a set of best practices for publishing and connecting
structured data on the Web,” Tom Heath maintains. “Key technologies that support Linked Data
are URIs (a generic means to identify entities or concepts in the world), HTTP (a simple yet
universal mechanism for retrieving resources, or descriptions of resources), and RDF (a generic
graph-based data model with which to structure and link data that describes things in the world)”
(Heath, n.d.). In computing, linked data is a “method of publishing structured data so that it can
be interlinked and become more useful through semantic queries”¹. Transforming the digital

¹ https://en.wikipedia.org/wiki/Linked_data
collections metadata into linked data allows libraries to publish the data “in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets” (Bizer, Heath, & Berners-Lee, 2009, p. 2)—which, in turn, increases “the relevance of libraries within the wider information ecosystem”\(^2\).

There is a difference between linked data and linked open data. “The label "Linked Open Data" is widely used, but often to refer to Linked Data in general, rather than to Linked Data that is explicitly published under an open license,” Heath points out. “Not all Linked Data will be open, and not all Open Data will be linked. Therefore care should be taken to use the appropriate term, depending on the licensing terms of the data in question” (n.d.). The most well-known application of linked open data is the Linking Open Data project, “a grassroots community effort founded in January 2007 and supported by the W3C Semantic Web Education and Outreach Group” (Bizer, Heath, & Berners-Lee, 2009, pp. 4-5). Europeana\(^3\) is another example of successful LOD implementation; UNVL Libraries followed Europeana’s example, in particular, having adopted its data model in the LOD project discussed in this paper.

The linked data movement has gained momentum in the library field since its inception: it “has created a Linked Data Cloud where data can be created locally and linked to existing data accessible globally in the cloud,” Southwick and Lampert report. “This movement is already being embraced by key organizations in the library field. For example, OCLC has begun efforts to transform WorldCat records into linked data. Since Uniform Resource Identifiers (URIs) are one of the foundations of linked data, Library of Congress has added URIs in their authority files and subject headings” (2013, p.1).

\(^2\) [http://www.oclc.org/data.en.html](http://www.oclc.org/data.en.html)
Part III. Linked Open Data Implementation at UNLV Libraries

“The UNLV’s LOD Project emerged from a UNLV Libraries study group created in April 2012,” Southwick notes. “The group comprised professionals from various areas of the University Libraries with the initial goal of understanding LOD concepts and its potential benefits to libraries” (2015, p. 12). Recognizing the importance of linked data in helping to better serve user information discovery needs, UNLV Libraries “designed and initiated a pilot project to apply linked data concepts to the practical task of transforming a sample set of […] CONTENTdm digital collections metadata into future-oriented linked data” (Southwick & Lampert, 2013, p. 1).

The Libraries realized the Linked Open Data Project in **five phases**: planning, design, implementation, publishing LOD, and LOD consumption (Southwick, 2015, pp. 12-31).

During the **planning phase**, the team designed a LOD prototype, focusing on testing technologies allowing transforming the local metadata into linked open data. The team also used the prototype to seek support from the University administration and secure funding to sustain the project.

The **design phase** involved selection of LOD technologies, research and selection of Data Model, mapping, and definition of rules for assigning URIs. The team employed several **LOD technologies**. In order to transform CONTENTdm metadata into LOD, OpenRefine[^4] was used, combined with its RDF extension (RDF Schema Alignment[^5]). For publishing and managing the linked data—the RDF files containing relational triples (Subject (*resource*), Predicate (*property*), and Object (*value*))—the Libraries used OpenLink Virtuoso Universal

[^4]: http://openrefine.org/
[^5]: http://refine.deri.ie/
Server\(^6\). For visualizing, the team used RelFinder\(^7\). Finally, SPARQL Protocol and RDF Query Language were implemented to query RDF data sets (Lampert & Southwick, 2013, p. 246; Lampert, 2015).

The Libraries adopted the **Europeana Data Model**\(^8\) (EDM) due to the similarity of content and relevance of classes and properties between the digital collections of Europeana and UNLV Libraries. “Another important factor that contributed to our decision is that EDM re-uses well-defined vocabularies for properties such as Dublin Core (dc/dcterms), Friend of a Friend (FOAF), etc.,” the implementing team reports (Lampert & Southwick, 2013, p. 244). Upon identifying a data model, librarians developed a mapping scheme among the CONTENTdm metadata, and the classes (core and contextual resources) and properties (Dublin Core and EDM specific properties) defined in the data model. While defining URI rules, “as a strategy to increase linkage among our data and to other data sets, we decided to create URIs only for “things” that have not yet received URIs by other data providers,” Southwick clarifies (2015, p.18). The URI syntax was designed in accordance with Wide Web Consortium linked data URI principles:

\[
\text{<Namespace}\text{^9} \text{<class of the “thing” being described> / <local unique identifier>}
\]

The **implementation phase** included export and cleaning of metadata; reconciliation of metadata with controlled vocabularies; preparing metadata for transformation, assigning URIs; and transformation of metadata into LOD (Southwick, 2015, p. 19). “The actual transformation of digital collections metadata into LOD is accomplished by implementing the mapping between the digital collections metadata elements and the EDM, as well as by creating or using existing

---

\(^6\) http://virtuoso.openlinksw.com/
\(^7\) http://www.visualdataweb.org/relfinder.php
\(^8\) http://pro.europeana.eu/share-your-data/data-guidelines/edm-documentation
\(^9\) http://ld.library.unlv.edu>
URIs. These actions will create the triples,” Southwick explains. “The final step in the transformation is to generate the RDF files” (2015, p.24).

The project’s publishing phase has been underway. The RDF files produced during the implementation phase, to form LOD sets from several collections—are uploaded onto an internal server, for final cleanup. Once finalized, the datasets will be published on the OpenLink Virtuoso Universal Server. The final steps include generating set-level metadata, linking to other data sets, providing provenance to metadata, providing licensing information, mapping local vocabulary terms to other vocabularies, and setting alternative ways to access the data sets (SPARQL endpoint and RDF dumps) (Southwick, 2015, p. 28).

The consumption phase has involved experiments with RDF files visualization tools—Pivot Viewer and RelFinder. “These experiments were very important for consolidating our knowledge on how to manage and use LOD,” Southwick concludes. “We now are ready to step forward and design a more comprehensive user interface” (2015, p. 31).

Part IV. Results

Evaluation and challenges

Aimed at evaluating benefits and challenges of transforming CONTENTdm records into linked data, the team evaluated the LOD data sets by “comparing query results from traditional records vs. linked data, assessing technology solutions, and analyzing impacts on workflow and metadata design” (Southwick & Lampert, 2013. p. 2). In order to test the implemented LOD sets with a peer institution, specifically as to how well the LOD tools can work with other data sets, UNLV partnered with North Carolina State University (NCSU) on a Data Exchange Project. The partners identified several challenges stemming from the automated reconciliation of two data sets, including unexpected matching results, which required further investigation. Additional
challenges involved time investment required to clean controlled vocabularies from legacy collections and training necessary for metadata creation and workflow redesign. Consistency in the use of standards was found to be of particular importance for mapping the record fields to linked data model. The partners presented the first testing results at the 2015 ALA Annual Conference in San Francisco, CA (Hanson, Lampert, & Southwick, 2015).

Benefits

The main benefit of linked data is in its ability to link data from different sources. “These may be as diverse as databases maintained by two organizations in different geographical locations, or simply heterogeneous systems within one organization that, historically, have not easily interoperated at the data level,” Bizer, Heath, and Berners-Lee maintain (2009, p. 2). Prior to LOD implementation, UNLV Digital Collections represented an example of the latter, as detailed in the User Needs section of this paper.

As a result of implementing LOD, UNLV Libraries have gained several important benefits, all of which are critical in the user information needs fulfilment. First, adoption of the LOD concepts and technologies provided the Libraries’ Digital Collections an opportunity “to break up the silos in which [local] digital collections metadata reside,” and, second, to “interconnect [local] data with relevant data from other data providers,” increasing the search results’ precision—Southwick points out (2015, p. 3). “Lastly, the interconnection of data” provided users with opportunity “to seamlessly search relevant information regardless of where it was generated or who generated it” (ibid.).

As a result of LOD implementation, the Libraries have begun revising their digital collection management policy. “For example, we have been revising our metadata practices to facilitate a more immediate, real-time generation of metadata into LOD. In addition, we are
adopting new processes for our daily activities concerning the management of metadata: generation and maintenance of local URIs to assign to our unique things, rigorous management of local controlled vocabularies, commitment to monitor the development of new data sets to which our data may be linked,” Southwick reports. “Although this management is local to our organizations it may have enormous implications for the ability to UNLV’s data set to integrate into the global LOD context” (2015, p.34).

Conclusion

“Linked data, explained as the transition from Web documents to Web data, will greatly benefit information search and discovery and will influence user experiences and expectations of contextual information,” Lampert and Southwick predict. “Linked data will also impact future library workflows and require new perspectives on ingesting, managing, distributing, and aggregating large data sets created from library metadata” (2013, p. 230).

While the UNLV team did encounter several challenges during the process of LOD implementation, the results were positive in many ways. Based on the critical evaluation of the LOD implementation results, it can be said with confidence that by transforming the diverse digital collections sources and types of data from CONTENTdm into the future-oriented linked data, UNLV Libraries have truly come a step closer in achieving their vision of defining the new academic research library, and bringing its users and information together in innovative ways.
References


