Development of an RDF-Enabled Cataloguing Tool

Lucy McKenna[1], Marta Bustillo[2], Tim Keefe[3], Christophe Debruyne[1], and
Declan O'Sullivan[1]

[1] ADAPT Centre, Trinity College Dublin, Ireland,
lucy.mckenna@adaptcentre.ie
[2] UCD Library, University College Dublin, Ireland,
[3] Digital Resources and Imaging Services, Trinity College Dublin, Ireland

Abstract. By generating bibliographic records in RDF, libraries can publish and interlink their metadata on the Semantic Web. However, there are currently many barriers which prevent libraries from doing this. This paper describes the process of developing an RDF-enabled cataloguing tool for a university library in an attempt to overcome some of these obstacles.

Keywords: semantic web, linked data, MODS, RDF, library, interface design, usability testing

1 Introduction

The Digital Resources and Imaging Services (DRIS) department of the Library of Trinity College Dublin (TCD) hosts the Digital Collections Repository of the university. This repository provides open access to TCD's collection of digitised cultural heritage materials which includes manuscripts, letters, books, images, and other archival materials. DRIS aims to publish the bibliographic data of its collections as RDF in order for these materials to be discoverable on the SW, increasing the visibility and use of the library's resources. Additionally, RDF metadata published by DRIS could be interlinked with Linked Data (LD) emerging from other institutions, facilitating library users to access a web of related data from a single information search [1].

2 Libraries and Linked Data

Although not yet widely used, libraries are publishing bibliographic metadata as RDF in increasing quantities [1, 2]. However, librarians have reported a number barriers in using LD to its full potential including that LD software is not tailored to the specific needs and expertise of librarians but rather technical experts. Other reported challenges included a lack of authority control on the SW, difficulties establishing interlinks, and few examples of useful applications of LD in the library domain that would justify the allocation of time and resources to

its generation [3, 4]. These challenges were experienced by DRIS and prevented the library from publishing its metadata to the SW. As such a bespoke RDF-enabled cataloguing interface was developed for DRIS. The aim of the interface was to explore whether such a tool could be used by DRIS to successfully generate MODS-RDF records for a small sample of records thus demonstrating the potential for LD software specifically designed for library use.

3   MODS and MADS

The Metadata Object Description Schema (MODS) is an XML schema for a bibliographic element set that can be used for the purpose of cataloging digital resources [5]. The full schema consists of 20 top-level elements, for example TitleInfo and Name, which are used to provide information on the title and creator of a work. The majority of MODS elements contain subelements, such as title, subtitle, and namePart, as well as attributes which describe the metadata itself, for example, the authority source from which a title or name was taken, or the language used when cataloguing.

MODS was selected as the output schema for the tool as it was sufficiently detailed for DRIS's cataloguing purposes and a MODS-RDF ontology was already available [6]. Additionally, a set of MODS implementation guidelines was developed by the Digital Library Federation's (DLF) Aquifer Initiative thus allowing for the standardisation of MODS records [7].

The Metadata Authority Description Schema (MADS) [8] can serve as a companion to MODS to provide metadata regarding the authority sources used in a record when describing names, organisations, genres, or subjects for example. Like MODS, a MADS-RDF ontology already exists [9]. Both MODS and MADS share a number of subelements, such as those in TitleInfo, Name and Subject. The schemas also share all attributes. Interestingly the MODS-RDF ontology excludes all elements it has in common with MADS. As such, in order to generate a full MODS record in RDF, both ontologies must be used.

4   Interface Design and Testing

A semi-structured interview was carried out with the DRIS metadata cataloguer in order to establish a set of tool requirements, and a mock-up of the cataloguing interface was subsequently developed. User requirements included:

– Facilitating cataloguing efficiency by automating input where possible.
– Publishing MODS records that meet DLF-Aquifer requirements by forcing data entry for certain fields and constraining data entry options for others.
– Further constraining data entry options as per the specific needs of DRIS.
– Providing additional administrative data entry fields.

The completed interface was programmed to initially constrain data entry options to only those elements and subelements which were identified as required

fields by the DLF. This was done to ensure that the minimal data requirements for each record were met prior to the addition of supplementary metadata. Once these fields were complete, data entry options expanded to include recommended and optional fields.

Data entry fields and dropdown menu options were programmed to dynamically alter based on prior selections made during the cataloguing process. This ensured that data entry options were restricted to DLF recommendations. For example, in the Name element, DLF require that the resource creator's name should be taken from the Name and Title Authority Source Codes maintained by the Library of Congress (LOC). Thus the list of options in the authority menu was constrained to these sources, this was then further constrained to display only the sources used by DRIS. Data entry fields also self-populated based on prior selections allowing for a more efficient cataloguing process. For example, again in Name, after selecting an authority source the Authority-URI field self-populated. This also highlights how the tool was capable of accepting URIs to other LD datasets - a first step in the LD interlinking process.

The interface was tested by observing the DRIS metadata cataloguer using the tool to create a bibliographic record. Although results indicated some issues with the interface layout, the librarian felt that the tool would be useful for creating more authoritative RDF datasets and that it could facilitate increased LD generation by librarians rather than technical experts alone.

5   Record Generation

Data from the interface was stored in a relational database. In oder to uplift this data to RDF an R2RML mapping was developed based on the MODS and MADS RDF ontologies. R2RML is a W3C Recommendation for declaring mappings from relational databases to RDF datasets [10]. In the process of adding MADS to the mappings it was noted that, unlike MODS-RDF where properties are represented individually, some MADS-RDF properties were grouped in collections including the subelements in TitleInfo and Name. Collections are a special RDF construct used to represent lists. This grouping allows for labels, such as title and subtitle, or first and last names, to be reconstructed with all elements in the correct order. However, at the time of the project, R2RML did not support the mapping of RDF collections, thus some metadata, such as subtitle, and more than one namePart were omitted. Despite this setback, semi-complete RDF records were generated for a small sample of DRIS's materials. A number of SPARQL (RDF query language) queries were successfully run over the RDF dataset including typical searches by author, date, and genre, as well as more interesting and detailed searches by ISO Language and Country Codes, authority sources, controlled vocabulary terms, and URIs.

This issue inspired a separate project in which an R2RML expansion supporting the mapping of RDF Collections (and Containers) was developed [11]. This expansion facilitated the uplift of all metadata in the database to RDF, allowing for the publication of complete MODS records.

6 Conclusions and Future Directions

Providing librarians with bespoke LD tools would allow for increased publication of rich LD datasets. It is likely that LD generated by librarians would be treated with increased credibility and thus used more frequently as libraries are viewed as trustworthy and authoritative sources of information. LD created by librarians will follow specific and standardised bibliographic schemas, and use long established authorities and controlled vocabularies to describe resources. This would increase the level of authority control on the SW, allowing for similar entities to be identified consistently across the SW leading to richer search results.

Future research will explore how to engage librarians in the process of interlinking with LD datasets published by other libraries and related institutions rather than just large scale authorities (LOC) and LD datasets (DBpedia). This would allow library users to access larger amounts of related data from single information search.

Acknowledgments

References

1. Hastings, R.: Linked Data in Libraries: Status and Future Direction. In: Computers in Libraries, 35, 12–16 (2015)
2. Mitchell, E. T.: Library Linked Data: Early Activity and Development. In: Libr Technol Rep, 52, 5–33 (2016)
3. Hallo, M., Lujan Mora, S., Trujillo Mondejar, J. C.: Transforming library catalogs into Linked Data. In: ICERI (2013)
4. OCLC, (2017), Linked Data Survey, `http://www.oclc.org/research/themes/data-science/linkeddata.html`
5. Library of Congress, (2017), MODS, `http://www.loc.gov/standards/mods/`
6. Library of Congress, (2012), MODS-RDF , `http://www.loc.gov/standards/mods/modsrdf/v1/modsrdf.owl`
7. Digital Library Federation, (2009), MODS Implementation Guidelines `https://wiki.dlib.indiana.edu/download/attachments/24288/DLFMODS_ImplementationGuidelines.pdf`
8. Library of Congress, (2017), MADS, `http://www.loc.gov/standards/mads/`
9. Library of Congress, (2017), MADS-RDF , `http://www.loc.gov/standards/mads/rdf/mads-ontology-20101119.owl`
10. W3C, (2012), R2RML, `https://www.w3.org/TR/r2rml/`
11. Debruyne, C., McKenna, L., O'Sullivan, D.: Extending R2RML with Support for RDF Collections and Containers to Generate MADS-RDF Datasets. In: TPDL (2017)