Greenberg, J., Méndez Rodríguez, E. & Bueno de la Fuente, G. (2013). Linking and opening vocabularies LHT special issue: Editorial. *Library Hi Tech*, 31 (4), pp. 569-574.

Linking and opening vocabularies *LHT* special issue: Editorial

Jane Greenberg[1]

School of Information and Library Science,
University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Eva Méndez Rodríguez

Dpto. Biblioteconomía y Documentación, Universidad Carlos III de Madrid, Madrid, Spain, and

Gema Bueno de la Fuente

Library and Information Science, University Carlos III of Madrid, Madrid, Spain

ABSTRACT

Purpose

- This editorial underscores the importance of linked data and linked open data (LD/LOD) in contemporary librarianship and information science. It aims to present the motivation for this special issue of *Library High Tech (LHT),* specifically the theme of linking and opening vocabularies (LOV) as a component of the LOD landscape. The editorial also seeks to describe the content of the papers and their contributions in the areas of current practice, trends, and R&D (research and development).

Design/methodology/approach

– The editorial is organized into four main sections: Introductory aspects about linking open data; the context and inspiration for this special issue of *LHT*; a description of the accepted papers highlighting their contribution for LOV; and a conclusion.

Findings

– The editorial presents several reflective insights for advancing LOD and LOV practices that may lead toward a more integrated semantic web.

Originality/value

– The editorial presents a synthesis of the three editors' views on approaching the subject of linking open vocabularies. The framework for the editorial is reflected in the original call for papers that supported this special issue.

KEYWORDS

Linked open data, Linked data, Vocabularies, Semantic web, Linking and opening vocabularies

---

[1] Jane Greenberg can be contacted at: janeg@email.unc.edu

Introduction

Linked Data (LD), Linked Open Data (LOD) and generating a web of data, present the new knowledge sharing frontier. In a philosophical context, LD is an evolving environment that reflects humankinds' desire to understand the world by drawing on the latest technologies and capabilities of the time.

Viewed pragmatically, LD is about using the Web to connect related data that was not previously linked or using the Web to lower the barriers to linking data currently using other methods (Linkeddata.org, 2013). LD embodies a movement and a "best practice" for making data open, accessible machine-understandable, and for connecting data in a particular way, using Semantic Web Technologies (RDF, URIs, and SPARQL).

LD, while seemingly a new phenomenon did not emerge overnight; rather it represents the natural progression by which knowledge structures are developed, used, and shared. Linked Open Vocabularies, increasingly referred to as LOV (Vatant, 2012), is a significant trajectory of LD. LOV targets vocabularies that have traditionally been contextualized as Knowledge Organizations Systems (KOS). This special issue of *Library High Tech* (*LHT*) focuses on LD and LOV in the current semantic information landscape. The introduction that follows includes three sections. Section 2, History and Context, sets the stage for LD and LOV; Section 3, This issue of *LHT*, presents the selected contributions, which include two examples of current LD practice, and three pieces representing LD R and D activities. Section 4, Conclusion, closes this introduction with several observations.

History and context

Early recorded history presents evidence of knowledge structures that document association, comparisons, and distinctions. The marking of astronomical observations and the history of calendars, from the Sumerian tablets of the 3000 BC to later developments, such as the Gregorian calendar (Richards, 2000), present obvious examples. As knowledge about the universe has evolved, religious scholars, philosophers, and mystics alike outlined new frameworks; and these schemes are reflected in our contemporary knowledge systems. Moreover, the documentation of

knowledge has enabled disciplines, domains, and communities of practices to evolve, thrive, and spur new areas of study.

Millenniums later, we are experiencing a thriving digital ecosystem fed by networked communication. It is natural to consider how to leverage knowledge structures and benefits from these capabilities. Questions abound asking – what is the best and most feasible way to present information associations, comparisons, and other connections; how can technological innovation enable a deeper understanding of our world? A logical step has been enable open access to vocabularies that record the disciplinary languages. A complementary and significant step has been enabling machine manipulation and processing of such vocabularies through efforts such as LOV project[1], Open Metadata Registry (OMR)[2], and the NC Bioportal[3] (also reviewed in Me´ndez and Greenberg, 2012).

By marking-up standard thesauri, classificatory systems, and ontologizing a discipline's language according to Semantic Web standards, new automatic capabilities for linking knowledge can be realized.   To follow suit, it is one thing to put a vocabulary on the web, and it is another thing to make it for the web. The Simple Knowledge Organization System (SKOS) and Web Ontology Language (OWL), Semantic Web standards, transform traditional KOS, making them for the web. Rendering a vocabulary following these standards enables new opportunities for knowledge discovery and sharing across our digital universe. This progress is also reflected in closed systems that use LD – even if the vocabularies   are not available for machine consumption on the global web. In this case, knowledge sharing has more restrictive boundaries, although the advancements within a community can be significant. This issue of *LHT* is concerned with the linked data and the potential to weave an integrated web of data, as noted in   tow of the co-editors' earlier work, *Knitting the Semantic Web (2007)*. The sections that follow explain the inspiration for this issue of *LHT*, summarize the contributions, and provide several concluding remarks.

This issue of *LHT*

This special issue was inspired, chiefly, by the LOV symposium and HIVE (Helping Interdisciplinary Vocabulary Engineering) workshop held in Madrid in June 2012 (http://bit.ly/LOV-Symposium-Wiki) (Melgar *et al.*, 2013). The two-day event drew nearly 60 participants and was international in scope (participants were from: Argentina, Brazil, Germany, Netherlands, Poland, Portugal, Spain, the UK, and the US), with another 35-40 attending virtually). The symposium component specifically addressed linked data for LOV. Two papers in this *LHT* issue are drawn from this event, with the additional selections submitted in direct response to a call inspired by the event. The HIVE workshop, cosponsored by Tecnodoc group[4] and the Metadata Research Center (MRC)[5], enabled experimentation with both HIVE-ES[6] and HIVE[6]. Hands on activities help attendees gain experience with linked data, and understand how the fundamentals of knowledge organization and the structures that facilitate can be applied and use.

The motivation of this special issue was to collect the spirit of that symposium and HIVE workshop, and allow participants to capture their ideas and submit a formal paper for this journal. We also invited participations from the broader LOD community, particularly to capture current R&D. The topics included in the *LHT* call for submissions identified important issues and trends, and they are listed here:

- Technological or political issues relating to linking and opening vocabularies.
- Knowledge Organization Systems (e.g., ontologies, taxonomies, authority files, folksonomies, and thesauri) and Simple Knowledge Organization Systems (SKOS).
- Ontology design and development/Integration of metadata and ontologies.
- Vocabularies and the Semantic Web (metadata and applications).
- Vocabularies for scientific data, e-Science and grid applications.
- Vocabulary registries and registry services.

- Alignment and mapping diverse vocabularies.
- Social tagging and user participation in building vocabularies.
- Management of social tagging vocabularies to improve metadata systems.
- Metadata Vocabulary-Quality.
- Creation and management of sustainable vocabularies. Vocabulary curation.
- Research tools and solutions to free/open vocabularies.
- Linked Open Vocabularies and the LOD Cloud.

Following the review process, we selected five articles for this special issue of *LHT*. Two contributions represent current LD practice, and three pieces represent R&D activities. These contributions are discussed further in the following section.

Current linked data practice and trends
Reading about LD is one task, but to implement a full-scale national Library Linked Data (LLD) project is a significant undertaking. This issue opens with a piece, entitled, "MARiMbA: An insight into library linked data", by Vila-Suero and Gómez-Pérez. The work reports on an effort to convert the National Library of Spain's MARC 21 records into RDF/OWL classes, and render these bibliographic entities as LLD. The article documents the process and work-cycles, and introduces the MARiMbA application. This effort reflects activities in Sweden and the USA, where national and university libraries have published bibliographic data and vocabularies as LLD. The National Library of Spain has also applied the Functional Requirements for Bibliographic Records (FRBR), and other IFLA standards (ISBD, FRAD, FRSAD), to the catalog via this process. The method used in this project, including the MARiMbA application, constitutes an important contribution for LLD. This article also documents the value of collaboration and how domain experts and librarians, working together, can facilitate quality work. The second contribution in this cluster is a work by Delgado, Hilera, and Ruggia, "Proposal of a controlled vocabulary to solve semantic interoperability

problems in social security information exchanges". This paper highlights EU vocabulary needs and the opportunity for addressing semantic interoperability problems in social security data exchanges. The work introduces the European Interoperability Strategy (EIS) and the Electronic Exchange of Social Security Information (EESSI) project. The emphasis on interoperability is crucial in the multilingual and heterogeneous contexts, given the composition of social security data in the European countries. The paper thoughtfully reviews how classical challenges of controlled vocabularies are magnified in a heterogeneous or broad-based community. The article underscores the need to address multilingual challenges for e-government at the pan-European level is discussed.

Linked data, linking vocabularies and R&D
The second selection of articles clusters Linked Data research and development. The first contribution, "Building a platform for linked open thesauri" by Zou, presents a new service, with a new approach describing, publishing and accessing linked open thesauri. The author is adamant in his message that publishing a thesaurus as linked data are not as simple as posting a SKOS formatted RDF file on the web. Generating a LD vocabulary requires knowledge of different description levels, carefully designed URIs, and properly designed terminology services. Indeed, a linked open thesaurus should be easy to discover, because the purpose of publishing thesauri is to make them usable for broader audiences. The work presented demonstrates an innovate approach that complements traditional registries, and enabling the synchronization among registries and individual thesauri.

   The next piece, "Evaluation of semantic search in the semantic web," by Morato, Sanchez-Cuadrado, Dimou, Yadav, and Palacios, integrates theoretical and technical LOD considerations. Limitations of current semantic navigation and querying for LOD are discussed. The article points to limitation with vocabulary selection criteria and with interoperability among metadata element sets. The paper proposes an evaluation method using the DESMET methodology to address these

short comings. DESMET is a widely used procedure in Software Engineering that can aid in evaluating retrieval systems effectiveness for systems with semantic documents. The article demonstrates this approach; and the results show that semantic search engines lack of a minimal set of functionalities. The discussion highlights the need for a more detailed and useful metadata descriptions. The authors' conclusion summarizes the challenge for LOV services and tools, and emphasizes for criteria capturing trustworthiness and quality. The last article in this issue, "Requirements for vocabulary preservation and governance", is by Vandenbussche, Baker, and Vatant. This contribution presents a theoretical approach to the very real problem of vocabulary governance and preservation. Clearly, permanence and maintenance are requirements for LOD services to work well. LOD services also necessitate vocabulary preservation, management, and growth, which are known traditional KOS challenges. This article relates to Morato *et al.* paper, in the sense that it proposes a solution to current challenges via shifting practices of vocabulary managers. Solutions considered include pushing forward policy and versioning, and enabling better discoverability and usability of RDF vocabularies. The paper proposes a set of recommendations to implement these solutions. The issues and recommendations explored in this paper are reflected in the upcoming special session, "Long-term Preservation and Governance of RDF Vocabularies"[7], to be held at the DC-2013 /iPres2013 in Lisbon, Portugal.

Conclusion

The LOV Symposium and HIVE workshop in June 2012 brought together a group of international scholars as well as individuals interested in learning more about Linked Open Data and how this trend can impact vocabulary use. The event was inspiring for all, and helped to generate dialog across projects and platforms. This special issue of *LHT* acknowledges LOV Symposium and HIVE workshops, and the inspiration for a themed issue. More importantly, this issue of *LHT* provides as a record in time gathering current LD practice, trends, and

R&D. The articles gathered here convey the following points:

- National Libraries have a significant role to play in the world of LLD- linked library data, and LOD in general.
- Multilingual aspects with LD magnify traditional KOS challenges, although LOD may ultimately help solve such challenges long-term.
- Sharing LOD processes and outcomes is important, because LOD efforts are not a one-size fits all.
- Evaluation frameworks are needed to advance LOD.
- LOD preservation and governance is important to sustain LD, LOD, and LOV efforts.

The collection of articles that follow provide insight into these issues via discussion and sharing of current projects. The articles also highlight current, pressing LD problems that need to be addressed for a more successful, global LD world.

Scholarship of all dimensions, from project reporting and bona fide scientific studies, to theoretical exploration, is essential for advancing our knowledge on any topic. The overall goal of this issue is to enable scholarship of those engaged in LOD, and to provide a record that can help advance our knowledge and practice of LD. It is the editors' firm belief that via linking and opening vocabularies we can build a global Semantic Web.

Notes

1. LOV project (linked open vocabularies), available at: http://labs.mondeca.com/dataset/lov
2. NCBO-BioPortal, available at: http://bioportal.bioontology.org
3. OMR (Open metadata registry), available at: http://metadataregistry.org
4. Tecnodoc group, available at: http://tecnodoc.uc3m.es
5. Metadata Research Center, School of Information and Library Science, University of Carolina at Chapel Hill, available at: http://ils.uc.edu/mrc
6. HIVE-ES demo server, available at:

http://klingon.uc3m.es:8080/home_es.html?locale ¼ es

7. "Long-term Preservation and Governance of RDF Vocabularies" session, DC-2013 /iPres2013 in Lisbon, Portugal, available at: http://dcevents.dublincore.org/IntConf/index/pages/view/ vocPres

References

Greenberg, J. and Méndez, E. (2007), *Knitting the Semantic Web*, Haworth Press, New York, NY. Melgar, E.L., Gómez, N. and Chapman, J. (2013), "A report of the LOV Symposium (Linking and Opening Vocabularies) and SKOS-2-HIVE Workshop: two linked data/open vocabularies events in Madrid, Spain", *Cataloguing & Classification Quarterly*, Vol. 51 No. 6, available at: http://catalogingandclassificationquarterly.com/ccq51nr6.html

Méndez, E. and Greenberg, J. (2012), "Linked data for Open Vocabularies and HIVE's global framework", *El profesional de la Información*, Vol. 21 No. 3, pp. 235-244, available at: www. elprofesionaldelainformacion.com/contenidos/2012/mayo/03_esp.pdf

Richards, E.G. (2000), *Mapping Time: The Calendar and Its History*, Oxford University Press, Oxford.

Vatant, B. (2012), "). LOV stories, Part 2: gardeners and gatekeepers. The wheel and the hub: tracks in the knowledge commons [blog], 22-03-2012", available at: http://blog.hubjects. com/2012/03/lov-stories-part-2-gardeners-and.html

Further reading

Phipps, J. and Hillman, D. (2011), "The Open Metadata Registry: an update", *Bulletin of the American Society for Information Science and Technology*, Vol. 37 No. 4, available at: www.asis.org/Bulletin/Apr-11/AprMay11_Phipps_Hillmann.pdf