Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-based Data Models

Getaneh Alemu

Brett Stevens

Penny Ross and

Jane Chandler
University of Portsmouth
Portsmouth, United Kingdom

Meeting: **92 —** *New futures for bibliographic data formats: reflections and directions* **— UNIMARC Core Activity**

Abstract:

Contemporary metadata principles and standards tended to result in document-centric rather than data-centric; human-readable rather than machine-processable metadata. In order for libraries to create and harness shareable, mashable and re-usable metadata, a conceptual shift can be achieved by adjusting current library models such as Resource Description and Access (RDA) and Functional Requirements for Bibliographic Records (FRBR) to models based on Linked Data principles. In relation to technical formats, libraries can leapfrog to Linked Data technical formats such as the Resource Description Framework (RDF), without disrupting current library metadata operations. This paper provides six key recommendations for libraries and standards agencies. These include rising to the challenges and embracing the opportunities presented by current technological trends, adopting minimal requirements of Linked Data principles, developing ontologies, deciding on what needs to be retained from current library models, becoming part of the Linked Data cloud, and developing mixed-metadata (standards-based and socially-constructed) approaches. Finally, the paper concludes by identifying and discussing five major benefits of such metadata re-conceptualisation. The benefits include metadata openness and sharing, serendipitous discovery of information resources, identification of zeitgeist and emergent metadata, facet-based navigation and metadata enriched with links.

Keywords: *metadata, metadata standards, MARC, Linked Data, RDF, socially-constructed metadata, mixed-metadata approaches.*

Introduction

The principles, standards and protocols currently in use in libraries for supporting metadata functions have had an enduring history (Denton, 2007; Dunsire, 2009; IFLA, 2009; Lubetzky, 1953; Wright, 2007). The history of library cataloguing spans many thousands of years, dating back to the days of the Library of Alexandria in the third century B.C. where the librarian Callimachus prepared a systematic bibliography of Greek literature (Day, 2005; Wright, 2007). Morville (2005) takes the history of cataloguing even farther back in time to the Assyrian Empire where King Assurbanipal built a library of 30,000 clay tablets in 650 B.C. However, the history of standardised efforts at cataloguing started in the 19th century only (Coyle & Hillmann, 2007; Denton, 2007; Weinberger, 2005, 2007; Wright, 2007). The main pioneers of modern cataloguing include Sir Anthony Panizzi, Charles Cutter, Melville Dewey, Paul Otlet, S.R. Ranganathan and Seymour Lubetzky (Denton, 2007; Lubetzky, 1953; Wright, 2007). Drawing upon the works of these pioneers, regional and international consortia embarked on projects aimed at instituting rigorous cataloguing principles and rules, which subsequently resulted in such standards publications as the Paris Principles, 1961; the Anglo-American Cataloguing Rules (AACR), 1967; MAchine-Readable Cataloguing (MARC), in the late 1960s; International Standard Bibliographic Description (ISBD) for Monographic Publications,1971; Functional Requirements for Bibliographic Records (FRBR), 1996; and Resource Description and Access (RDA), 2010 (Denton, 2007).

However, as the size of collections in digital libraries continues to grow, contemporary metadata principles and standards have come under intense scrutiny (Alemu, Stevens, & Ross, 2012; Coyle, 2010; Coyle & Hillmann, 2007; Lagoze, 2010; Mathes, 2004; Shirky, 2005; Veltman, 2001; W3C, 2011; Weinberger, 2005, 2007). Critics contend that current library standards have brought along some of the constraints inherent in the traditional card catalogue system and, hence, are liable to failure to scale as well as to interoperate within the present day information landscape (Coyle, 2010; Coyle & Hillmann, 2007; Weinberger, 2005, 2007). The metadata generated through the use of contemporary metadata standards and technical formats is mainly attuned to human consumption rather than machine processing. This is attributed to both conceptual and technical limitations of the standards and technologies used in libraries. Coyle (2010) argues that “the library catalogue has been the sole context for library data” hence, fails to interoperate with external information providers.

The challenges for principles, standards and protocols could be looked at from two perspectives: conceptual and technical. The conceptual underpinnings of contemporary metadata standards, such as FRBR and RDA, have arguably resulted in metadata records as documents attuned to human consumption rather than machine processing (Coyle & Hillmann, 2007). Furthermore, when such metadata principles are implemented using technical formats such as MARC, the resultant records exhibit problems of metadata duplication, data inconsistency, lack of granularity and complexity (Coyle, 2010; Coyle & Hillmann, 2007; Day, 2000; Guenther & McCallum, 2003; Tennant, 2002). Even though the call for an end to the use of MARC has been proclaimed as long overdue (Tennant, 2002), the standard remains the dominant metadata structure used in libraries to this day. This can directly be attributed to several causes, including the fact that MARC is deeply embedded in library systems and functions and, thus making any changes would simply become too difficult and expensive; or that MARC, after all, is ‘adequate’ enough for libraries and serves its purposes; or it may be that alternative formats, including eXtensible Markup Language (XML), fail to deliver the additional functionality required to merit and justify the changeover. However, there are still grave doubts regarding the adequacy of MARC, espoused by several metadata experts who assert that the standard is not suitable for machine process-able and actionable metadata (Coyle, 2010; Coyle & Hillmann, 2007; Wallis, 2011a, 2011b).

Suggested alternatives include the adoption of Linked Data. As its inventor, Berners-Lee, et al (2001), state, the original web has conceptually been document-centric in which the links do not carry any semantics with them and when implemented with technical formats such as Hyper-Text Mark Up Language (HTML), the resultant web pages are more attuned to human consumption, rather than machine processing. Conversely, whilst the web has been exceptionally efficient for sharing documents and creating possibilities for collaboration, a document requires human intervention for understanding its semantics once it is presented and displayed on user's computer screen. Put simply, machines cannot make sense of such documents. To alleviate this limitation, Berners-Lee, et al (Berners-Lee, 1997; Berners-Lee, 1998; Berners-Lee, et al., 2001) came up with the concept of Linked Data (Berners-Lee, et al., 2001). The technologies to implement Linked Data include RDF, RDFS, SPARQL, and OWL (Allemnag & Hendler, 2008; Berners-Lee, 1998; Decker et al., 2000; W3C, 2004a, 2004b). According to Allemnag and Hendler (2008) "the main idea of [Linked Data] is to support a distributed web at the level of the data rather than at the level of presentation [documents]. Instead of having one webpage point to another, one data item can point to another, using global references called URIs". Despite the growing interest in Linked Data, there is still an ongoing, widespread debate as to whether incremental changes made to traditional library-centric conceptual principles such as RDA and FRBR. The same question also holds true whether the change in technical formats from MARC to RDF should be evolutionary or whether there is compelling reason for abandoning some or all of these traditional formats altogether and adopting these new approaches (Coyle, 2010; Coyle & Hillmann, 2007; Marcum, 2011; Styles, 2009; Styles, Ayers, & Shabir, 2008; W3C, 2011; Wallis, 2011a, 2011b).

Another important challenge that standards authorities and libraries should consider, in relation to current metadata re-conceptualisation, is the implication of Web 2.0 approaches in the creation and utilisation of metadata in digital libraries. Even though, the role of Web 2.0 for libraries has widely been acknowledged, as Evans (2009) notes "*its implications are not yet fully realised, especially for librarianship*" perhaps because, as Lagoze (2010) argues, there exists a conceptual incompatibility between traditional library models and the emergent Web 2.0 approaches. Socially-constructed metadata (Web 2.0) approaches are criticised for being flat, one-dimensional and plagued with inconsistencies (Gruber, 2007, 2008), whilst on the other end standards-based metadata approaches are criticised for their rigid hierarchies and failure to represent the vocabularies of users (Shirky, 2005; Veltman, 2001; Weinberger, 2007). Conversely, Gruber (2007) and Morville (2005) contend that expert-controlled and user-generated metadata should not be considered as opposing approaches suggesting that they not only co-exist but also mutually shape each other. In view of optimally benefiting from the capabilities of Linked Data (Web 3.0), a mixed metadata approach that consists of not only the physical description of information objects (such as author, title, ISBN, subject, format, etc) but also incorporate elements describing its socio-cultural facets (user tags, comments, reviews, links, ratings, recommendations) is crucial. In other words, the representation of the social space of metadata (Web 2.0) should be considered as equally important as the recording of standardised and objectivistic metadata elements (Web 1.0) that have hitherto been used to characterise the physical characteristics of information objects. This is especially relevant for metadata richness where both user-generated (socially-constructed) metadata is harnessed along with the standards-based, librarian created metadata.

This paper discusses the challenges of contemporary metadata principles and standards and suggests the benefits and implications of making a conceptual shift from document-centric to data-centric metadata approaches. Two major broad classes, namely metadata principles (such as RDA and FRBR) on one hand, and technical record formats (such as MARC) on the other, are identified and, recommendations are suggested as to which of these classes should

be modified, adapted, or retained so as to comply with Linked Data principles.

Linked Data principles

Linked Data, as the name indicates, is a data model that identifies, describes, links and relates structured data elements, analogous to the way relational database systems function, albeit the fact that the former operates at a web scale. The overall purpose of Linked Data is facilitating the re-usability, cross-linking, integration and sharing of data (Berners-Lee, 2009; Shadbolt, 2010; W3C, 2011). Berners-Lee (2007) notes that "adding a page provides content, but adding a link provide the organization, structure and endorsement to information on the Web which turn the content as a whole into something of great value."

It is important to note that Linked Data is a meta-model wherein it provides a framework to defining, designing, developing and maintaining schemas and vocabularies of any kind and size in a given domain. This in effect means institutions, such as libraries, need not necessarily abandon existing metadata standards, controlled vocabularies, authority lists and legacy metadata. However, in order to interoperate with the world of Linked Data cloud, institutions are expected to adopt the underlying Linked Data.

Linked Data principles start from the most fundamental component of Linked Data, i.e., the use of globally unique URI as names for distinctively denoting such things as information objects, people, places, and events (Berners-Lee, 2009). Other essential principles include an ability to dereference a given URI, using Hypertext Transfer Protocol (HTTP), in order to retrieve relevant information as to what this particular URI refers to; use of data formats, such as RDF/XML, use of vocabulary definition languages, such as RDFS and OWL, and use of query language known as SPARQL; and last but not least is the incorporation of inbound and outbound links within and outside the data sets, thereby enriching the data and providing contextual significance.

Linked Data for library metadata modelling

Linked Data and its associated technologies have significant roles for metadata modelling, encoding, representation and sharing. The utilisation of URIs for metadata element names, labels, and relations alleviates naming and identification conflicts in the use of elements. It has been indicated that RDF's simple data model enables the creation of semantic links among information resources (Coyle, 2010; Coyle & Hillmann, 2007; Day, 2000, 2003a, 2003b; Helen, 2010; Nilsson, 2010; Rothenberg, 2008; Styles, et al., 2008; W3C, 2011; Wallis, 2011a; Wilson 2010).

Closed- versus open-systems assumption

Linked Data principles enable and operate in an open, dynamic, and interactive system. However, library standards have mostly operated in a closed and static environment detached from the general web information landscape. According to Allemnag and Hendler (2008), Linked Data offers a distributed data model whereby "anyone can say anything about any topic" resulting in "variations and disagreements" about the meaning of entities. As the authors attest "an open world in this sense is one in which we must assume at any time that new information could come to light, and we may draw no conclusions that rely on assuming that the information available at any one point is all the information available".

In a digital library context, the open-systems assumption lends itself amenable to new metadata contributions from diverse points of views, thereby immensely benefiting from what are known as network effects. Employing URIs, ontologies - which are explicit specification of concepts in a given domain using vocabularies and relations between them (Gruber, 1993) - can be mixed, matched and merged.

Current status of Linked Data use in libraries

There seems to be an urge, especially among public sector organisations such as national libraries, to make accessible their bibliographic data freely and openly (Wilson, 2009). 2011 witnessed the publication of two definitive reports, both addressing the importance of opening library metadata in formats that are easily accessible and re-usable (Library of Congress, 2011; W3C, 2011). As Haslhofer and Isaac (2011) point out Europeana and its affiliated national and regional libraries have consented to go the Linked Data way. The British Library has already developed a Linked Data model (The British Library, 2011). Similarly, Europeana has revised its metadata model from the Europeana Semantic Elements specification to the Europeana Data Model (EDM), the latter being more attuned to Linked Data (Doerr et al., 2012; Haslhofer & Isaac, 2011).

Wilson (2010) points out the decision by the British Library to release its bibliographic data freely is in accordance with the commitment of the Government of the UK towards transparency and accountability. The author also notes that part of the move to Linked Data can be attributed to the growing expectation, expressed by library users, to see the library keeping up with cutting-edge technological trends. The British Library data model incorporates and uses several existing URI-based vocabularies and ontologies such as the Virtual International Authority File (VIAF), Library of Congress Subject Headings (LCSH), Lexvo (URI referenced controlled list of characters, words, terms), GeoNames (geographical database), MARC country and language codes, Dewey.info (top level classes of Dewey Decimal Classification) and RDF book mash-up (information about books and their authors) (The British Library, 2011).

Slow adoption of Linked Data for library metadata

The W3C Library Linked Data Incubator Group (2011) report acknowledges the slow pace with which libraries are adopting Linked Data, whilst providing strong cases and recommendations to libraries to embrace the Linked Data principles. The report emphasises the importance of making bibliographic library data openly and freely accessible in a form that is "shareable, extensible, and easily re-usable" (W3C, 2011). It also points out that "library standards (MARC, Z39.50) are designed only for the library community," making it difficult for outsiders to reuse and recombine it with other data. This report is in complete accord with assessments made by Coyle (2010) and Styles, et al (2008), who highlight present-day library metadata development and adoption challenges and advocate addressing these challenges by embracing web-based standards that favour re-usability, cross-linking and sharing of metadata.

Shadbolt and Hall (2006) commented that uptake is "about reaching the point where serendipitous reuse of data, your own and others' becomes possible." They regret that there has been little effort expended at promoting a more widespread use of the Semantic Web, which can partly be attributed to the emphasis placed on "languages, formalisms, standards, and semantics". Weinberger (2012) believes that "while the original Semantic Web emphasised building ontologies that are 'knowledge representations' of the world, it turns out that if we go straight to unleashing an abundance of linked but imperfect data, making it widely available in standardised form, the Net becomes a dramatically improved infrastructure of knowledge".

Given the substantial amount of investments and effort that have already been expended by organisations, such as IFLA, the Library of Congress, OCLC and many other national libraries, on developing and maintaining the MARC record format, the transition to Linked Data is not a trivial manner. The fact that libraries have significant amounts of valuable metadata records in MARC formats is indisputable. For example, starting from the 1950s, the British Library has made its national bibliographic records available to outsiders through

subscription-based access, initially using printed formats and then, following the implementation of library automation, using the MARC standard (The British Library, 2011; Wilson 2010).

Wallis (2011b) observes that even though there now are a few national libraries and regional cultural heritage institutions (such as the CENL and Europeana) who have publicly declared their commitment to go the Open Linked Data route, fundamental challenges to re-using their data remain at large. He contends that, whilst converting legacy library records to RDF is crucial, the resultant datasets still retain the library-domain-specific languages and terminologies which would be "impenetrable for anyone who is not a librarian looking for useful data".

Challenges to adopt Linked Data in libraries

The road towards adoption of Linked Data is not without its challenges. When it comes to libraries, there are three main challenges that arise. Firstly, the extensive use of the MARC standard as the basis for current library management systems as well as the legacy metadata that has been created over the years. The MARC format, although dominant, is considered to be a record and document-centric metadata structure, rather than being an actionable data-centric format (Coyle, 2010; Coyle & Hillmann, 2007; Styles, 2009; Styles, et al., 2008). There are literally billions of records in MARC formats; an attempt at making the slightest move away from it would have huge implications in terms of resources. Even though the limitations of the MARC standard have been exposed, ever since the early years of 2000s, libraries have continued using it up until now (Tennant, 2002). The issue at present is not that libraries and standards bodies are unaware of MARC's limitations but perhaps the fact that alternative formats, such as XML, have not been more suitable as a replacement.

A second challenge, singled out by the W3C Library Linked Data Incubator Group (W3C, 2011), is the terminological disparity that exists between library and web-based standards. In relation to this, Styles (2009) cites ambiguities that may arise from using the FRBR model. Styles (2009) contends that "nobody talks about Works, Expressions and Manifestations [which are the core constructs of the FRBR model], so why describe our data that way?" Holding a similar view, Wallis (2011a), proposes a simplistic model that focuses on the object (item) such as a specific book than its abstractions as a Work, Expression and Manifestation. Both authors concur that such abstractions are not in everyday use, by publishers, cataloguers and library users. Wallis especially recommends that the library and the Linked Data community should work in concert and bridging such differences so as to facilitate the reusability and extensibility of library data by outsiders to its domain. Wallis argues that the new initiatives to develop library standards, such as RDA and FRBR, should cater for simplicity while exploiting the metadata richness that would be possible through the use of Linked Data.

The third and important challenge confronting potential adopters is the complexity of Linked Data technologies such as RDF/XML, RDFS, OWL and SPARQL. There is an apparent lack of tools and applications for creating Linked Data in libraries. Berners-Lee has remarked that "the [current] web has grown because it's easy to write a web page and easy to link to other pages" (Berners-Lee, 2007). In regard to this, it is imperative that the said technologies be made relatively easy to learn and use, analogous to the simplicity of creating HTML pages during the early days of the web. As things stand now, these technologies are generally too complicated for people outside the domain of the Linked Data community. For a wider uptake to take place, anyone with basic skills of website design should be able to create a page based on these Linked Data technologies. Last but not least, current efforts to adopt Linked Data principles and publishing bibliographic data focus on standards-based legacy metadata and tend to ignore socially-constructed (user-driven) metadata approaches.

Recommendations for making a conceptual shift
In what follows, six main recommendations are indicated in order for libraries and standards agencies to address the three main challenges identified above. These recommendations include:

- Libraries shall rise to the challenges presented by Linked Data
- Adopt the minimum requirements of Linked Data principles
- Develop ontologies: entities (classes), elements (properties) and values (instances)
- Deciding on what needs to be retained
- Become part of the Linked Data web
- Adopt a mixed-metadata approach

Libraries shall rise to the challenges presented by Linked Data
In light of current Web 2.0 and Web 3.0 paradigms, and with the aim of benefiting from the globally unique identification mechanism, cross-linking and re-usability features afforded by Linked Data principles, libraries, standards authorities and library-affiliated agencies such as IFLA should rise to the challenge and embrace the opportunities presented by current technological trends. In connection with this, Coyle suggests that "the change that we must address is that the Web is increasingly the source of information for searchers and researchers and that the library needs to be interconnected with that web of data". The report by the W3C's Library Linked Data Incubator Group (W3C, 2011), cited earlier, also recognised the need for a move from domain-specific library standards and record formats (such as MARC, Z39.50) to Linked Data standards and data formats such as RDF. Transferring legacy library metadata from its OPAC context to the web context, will, nonetheless, be a massive challenge, hence close cooperation and collaboration between the various stakeholders, such as libraries, archives, and museums, publishers, and standards agencies, is bound to be crucial.

Adopt the minimum requirements of Linked Data principles
There is a need for libraries and standards-authorities to adopt the minimum requirements of Linked Data principles. Hence, adapting legacy library metadata to Linked Data structures, using the RDF/XML serialisation format, will be mandatory. The W3C's Library Linked Data Incubator Group (2011) report indicates libraries, standards agencies and vendors would benefit "by broadening their scope or liaising with Linked Data standardization initiatives" (W3C, 2011). The Use of URIs, HTTP URIs, RDF and linking within and out of data sets would constitute conformance with these minimal requirements of Linked Data principles.

Develop ontologies: entities (classes), elements (properties) and values (instances)
There will also be a need for making a conceptual shift from perceiving library metadata as a document or record to what Coyle (2010) terms as actionable metadata, i.e., one that is machine-readable, mash-able and re-combinable metadata. To be able to accomplish this, it is mandatory that metadata models (ontologies) be developed beforehand. To this end, FRBR and RDA should assume centre stage, to serve as high-level library data models.

According to Horridge, Knublauch, Rector, Stevens, and Wroe (2004) ontologies help capture knowledge about a certain domain with concepts and relationships amongst them. As Noy and McGuinness (2000) describe ontologies enable sharing a common understanding of the structure of information among people or software agents; facilitate reuse of domain knowledge by making domain assumptions explicit, separating domain knowledge from operational knowledge, and aiding the analysis of domain knowledge. According to Berners-

Lee, et al (2001), the word ontology does not refer to “a theory about the nature of existence, of what types of things exist” instead the definition is adopted from its use by the artificial intelligence community. Hence the term ontology is defined as “a document or file that formally defines the relations among terms” (Berners-Lee, et al., 2001). According to Horridge, et al (2004) “ontologies are used to capture knowledge about some domain of interest”. However, the most widely cited definition of ontology is given by Gruber (1993) as “an explicit specification of a conceptualization” (Gruber, 1993).

According to Noy and McGuinness (2000), developing an ontology involves the process of defining classes, subclasses, properties and instances (values). It is strongly recommended that the development of library metadata ontologies should utilise the constructs as the ones adopted by Semantic Web technologies when specifying high level definitions of vocabularies. Furthermore, it is also proposed that the development of ontologies be centralised (authoritative) and preferably left for such established institutions as IFLA, Joint Steering Committee for Development (JSC) of RDA, the British Library, the Library of Congress, and OCLC, although the contributions of decentralised, but collaborative initiatives, by individuals and communities cannot be ruled out.

Deciding on what needs to be retained

A distinction needs to be made between library models and record formats, as libraries would benefit immensely by embracing the principles that underpin library metadata, as manifested in existing library-centric models such as RDA and FRBR. However, libraries could benefit by abandoning obsolete formats and search and retrieval protocols (e.g., MARC and Z39.50, respectively). In regard to this, the authors of this paper are persuaded that the call to abandon existing metadata models (such as RDA and FRBR) is against the ethos of the Semantic Web. Furthermore, it is recommended that international organisations such as the International Federation of Library Associations and Institutions (IFLA) as well as regional standards authorities and libraries should, in general, take the lead and exert concerted effort in initiatives aimed at making contemporary models and standards compatible with Linked Data principles. To optimally benefit this process, libraries should make the most of their knowledge and their expertise in controlled vocabularies such as LCSH, Medical Subject Headings (MeSH), and authority lists.

Continuing from the above, this paper has identified two prerequisites, namely, the benefits of re-conceptualising existing models, such as RDA and FRBR, so as to make them compatible with Linked Data principles; and libraries should convert existing MARC-compatible metadata records to RDF/XML serialisation formats, thereby slowly moving out of MARC and its derivatives. In other words, there is a need to make a conceptual shift from record-centric metadata formats to RDF-based data formats.

Become part of the Linked Data web

When datasets are released under the Linked Open Data principle, the process generates a globally connected Semantic Web wherein users and applications are able to identify, select, adapt, use and re-use data. The data cloud, as it stands at present, looks rather patchy since data belonging to governments, businesses, and libraries tend to be stored and maintained in disparate, walled silos (W3C, 2011). As the W3C (2011) report has remarked, “much of the content in today's Linked Data cloud is the result of ad-hoc, one-off conversions of publicly available datasets into RDF and is not subject to regular accuracy checks or maintenance updates.” This implies that much more concerted effort would be required to develop and maintain persistent and reliable namespaces, vocabularies (ontologies) as well as datasets. With a proper adoption of Linked Data principles, libraries would not only be part of this data cloud, but would also be the dominant players as they hold tremendous amount of legacy bibliographic and authority-list data, even though most of these wealth is presently locked up

in silos and library-specific formats. One of the ways for ensuring persistency and reliability of the data cloud is to reusing existing ontologies (vocabularies) and URIs. Some of the existing namespaces that include RDFS, OWL, Dublin Core, VIAF, the Friend of a Friend (FOAF), Schema.org, BLT (British Library Terms), Lexvo, GeoNames, MARC country and language codes, Dewey.info and the RDF Book Mashup.

Adopt a mixed-metadata approach

Concurring to the argument that ontologies (taxonomies) and Web 2.0 approaches would complement than oppose each other (Gruber, 2007; Morville, 2005; Wright, 2007), it is important that any metadata model re-conceptualisation leverages the best of these various paradigms. For instance, embedding collaboration and user participation in library OPACs makes the latter a Web 2.0 service. In a similar manner, Web 2.0 and Web 3.0 technologies are complementary, and hence could be leveraged to provide satisfying user experiences by harnessing the significant volume of data collected from the former and the structure (modelling) and technological capabilities afforded by the latter (Gruber, 2007, 2008). Gruber (2007) argues that socially-constructed metadata approaches are one-dimensional, and hence, plagued with inconsistency and lack organisation, whilst standards-based metadata approaches are criticised for failing to represent the vocabularies of users. Similarly, Morville (2005) also asserts that the social web and the Semantic Web would co-exist and mutually shape each other. Elaborating on the point, Morville (2005) contextualises the theory of "Pace Layering" from Stewart Brand's "How Buildings Learn" (Brand, 1994; Brand, 1999) and argues that "taxonomies and ontologies provide a solid semantic network that connects interface to infrastructure" whilst folksonomies are overlaid on the taxonomic metadata infrastructure providing it with the fast-moving and volatile vocabularies of users (see also Campbell & Fast, 2006; Smith, 2008). In moving from one paradigm to another, libraries would selectively decide which of the principles, standards and formats should be carried over to the next paradigm and which ones to abandon. For instance in moving from librarian-controlled (standards-based) to a socially-constructed metadata approach, it is found important that librarians lower the barriers to entry for new users, in order to enable them participate in the creation of metadata, which implies to relinquish the stringent metadata quality controls.

Benefits of linked library metadata for digital libraries

There are many benefits of adopting Linked Data principles in library standards, but five key ones are indicated and discussed below. These benefits include:

- Metadata openness and sharing
- Facilitate serendipitous discovery of information resources
- Linked metadata to identify resource usage patterns, zeitgeist and emergent metadata
- Facet-based navigation
- Metadata enriched with links

Metadata openness and sharing

One of the core competencies of Web 2.0 technologies is its architecture that facilitates participation (O'Reilly, 2005), which has lowered the barrier to entry for contributors. Central to this architecture of participation is the importance of embracing openness to sharing and collaboration, based on mutual trust (Alexander, 2006; Anderson, 2006, 2010; Shirky, 2005; Tapscott & Williams, 2010; Udell, 2004; Weinberger, 2005, 2007). Web 2.0 has a lot to do more with attitude and culture than technology (Miller, 2005). In other words, Web 2.0 is not, as such, a new invention; rather it is characterised by the participatory and collaborative

culture that has been built around Web 1.0 technologies, enabling users to become proactive content creators and consumers. In their bestselling book "Wikinomics", Tapscott and Williams (2010) assert that "due to the deep changes in technology, demographics, business, and the world, we are entering a new age where people participate in the economy like never before". They identify openness as one of the major principles underlying mass collaboration, along with peering (self organised peer networks of contributors), sharing, and acting globally, each of which expanding the possibilities to tap into a much larger pool of talent. They also note that "openness is associated with candour, transparency, freedom, flexibility, expansiveness, engagement and access". According to them, current economical, social and technological trends suggest that openness does not necessarily correlate to intellectual property infringements. They further contend that the culture of openness and continued recognition of its potential benefits has compromised the "conventional wisdom that says companies compete by holding their most coveted resources close to the chest" (Tapscott & Williams, 2010).

Alexander (2006) asserts that the multi-directional flow of information, between producers and consumers and across domains, servers, and machines, necessitates opening up information silos and fostering shared services. The author contends that even commercial sites, such as Amazon.com, permit their users to "harvest ISBN numbers from its listings". He attests that "openness remains a hallmark of this emergent movement, both ideologically and technologically." Contextualising this to libraries, Miller (2005) argues that the principles of Web 2.0 are predicated on the notion of the liberating data, which in turn allows data to be "exposed, discovered, and manipulated" in a multitude of ways, thereby creating unimaginable possibilities for repurposing and reusing the data.

Whilst Linked Data can be made usable without it necessarily being open (Cobden, Black, Gibbins, Carr, & Shadbolt, 2011; Shadbolt, 2010; W3C, 2011), as Berners-Lee (2010) emphasises, opening data brings forth numerous benefits to society. In his TedTalk, Berners-Lee (2010) re-iterates his vision of Linked Data by citing a number of international, regional and community-based initiatives and projects that have adopted Linked Data principles and thereby made their data openly available using URIs and RDF technologies. As the author highlights, open data can be re-used in an unimaginable number of ways. For instance, open government data enables tax payers to check how, where and for what purposes their money has been spent by their elected representatives. As Berners-Lee (2010) stresses the momentum for opening up data "has only just started".

The concept of open linked data has far reaching consequences for the way metadata in libraries is created, accessed, shared and re-combined. Linked library Metadata can be "freely usable, reusable, and redistributable [...] by using globally unique identifiers to designate works, places, people, events, subjects, and other objects or concepts of interest, libraries will allow resources to be cited across a broad range of data sources and thus make their metadata descriptions more richly accessible." Adopting Linked Data principles and making library metadata available for re-use would eliminate unnecessary duplication of data that is already available elsewhere, through reliable sources. It enables several libraries to cross-link between their resources, thereby facilitating collaboration. Finally, open Linked Data unlocks the prospect of library metadata visibility through links, whilst simultaneously allowing it to be linked to non-library information sources such as Google, Wikipedia, LibraryThing, CiteULike, and Amazon.

Facilitate serendipitous discovery of information resources

A significant number of notable scientific innovations have reportedly been the results of serendipitous discoveries made by scientists (Campa, 2008; Stoskopf, 2005; Zuckerman, 2010). Such unexpected and unanticipated discoveries could be pleasantly surprising to the

scientist. As Stoskopf (2005) notes such discoveries more often happen to the open and curious observer rather than to the scientist who may consider such cues as distractions. He asserts that "basic curiosity and observation are necessary precursors to scientific discovery. It should be recognized that serendipitous discoveries are of significant value in the advancement of science and often present the foundation for important intellectual leaps of understanding" (Stoskopf, 2005). The term serendipity was coined in 1754 by the English historian Horace Walpole (Merton & Barber, 2004). In one of his forty-six letters to his friend Horace Mann, Walpole wrote:

> "I shall endeavour to explain to you: you will understand it better by the derivation than by the definition. I once read a silly fairy tale, called the three Princes of Serendip: as their Highnesses travelled, they were always making discoveries, by accidents and sagacity, of things which they were not in quest of: for instance, one of them discovered that a mule blind of the right eye had travelled the same road lately, because the grass was eaten only on the left side, where it was worse than on the right" (Merton & Barber, 2004).

Serendipitous information discovery has also relevance as a topic of research in library and information science; even though, research on the issue is very scant (Foster & Ford, 2003). As Foster and Ford (2003), note, current models regarding user information seeking behaviour (such as Ingwersen, 1996; Wilson and Walsh, 1996; Wilson, 1997; Kuhlthau, 1993; Saracevic, 1996 & Spink, 1997) provide no consideration of serendipity. As Foster and Ford (2003) report, the closest reference to the concept are the ones given in Wilson and Walsh, 1996 and Wilson, 1997, both of which treat the phenomena as "passive attention" and "passive search". Foster and Ford (2003), explore the topic in greater detail, asserting that that serendipitous access to information reinforces or strengths "the researcher's existing problem conception or solution" and helps the researcher to chart new territories in the quest for exploration and discovery. Whilst noting its relevance for fully understanding users' information seeking behaviour, Foster and Ford (2003) note the problematic nature of studying serendipity, which can partly be attributed to the difficulties related to systematic control as well as prediction of how it improves user experiences. As library services go online, users rely on searching databases for getting access to predetermined information sources, users would lose the opportunity for a serendipitous discovery of an information source which could have been gained through browsing the physical library shelf (Foster & Ford, 2003; Massis, 2011). It is to be noted that library databases, with a predetermined item search interface, are very unlikely to offer the possibility of accidentally pertinent search results, as that is not the way they are designed.

Open access to physical library shelves tend to make serendipitous discovery of a book or a journal possible. Those libraries that enforce closed access to the stacks of books offer very little in terms of such accidental discoveries of books on shelves. OPACs, which are designed to answer predetermined item searches (e.g., using author's name or title), are analogous to closed access practises to books on physical library shelves, as both rely on matching users' queries with records available in the collection when responding to a user's quest for an information source. Put simply, the element of pleasant surprising discovery which could be realised by physical library shelves could altogether vanish if digital libraries were to solely focus on matching user queries. This then begs the question, 'how can such serendipity be incorporated in the design of digital libraries?' Since access to information objects in digital libraries depends upon the use of metadata, part of the solution lies with the choice of metadata approaches. It is therefore important to explore how socially-constructed metadata approaches could be utilised along with legacy library metadata to improve user experiences through serendipitous discovery. In a similar vein, the role of Linked Data in fostering serendipity needs to be explored.

Through the release of Linked Data, libraries would benefit from the possibility of linking to myriad but relevant databases and sources hence allowing collocation of various information objects from disparate sources. The use of URIs helps in assembling such resources from heterogeneous sources. Hence, allowing users to jump seamlessly from one source to another re-introduces serendipitous access to digital libraries, in the same way that open library shelves provide for users.

Linked Metadata to identify resource usage patterns, zeitgeist and emergent metadata

The more a metadata system reflects the variations and coinages in the nomenclatures of objects (e.g., Tsunami, Czech Republic for the former Czechoslovakia or Russia for the former Soviet Union; or Mac for Macintosh), the better it supports discoverability of information objects. This is because "the vocabulary of the information professional/librarian may be at odds with the language of content creators and users, obscuring the very information the taxonomy should reveal" (Barbosa, 2008). Since the main objectives of affixing metadata to information objects it to enable its discoverability, metadata should be judged by its relevance in meeting this objective, rather than the way it depicts an information object in an objective and ontological manner. Terminologies change over time, hence, the metadata should be able to be cognisant of and reflect such changes. In the age of the print catalogue, libraries tried their best to update and maintain the currency of their card catalogues. The resource implications of updating records meant that even though new terminologies and nomenclatures had been introduced and had also been brought to the attention of librarians, the necessary updates had to be delayed for a significant amount of time. This problem is not particularly unique to librarians, as the same holds true when updating a factual error in a print encyclopaedia entry.

A digital library system underpinned by mixed metadata approaches (standards-based and socially-constructed metadata) and augmented by the additional semantics and technological capabilities of Linked Data (URI, RDF, OWL, SPARQL) would be better situated to provide a richer and more complete metadata descriptions of information objects, more accurately reflecting the various interpretations and terminologies employed by users. Additionally, Linked Data would help facilitate the possibility of an analysis of the usage patterns of resources (such as number of clicks, total downloads, frequently used tags (tag clouds), average ratings, and highly recommended resources). For example, currently LibraryThing.com provides features such as Zeitgeist "the spirit of the times" and presents analyses of most favourite books by way of average ratings, tag clouds and most reviewed books.

Linked Data as a data model and set of technological frameworks does not stand by itself and serves little or no purpose without the underlying data. Hence, incorporating both the legacy library metadata that has been created by librarians and socially-constructed metadata generated by users is crucial, both being equally important. Incorporating the latter is becoming especially relevant, in view of the need for identifying usage patterns and Zeitgeist. Weinberger (2005) asserts that "an author is an authority when it comes to what she intended her work to be about, but not about what it means to others. When it comes to searching, what a work means to the searcher is far more important than the author's intentions". The modelling and technological edges offered by Linked Data would substantially contribute to the success of activities aimed at identifying resource usage patterns, the Zeitgeist and emergent vocabularies employed by users, which could help libraries make strategically-correct decision when acquiring resources (purchase of books, journal subscriptions), which again would all the more justify the return on investments.

Facet-based navigation

Faceted browsing refers to the use of different dimensions and properties when navigating an information system, with the aim of improving discoverability (Morville, 2005). According to Morville (2005), the idea of using facets in libraries was popularised by S.R. Ranganathan. The use of facets contrasts with the use of taxonomies where an item goes to only a single location in the hierarchy, limiting alternative ways of navigation. Concurring with Ted Nelson's concept of intertwingularity, Weinberger (2007) argues against the notion of pursuing a single, pre-determined and perfect navigation structure. He writes: "people keep pretending they can make things deeply hierarchical, categorisable, and sequential when they can't. Everything is deeply intertwingled". The term intertwingularity, coined by Ted Nelson (who is also credited for the term hypertext), refers to the interrelatedness and interdependency of ideas (concepts), thus suggesting that any attempt to categorisation is artificial (for more on Ted Nelson's intertwingularity see: Weinberger, 2007; Wright, 2007). Digital information objects, Weinberger contends, could be sorted, categorised and presented in a multitude of ways based on users' needs and context (Weinberger, 2007). To quote Weinberger (2007) at length:

> "*In the third order of order [in full-text digital libraries], a leaf [an information object] can hang on many branches, it can hang on different branches for different people, and it can change branches for the same person if she decides to look at the subject differently. It's not that our knowledge of the world is taking some shape other than a tree or becoming some impossible-to-envision four-dimensional tree. In the third order of order, knowledge doesn't have a shape. There are just too many useful, powerful, and beautiful ways to make sense of our world*."

In his book, "Too Big to Know" (2012), Weinberger suggests that "the solution to the information overload problem is to create more metadata" whilst indicating the importance of filtering relevant results based on various facets. In his book, "The Laws of Simplicity", Maeda (2006), states that "simplicity is about subtracting the obvious, and adding the meaningful." He writes "more appears like less by simply moving it far, far away. Thus an experience is made simpler by keeping the result local, and moving the actual work to a far away location."

In principle, digital libraries are free from the constraints imposed by the physical shelf in conventional libraries, hence enabling users to filter information in a multitude of ways using keywords and categories as facets in the discovery of information objects. Through the power of the link structure of Linked Data and the metadata richness that would be generated through the use of both standards-based and socially-constructed metadata approaches, a multitude of facets would be used to filter the content presented to users. Such metadata richness may as well utilise user profiles so that digital library systems would be able to contextualise and customise search results. Currently, Last.fm, LibraryThing, and Amazon.com are some of the examples where faceted navigation and metadata filtering are utilised.

Metadata enriched with links

At present, the links to metadata on an OPAC display can only go up to a certain point and soon reach a dead end when a data item is not linked further. For example, current metadata interfaces allow a user to search for a specific book, click on the author's name and see his publications, and, maybe find related/similar books. They, usually, do not offer a seamless link to the author's biography page or from there to another page, such as one on Wikipedia, Google books, back to the library listing, etc. A richly described book in a Linked Data environment could be linked to anything that is related to, it would be up to the user to make a decision as to whether to stop the navigation, once one's information need is satisfied, or

for the applications to provide contextual and relevant information. In other words, each piece of metadata is enriched with a link unless in fact such metadata is of type literal such as age of a person where it is not necessary to link it to another page.

Metadata enriched with contextual and relevant links would allow users to seamlessly navigate between disparate library databases and external information providers such as other libraries, and search engines. By globally and uniquely identifying entities (such as works, people, places, and events), metadata elements or properties (author, title, subject, relations) and corresponding values (instances), Linked Data offers a multitude of ways to enrich information objects with metadata which would facilitate information discovery and improve the experience of the user in digital libraries.

Conclusions

Contemporary library standards and models carried forward some of the constraints inherited from the traditional card catalogue system. The metadata generated, using these models, is mainly attuned to human consumption rather than machine processing. Linked Data is considered to be the solution to enable a data-centric and machine process-able metadata. However, the adoption of Linked Data in the actual creation and utilisation of library metadata is yet in its infancy.

It is suggested that existing library models, such as RDA and FRBR, are re-conceptualised in view of making them compatible with Linked Data principles; and libraries should convert existing MARC-compatible metadata records to RDF/XML serialisation formats, thereby slowly moving out of MARC and its derivatives. In regard to this, it is argued that, the call to abandon existing metadata models (such as RDA and FRBR) is against the ethos of the Linked Data where it allows disparate standards and controlled vocabularies to “happily” co-exist in the same sphere, provided that these resources (objects) are referenced using globally unique URI and that a high level data model known as RDF is employed. In connection with this, centuries-old best practises and lessons learnt by libraries whilst developing, maintaining and preserving controlled vocabularies (such as authority names and subject headings) along with the wealth of experience determining users needs, should be leveraged in every initiative at adopting Linked Data principles. Proper adoption of Linked Data for library metadata, it is suggested, would help the library develop novel library applications and services such as faceted browsing, serendipitous browsing, metadata zeitgeist and identification of emergent trends and user vocabularies.

References

Alemu, G., Stevens, B., & Ross, P. (2012). Towards a Conceptual Framework for User-Driven Semantic Metadata Interoperability in Digital Libraries: A Social Constructivist Approach. *New Library World, 113*(1/2), 38-54.

Alexander, B. (2006). Web 2.0: A New Wave of Innovation for Teaching and Learning? *EDUCAUSE Review, 42*(2), 32-44.

Allemnag, D., & Hendler, J. (2008). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Amsterdam: Morgan Kaufmann.

Anderson, C. (2006). *The Long Tail: How Endless Choice is Creating Unlimitted Demand*. London: Random House Business Books.

Anderson, C. (2010). *Free : How Today's Smartest Businesses Profit by Giving Something for Nothing*. London: Random House Business Books.

Barbosa, D. (2008). Taxonomy Folksonomy Cookbook. Retrieved from http://solutions.dowjones.com/cookbook/ebook_sla2008/cookbookebook.pdf

Berners-Lee, T. (1997). Metadata Architecture. Retrieved from http://www.w3.org/DesignIssues/Metadata.html

Berners-Lee, T. (1998). What the Semantic Web can Represent. Retrieved from http://www.w3.org/DesignIssues/RDFnot.html

Berners-Lee, T. (2007). Testimony of Sir Timothy Berners-Lee Before the United States House of Representatives Committee on Energy and Commerce Subcommittee on Telecommunications and the Internet. Retrieved from http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html

Berners-Lee, T. (2009). Linked Data. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T. (2010). Tim Berners-Lee: The Year Open Data Went Worldwide. from http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web: A New form of Web Content that is Meaningful to Computers will Unleash a Revolution of New Possibilities. *The Scientific American, 284*(5), 34-43

Brand, S. (1994). *How Buildings Learn: What Happens after They're Built*. London: Viking.

Brand, S. (1999). *The Clock of the Long Now: Time and Responsibility* (1st ed.). New York: Basic Books.

Campa, R. (2008). Making Science by Serendipity. A review of Robert K. Merton and Elinor Barber's the Travels and Adventures of Serendipity. *Journal of Evolution and Technology 17*(1), 75-83.

Campbell, D. G., & Fast, K. V. (2006). *From Pace Layering to Resilience Theory: The Complex Implications of Tagging for Information Architecture.* Paper presented at the The 7th Information Architecture Summit, Vancouver, British Columbia, Canada.

Cobden, M., Black, J., Gibbins, N., Carr, L., & Shadbolt, N. (2011). *A Research Agenda for Linked Closed Data*. Paper presented at the Second International Workshop on Consuming Linked Data. Retrieved from http://eprints.ecs.soton.ac.uk/22711/3/position.pdf

Coyle, K. (2010). Library Data in a Modern Context. *Library Technology Reports 46*(1), 5-13.

Coyle, K., & Hillmann, D. (2007). Resource Description and Access (RDA): Cataloging Rules for the 20th Century. *D-Lib Magazine, 13*(1/2).

Day, M. (2000). Resource Discovery, Interoperability and Digital Preservation: Some Aspects of Current Metadata Research and Development. *VINE, 36*(117), 35-48.

Day, M. (2003a). *Integrating Metadata Schema Registries with Digital Preservation Systems to Support Interoperability: A Proposal.* Paper presented at the International Conference on Dublin Core and Metadata Applications, Seatle.

Day, M. (2003b). *Preservation Metadata Initiatives: Practicality, Sustainability, and Interoperability*. Paper presented at the ERPANET Training Seminar on Metadata in Digital Preservation. Retrieved from http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf

Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., et al. (2000). The Semantic Web: The Roles of XML and RDF. *IEEE Internet Computing, 15*(3), 63-74.

Denton, W. (2007). FRBR and the History of Cataloging. In A. G. Taylor (Ed.), *Understanding FRBR: What it is and How it Will affect Our Retrieval* (pp. 35-57). Westport, Connecticut:

Libraries Unlimited.
Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., & Van de Sompel, H. (2012). *The Europeana Data Model (EDM).* Paper presented at the World Library and Information Congress: 76th IFLA General Conference and Assembly, Gothenburg, Sweden.
Dunsire, G. (2009). *UNIMARC, RDA and the Semantic Web* Paper presented at the World Library and Information Congress: 75th IFLA General Conference and Council Retrieved from http://www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf
Evans, W. (2009). *Building Library 3.0: Issues in Creating A Culture Of Participation*. Oxford: Chandos.
Foster, A., & Ford, N. (2003). Serendipity and Information Seeking: an Empirical Study. *Journal of Documentation, 59*(3), 321-340.
Gruber, T. (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies, 43*, 907-928.
Gruber, T. (2007). Ontology of Folsonomy: A Mash-up of Apples and Oranges. *International Journal on Semantic Web & Information Systems, 3*(2).
Gruber, T. (2008). Collective Knowledge Systems: Where the Social Web Meets the Semantic Web. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 6*(1), 4-13.
Guenther, R., & McCallum, S. (2003). New metadata standards for digital resources: MODS and METS. *Bulletin of the American Society for Information Science and Technology, 29*(2).
Haslhofer, B., & Isaac, A. (2011). *data.europeana.eu - The Europeana Linked Open Data Pilot.* Paper presented at the DC-2012 International Conference on Dublin Core and Metadata Applications, The Hague, The Netherlands.
Helen, W. (2010). Linked Data and Libraries. *Catalogue & Index*(160), 2-5.
Horridge, M., Knublauch, H., Rector, A., Stevens, R., & Wroe, C. (2004). A Practical Guide to Building OWL Ontologies Using the Prot´eg´e-OWL Plugin and CO-ODE Tools Edition 1.0 Available from http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf
IFLA. (2009). *Functional Requirements for Bibliographic Records: Final Report*: International Federation of Library Associations and Institutions.
Lagoze, C. (2010). *Lost Identity: the Assimilation of Digital Libraries into the Web.* Cornell University, School of Information Science.
Library of Congress. (2011). *A Bibliographic Framework for the Digital Age*. Washington, D.C.: Library of Congress.
Lubetzky, S. (1953). Development of Cataloging Rules *Library Trends 2*(2), 179-186.
Marcum, D. (2011). *A Bibliographic Framework for the Digital Age*. Washington, D.C.: Library of Congress.
Massis, B. E. (2011). "Serendipitous" browsing versus library space. *New Library World, 112*(3/4), 178-182.
Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata Retrieved from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html
Merton, R. K., & Barber, E. (2004). *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*. Princeton: Princeton University Press.
Miller, P. (2005). Web 2.0: Building the New Library *Ariadne, 45*.
Morville, P. (2005). *Ambient findability*. Sebastopol, CA: O'Reilly.
Nilsson, M. (2010). *From Interoperability to Harmonization in Metadata Standardization: Designing an Evolvable Framework for Metadata Harmonization.* Unpublished PhD, School of Computer Science and Communication, KTH, Stockholm.
Noy , N. F., & McGuinness, D. L. (2000). Ontology Development 101: A Guide to Creating Your First Ontology. Retrieved from http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html
O'Reilly, T. (2005). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Retrieved from http://oreilly.com/web2/archive/what-is-web-20.html
Rothenberg, J. (2008). Interoperability as a Semantic Cross-cutting Concern *Interoperabiliteit: Eerlijk zullen we alles delen*. Den Haag.
Shadbolt, N. (2010). *data.gov.uk - The Linked Data Revolution*. Paper presented at the Innovating Through Information Lecture Series. Retrieved from http://eprints.ecs.soton.ac.uk/18787/

Shirky, C. (2005). Ontology is Overrated: Categories, Links, and Tags. *Clay Shirky's Writings About the Internet*. Retrieved from http://www.shirky.com/writings/ontology_overrated.html
Smith, G. (2008). *Tagging : People-Powered Metadata for the Social Web*. Berkeley, CA: New Riders.
Stoskopf, M. K. (2005). Observation and cogitation: how serendipity provides the building blocks of scientific discovery. *Institute for Laboratory Animal Research, 46*(4), 332-337.
Styles, R. (2009). Bringing FRBR Down to Earth. http://dynamicorange.com/2009/11/11/bringing-frbr-down-to-earth/
Styles, R., Ayers, D., & Shabir, N. (2008). *Semantic MARC, MARC21 and the Semantic Web*. Paper presented at the Linked Data on the Web (LDOW2008). Retrieved from http://events.linkeddata.org/ldow2008/papers/02-styles-ayers-semantic-marc.pdf
Tapscott, D., & Williams, A. D. (2010). *Wikinomics: How Mass Collaboration Changes Everything* (Expanded ed.). London: Penguin Books.
Tennant, R. (2002). MARC must die. *Library Journal*. Retrieved from http://www.libraryjournal.com/article/CA250046.html
The British Library. (2011). Free Data Services. Retrieved December 14, 2011, from http://www.bl.uk/bibliographic/datafree.html
Udell, J. (2004). Collaborative Knowledge Gardening with Flickr and del.icio.us, Social Networking goes Beyond Sharing Contacts and Connections. *InfoWorld*. Retrieved from http://www.infoworld.com/d/developer-world/collaborative-knowledge-gardening-020
Veltman, K. H. (2001). Syntactic and Semantic Interoperability: New Approaches to Knowledge and the Semantic Web. *New Review of Information Networking, 7*(1), 159-183.
W3C. (2004a). OWL Web Ontology Language Overview. from http://www.w3.org/TR/owl-features/
W3C. (2004b). RDF Primer. *W3C Recommendation*. Retrieved from http://www.w3.org/TR/2004/REC-rdf-primer-20040210/#basicconcepts
W3C. (2011). *Library Linked Data Incubator Group Final Report*: W3C.
Wallis, R. (2011a). Library of Congress to Boldly Voyage to Linked Data Worlds. http://consulting.talis.com/2011/11/library-of-congress-to-boldly-voyage-to-linked-data-worlds/
Wallis, R. (2011b). Will Europe's National Libraries Open Data In An Open Way? Retrieved from http://dataliberate.com/2011/09/will-europes-national-libraries-open-data-in-an-open-way/
Weinberger, D. (2005). Tagging and Why it Matters. Retrieved from http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/07-WhyTaggingMatters.pdf
Weinberger, D. (2007). *Everything is Miscellaneous*. New York: Times books.
Weinberger, D. (2012). *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. New York: Basic Books
Wilson , N. (2010). *Linked Data Prototyping at the British Library*. Paper presented at the Talis Linked Data and Libraries event. Retrieved from http://blogs.talis.com/nodalities/neil-wilson-lightning-talk-%E2%80%93-linked-data-and-libraries-2010
Wright, A. (2007). *Glut: Mastering Information Through the Ages*. Ithaca: Corenell University Press.
Zuckerman, E. (2010). Desperately Seeking Serendipity. Retrieved from http://www.chi2011.org/program/plenaries.html

About the authors

Getaneh Alemu is a PhD student in the School of Creative Technologies, and Dr Brett Stevens (Principal Lecturer, School of Creative Technologies), Dr Penny Ross (Senior Lecturer, School of Computing) and Jane Chandler (Principal Lecturer, Faculty of Creative and Cultural Industries), all at the University of Portsmouth, Portsmouth, UK. Getaneh Alemu is the corresponding author and can be contacted at: Getaneh.Alemu@port.ac.uk