

Introduction to a special issue on digital libraries and the Semantic Web: context, applications and research

George Macgregor

Information Strategy Group, Liverpool Business School,
Liverpool John Moores University, Liverpool, UK

Abstract:

Purpose – The purpose of this article is to introduce the papers in the special issue which explores some of the potential, opportunities and challenges to be found in greater library and information science alignment with semantic web developments.

Design/methodology/approach – The article is a general review of the papers in the issue.

Findings – For many digital libraries or cultural institutions, the semantic web offers an opportunity to better expose valuable digital resources pertaining to research, culture or history, using common standards and technologies in a collaborative and “joined up” way. The papers in this issue “paint a rainbow”, exploring the issues through elements of case studies, reviews research and conceptual expositions and viewpoints.

Originality/value – The article emphasises how the practical implications of semantic web research or developments for digital libraries and repositories is important for LIS professionals.

Keywords Digital libraries, Worldwide web, Information management

Paper type Viewpoint

Digital library:

Digital libraries are organisations that provide the resources, including the specialised staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities (Digital Library Federation, 1998).

The Semantic Web:

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and “understand” the data that they merely display at present (Berners-Lee et al., 2001).

Digital libraries are now a “mature information service application” (Bearman, 2007), the parameters of which have been delineated by continuous definition and conceptualisation, almost since the advent of the web itself. By contrast, the Semantic Web remains a project in its infancy; an attempt to augment resources on the web with metadata about their semantics, not just their syntax. Making the web as readable to machines as it currently is to humans is the Holy Grail. The prospect of a machine-readable web creates the potential for a plethora of intelligent, information rich applications, the scope of which

precludes a dozen special issues (see Legg, 2007 for a brief review). There clearly exists numerous opportunities for digital libraries to use what the Semantic Web can offer. Perhaps the most seductive of these relates to improved user resource discovery across distributed heterogeneous collections (via search disambiguation or inference) and improved data interoperability, among many others.

This special issue presents a collection of papers aimed at exploring some of the potential, opportunities and challenges to be found in greater LIS alignment with Semantic Web developments. For many digital libraries or cultural institutions, the Semantic Web offers an opportunity to better expose valuable digital resources pertaining to research, culture or history, using common standards and technologies in a collaborative and “joined up” way. Semantic Web technologies are capable of enhancing digital libraries or repositories by facilitating improved navigation and retrieval within heterogeneous document environments, user profiling, personalisation and contextualisation, improved user interfaces and human-computer interaction. Such technologies also have the potential to solve or aid the management of problems relevant to digital libraries and the LIS community generally, such as semantic interoperability, advanced metadata and information integration, the management of large corpora of heterogeneous digital resources, and so forth.

Given the potential benefits to be gleaned and the clear synergy between LIS and the Semantic Web, much of the digital library community has been slow to assimilate Semantic Web developments, or has chosen to ignore them altogether. Whether this has been a consequence of a perception that it is the preserve of the W3C and irrelevant to the community, or whether both communities are talking at cross-purposes remains a moot point. Fortunately, it is possible to write that this position has been altering over recent years. Greenberg and Méndez (2007a) suggest the change was precipitated by the emerging notion of “Library 2.0”, first proposed by Miller (2005) and further developed in 2006 (Miller, 2006). Miller’s “call to arms” may use Web 2.0 as the hook, but its detail has awoken a wider recognition within LIS that it has to be more proactive in delivering valuable content to users and must – more generally – participate in the evolution of the web itself.

This special issue was announced in early 2007. Since then Greenberg and Méndez (Greenberg and Méndez, 2007b) have edited a special issue of *Cataloging and Classification Quarterly* on a similar theme, published simultaneously as an indispensable monograph. This special issue of *Library Review* attempts to continue the collection of papers presented by Greenberg and Méndez and build upon similarly themed conferences (Prasad and Madalli, 2007), the latter of which originally inspired this special issue of *Library Review*. Since this issue is using previous literature as a springboard, some basic Semantic Web concepts may elude exposition. Those readers requiring further detail are encouraged to consult Greenberg and Méndez (2007b), part I of which includes papers on the Semantic Web building blocks, or Legg (2007).

Digital libraries and the Semantic Web: context, applications and research

The papers contained in this special issue comprise a rainbow painted in the splendour of the Emerald paper categories[1]; encompassing elements of case studies, general reviews, research and conceptual expositions, and viewpoints.

This issue begins with a *Library Review* staple: the ANTAEUS column. Nick Joint provides musings on the special issue theme and considers the impact of Semantic Web developments on digital libraries and the practitioner librarian. Joint outlines some of the broad issues associated with the Semantic Web, together with a simple explanation of some basic Semantic Web principles. He concludes that the Semantic Web is of fundamental importance to LIS practitioners (particularly for digital repository developments and the increased exposure of open access materials) and argues that if the true Semantic Web dream is to be realised, the LIS community will have to play an active role. However – in true ANTAEUS style – Joint also provides a cautionary note and reminds us that the futuristic nature of

Semantic Web/digital library interactions requires practitioner-oriented research to ensure the implementation of meaningful and practical applications.

The LIS community has much in common with developments in the Semantic Web. This is perhaps most obvious in the Semantic Web use of ontology; an area that resonates squarely with the study of other Knowledge Organisation Systems (KOS), such as classification schemes, taxonomies, thesauri or subject heading lists. Most of the papers in the special issue explore – or at least touch on – the use of such techniques as a means of addressing issues of semantic heterogeneity.

Emanating from the STITCH project[2] and highly active in the development of the Semantic Web itself, Antoine Isaac, Henk Matthezing, Stefan Schlobach and Claus Zinn demonstrate how Semantic Web techniques can ameliorate semantic interoperability issues within the cultural heritage domain, providing users with integrated and seamless access to heterogeneous collections. Isaac et al. deploy the Simple Knowledge Organisation System (SKOS)[3] and automatic “vocabulary alignment” methods to facilitate semantic user searching and navigation across collections. Whilst Isaac et al. concede that these techniques require further refinement and experimentation, they conclude that it is the future for digital libraries and cultural heritage collections; successfully shifting us from “separate islands of collections and vocabularies to better connected networks of cultural heritage knowledge”.

With a mission to combat hunger through improved knowledge exchange, the Knowledge Exchange and Capacity Building Division (KCE)[4] of the Food and Agriculture Organization of the United Nations (FAO)[5] continue to lead developments in digital library technologies. Margherita Sini, Boris Lauser, Gauri Salokhe, Johannes Keizer and Stephen Katz discuss the development of the AGROVOC concept server (CS). The OWL[6] compliant CS is designed as a “collaborative reference platform”, providing a collection of commonly used agricultural concepts, containing a plethora of terms, definitions and relationships between terms in multiple languages. Sini et al. also detail the CS Workbench, a CS module enabling the distributed and collaborative management of the CS, in turn facilitating the re-use and extension of agricultural knowledge for increased interoperability and improved user services.

Philipp Mayr, Peter Mutschke and Vivien Petras use the German science portal “vascoda”[7] as a test-bed for implementing techniques designed to ameliorate user term vagueness and to improve result rankings for users. They demonstrate a search term recommender system employing terminology mappings and query expansion techniques. Approaches derived from scientometrics and network analysis are also deployed to better rank user result sets. Noting various synergies with Friend-of-a-Friend (FOAF)[8], Mayr et al. rank results according to the core journals of specific domains of knowledge (Bradfordizing) and by the centrality of authors in co-authorship networks. It is the conclusion of Mayr et al. that greater intersection between the digital library and the Semantic Web communities will produce improved analytical tools and interfaces for the presentation of information better adapted to users' needs; something that Mayr et al. intend to pursue in further work.

As we have noted, the LIS use and deployment of KOS resonates with the Semantic Web use of ontology. This, in turn, is an area that resonates with the work of S.R. Ranganathan, a founding father of modern library science and key founder of the Documentation Research and Training Centre (DRTC), Bangalore[9]. Continuing Ranganathan's legacy and emanating from DRTC itself, Prasad and Devika Madalli use their paper to propose a faceted infrastructure model for semantic digital libraries. This conceptual model – to be implemented and evaluated as part of a wider research study – employs a “centre-out approach”; the centre being semantic representations (based on faceted ontologies) of information. Prasad and Madalli also provide us with a useful reappraisal of why digital libraries have to be cognisant of Semantic Web developments and seize this new potential for enhanced user resource discovery.

As Isaac et al. remind us, digital libraries offer access to large amounts of heterogeneous resources. Semantic Web approaches to the management of these resources (e.g. through the use of ontology to ensure the consistent labelling and description of resources and the use of common technical standards) can ameliorate or resolve the interoperability issues often encountered by digital libraries. Resolving heterogeneity to ensure interoperability or to improve information retrieval is the focus Jürgen Krause, who picks up themes raised by Mayr et al. Krause closes the special issue with a detailed theoretical exposition, exploring the development of Semantic Web approaches in tandem with those of digital libraries. In particular, Krause focuses on the “Shell model”, an approach used in the German science portal “vascoda”, and conceptually analyses the two approaches which – although employing different techniques – seek the same goal (i.e. to resolve semantic heterogeneity). He places both under intense theoretical scrutiny, noting areas for optimism, but also those areas which require caution. Krause also notes the mismatch between the Semantic Web rationale for using controlled vocabularies instead of ontologies (by using alternatives such as SKOS, for example) and that of the LIS community, and argues that the both communities have to revert to “weaker semantic foundations” via SKOS in order to achieve the ultimate aim of the Semantic Web: interoperability.

It was the intention of the guest editor that this issue should emphasise the practical implications of Semantic Web research or developments for digital libraries and repositories; for research to be given an applied LIS focus in order to increase relevance to the usual Library Review readership. Considering all the papers together, it is hoped that the reader develops a fuller appreciation of the special issue theme. And it is hoped that both the digital library and Semantic Web community benefit from this activity. At the very least it is hoped that it stimulates further comment on areas of useful overlap between communities and how the aims of each can be furthered by increased collaboration.

Acknowledgements

The author wishes to extend his thanks to David McMenemy (Library Review editor) for his editorial guidance throughout the stewardship of this special issue, and to all those who kindly participated as peer reviewers. Their involvement was greatly appreciated.

Notes

1. Guidelines for writing structured abstracts: www.emeraldinsight.com/info/authors/writing_for_emerald/abstracts.jsp (accessed 01 December 2007).
2. STITCH Project: <http://stitch.cs.vu.nl/> (accessed 01 December 2007).
3. Simple Knowledge Organisation Systems (SKOS): <http://www.w3.org/2004/02/skos/> (accessed 01 December 2007).
4. Knowledge Exchanges and Capacity Building Division (KCE): www.fao.org/gi/gil/index_en.asp (accessed 01 December 2007).
5. Food and Agriculture Organization of the United Nations (FAO): www.fao.org/ (accessed 01 December 2007).
6. Web Ontology Language (OWL): www.w3.org/2004/OWL/ (accessed 01 December 2007).
7. vascoda: <http://www.vascoda.de/> (accessed 01 December 2007).
8. Friend-of-a-Friend (FOAF) Vocabulary Specification: <http://xmlns.com/foaf/spec/> (accessed 01 December 2007).
9. Documentation Research and Training Centre (DRTC): <http://drtc.isibang.ac.in/DRTC/> (accessed 01 December 2007).

References

- Bearman, D. (2007), "Digital libraries", in Cronin, B. (Ed.), Annual Review of Information Science and Technology, Vol. 41, Information Today Inc., New Jersey, NJ, pp. 223-72.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001), "The Semantic Web", Scientific American Magazine, May, available at: www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 (accessed 1 December 2007).
- Digital Library Federation (1998), "A working definition of digital library", available at: www.clir.org/diglib/dldefinition.htm (accessed 1 December 2007).
- Greenberg, J. and Méndez, E. (2007a), "Introduction: toward a more library-like web via semantic knitting", Cataloging & Classification Quarterly, Vol. 43 No. 3-4, pp. 1-8.
- Greenberg, J. and Méndez, E. (Eds) (2007b), Knitting the Semantic Web, The Haworth Information Press, Binghamton, NY (published simultaneously as Cataloging & Classification Quarterly, Vol. 43 No. 3-4).
- Legg, C. (2007), "Ontologies on the Semantic Web", in Cronin, B. (Ed.), Annual Review of Information Science and Technology, Vol. 41, Information Today Inc., New Jersey, NJ, pp. 407-51.
- Miller, P. (2005), "Web 2.0: building the new library", Ariadne, No 45, available at: www.dlib.org/dlib/april06/miller/04miller.html (accessed 1 December 2007).
- Miller, P. (2006), "Coming together around Library 2.0: a focus for discussion and a call to arms", D-Lib Magazine, Vol. 12 No. 4, available at: www.dlib.org/dlib/april06/miller/04miller.html (accessed 1 December 2007).
- Prasad, A.R.D. and Madalli, D.P. (Eds) (2007), Proceedings of the International Conference on Semantic Web and Digital Libraries (ICSD-2007), 21-23 February 2007, Documentation Research and Training Centre, Bangalore.