Tutorial: Linked Data for Digital Libraries

Uldis Bojārs[1], Nuno Lopes[2], and Jodi Schneider[2]

[1] National Library of Latvia
[2] Digital Enterprise Research Institute, National University of Ireland, Galway
uldis.bojars@gmail.com
nuno.lopes@deri.org
jodi.schneider@deri.org

This tutorial will empower attendees with the necessary skills to take advantage of Linked Data already available on the Web, provide insights on how to incorporate this data and tools into their daily workflow, and finally touch upon how the attendees' own data can be shared as Linked Data.

Linked Data has been embraced as the way to bring complex information onto the Web, enabling accessibility while maintaining the richness of the original data, making it ideal for digital libraries that want to increase their search visibility and interoperability. Thus digital libraries are giving increasing importance to Linked Data in a variety of ways – by creating metadata models (such as the Europeana Data Model and the Library of Congress Bibliographic Framework) and by publishing Linked Data from digital library projects (such as LC's Chronicling America Historic Newspapers, WGBH's public television video archives, ...), authority files (e.g. VIAF), and bibliographic catalogs (used by many national libraries, pioneered by the Swedish LIBRIS project since 2008). Other types of Linked Data available include crowdsourced information from Wikipedia and OpenStreetMap, published as Linked Data by the DBpedia and LinkedGeoData projects, respectively. Meanwhile, library-created Linked Data has been reused by Wikipedia: biography articles in the English and German-language Wikipedia projects now link to the `VIAF.org` name authorities.

Commonly used software systems (such as Evergreen, Hydra, and Omeka) are beginning to automatically produce Linked Data, and a deeper understanding of the principles of Linked Data is beneficial for making best use of these systems. For practitioners, this tutorial provides a greater understanding of what Linked Data is, and how to prepare digital library materials for conversion to Linked Data. For researchers, this tutorial updates the state of the art in digital libraries, while remaining accessible to those learning Linked Data principles for the first time. For library and iSchool instructors, the tutorial provides a valuable introduction to an area of growing interest for information organization curricula. For digital library project managers, this tutorial provides a deeper understanding of the principles of Linked Data, which is needed for bespoke projects that involve data mapping and the reuse of existing metadata models.

The introduction will start by explaining the motivation for using Linked Data for digital libraries, then will provide an overview of Linked Data and the concepts involved, including links as the basis for the Web, URIs as identifiers for objects, and the RDF data representation model. We will highlight some of the most widely used RDF vocabularies and ontologies, for example the Biblio-

graphic Ontology, Dublin Core, FOAF (short for 'Friend of a Friend ', for data about persons) and NeoGeo (for describing geographical data). Turning our focus to digital libraries we will explore the uptake of Linked Data in the cultural heritage sector and will look at some examples including the metadata standards and technologies used.

The second part of the tutorial will present tools that can be used to visualize, validate, and transform Linked Data. Participants will learn to discover and analyze existing Linked Data sets, with special focus on Library Linked Data. We will look in some detail at library information contained in the Linked Data cloud (`http://lod-cloud.net/`), for instance the Open Library data, as well as how Linked Data is used in a growing number of national libraries. We will provide examples of tools for translating MARC records to RDF as well as tools for translating spreadsheets to RDF (e.g. MARiMbA and the RDF extension to Google Refine). We will also present existing library and repository software that are using Linked Data. Also relevant for the third part of the tutorial, we will look at existing providers of geographical data like LinkedGeoData, which exports user-generated data from OpenStreetMap. We will briefly introduce participants to the SPARQL query language for RDF and show how public Web data can be integrated into your Linked Data applications.

The final part of the tutorial will pull together the information presented in the first two parts. We will explore in more detail two use cases: (1) the VIAF international authority file project and how it is being used outside libraries; and (2) how to publish and use Linked Data about maps.

In the first use case, we will look at a large-scale Linked Data project, VIAF, which is integrating multilingual information contributed by national libraries from every continent. We will also highlight how VIAF is being used outside libraries. In particular, we will look at how the Wikipedia Authority Control initiative acts as an entry-point to the web of linked authority data. The second use case will explore enhancing library records, using the online database of Irish place names, Logainm (`http://logainm.ie`), as an example of how Geographic Linked Data can be used as a controlled vocabulary for place names, for example to avoid misspellings or to disambiguate place names. We describe the creation of a new Linked Data dataset, pointing out the steps needed to build Linked Data applications. We will show how we can automatically leverage the resulting Linked Data version to enhance metadata records by searching for place names that are represented in digital maps.