Metadata and Semantics Research: a case of an international conference paving toward a data driven future

Getaneh Alemu[1], Emmanouel Garoufallou[2], Rania Siatri[3], Damiana Koutsomiha[4] and Sirje Virkus[5]

Abstract

This paper provides an overview of the 11th international conference on Metadata and Semantics Research (MTSR-2017) which was held in Tallinn, Estonia from November 28th to December 1st, 2017. The paper contextualises this with existing literature and concludes by offering insight toward the future of metadata. MTSR-2017 brought metadata experts from various domains including libraries, museums, archives, higher education and agriculture. The conference provided an opportunity for participants to share their knowledge and novel approaches in the implementation of metadata and semantics technologies across diverse types of information environments and applications. In libraries, it is indicated that contemporary standards-based metadata approaches fail to describe the ever increasing size of digital resources as well as the changing technologies and dynamic user requirements. The paper shows the need to re-conceptualise existing metadata principles and technical formats with emerging Linked Open Data frameworks. This is where, the theory of metadata enriching and filtering (Alemu & Stevens, 2015) fits in. It emerges that the future of metadata, ontologies and semantics is enriched, linked, open and filtered. In addition, ontologies need to reflect the diversity of interpretations inherent in human beings and the existence of multi-lingual, cross-cultural and multi-disciplinary content – hence they should be designed, developed and maintained with diversity, scalability and interoperability in mind.

Introduction

A conference is a platform or "agora" which brings people, knowledge and crafts together so that their ideas emerge, receive criticism and can be shared. It is a venue for networking with professionals, and in such a network, the sum is greater than its parts. As David Weinberger contends, knowledge resides in the network of experts who actively engage in conversations and not on a single individual's brain (Weinberger, 2012). This is especially true of the field of metadata, ontologies and semantics where the overarching aim ought to be bridging semantics gaps, obviating duplication of effort, offering adequate description of resources and providing explicit specifications of concepts to enhance discovery of data, information and knowledge. It was in light of this that the 11th international conference on Metadata and Semantics Research (MTSR-2017) brought together metadata, ontology and semantics experts from across the world to Tallinn, Estonia – a vibrant, digital and snowy city. The history of MTSR conference goes back to 2005. Since 2009, its conference proceedings were published in Springer's Communications in Computer and Information Science (CCIS) book series. Selected revised and extended papers are also published on the International Journal of Metadata, Semantics and Ontologies (Inderscience). In what follows, we aim to offer a summary of the ideas presented at the conference, identify emerging trends and provide an insight into future directions.

To begin with, it is appropriate to note the fact that the term metadata is a recent phenomenon, especially with the development of the web and the creation of the Dublin Core metadata standard in 1995 in Dublin, Ohio. However,

[1] Solent University, Southampton, UK

[2] Department of Library Science and Information Systems, Alexander Technological Educational Institute of Thessaloniki, Greece

[3] Department of Library Science and Information Systems, Alexander Technological Educational Institute of Thessaloniki, Greece

[4] American Farm School, Thessaloniki, Greece

[5] School of Digital Technologies, Tallinn University, Tallinn, Estonia

the notion of cataloguing itself goes back to the history of libraries (Wright, 2007). Library pioneers including Anthony Panizzi, Charles Cutter, Melville Dewey, S. R. Ranganathan and Seymour Lubetzky (Svenonius, 2000; Wright, 2007) played a major part in devising mechanisms to help users find books in a library. For these pioneers, the library catalogue has one essential function – to help users identify a particular book by title, author or subject and to do so efficiently. Such individual contributions were further developed and standardised by international initiatives such as the Anglo-American Cataloguing Rules (AACR), Paris Principles, MAchine-Readable Cataloguing (MARC), Resource Description and Access (RDA) and Library of Congress Subject Headings (LCSH). These efforts led to the development of a plethora of metadata standards (Coyle & Hillmann, 2007).

In his book, "Metadata: Shaping Knowledge from Antiquity to the Semantic Web", Gartner (2016) contends about the centrality of metadata in organising, finding, managing and archiving information. Meta in Greek refers to after or beyond about things, events or people. The meaning of metadata goes well beyond its catchy but circular definition of "data about data" (Gartner, 2016). He discusses topics covering taxonomies, controlled vocabularies, faceted classification, ontologies and folksonomies. He chronicles a shift from strict hierarchal information organisation systems to networked and interlinked metadata systems. He cites examples such as FOAF and SKOS – which use RDF to represent and encode metadata. As Gartner explains, metadata has three fundamental components – semantics (e.g. Dublin Core metadata fields), syntax (e.g. MARC, XML and RDF) and content rules (e.g. RDA). He notes "human knowledge is built on 'about-ness' and it is through our interpretation of what the world is 'about' that most of our intellectual endeavours are based. Without metadata we cannot have knowledge". Gartner notes the challenges to harness user-generated metadata along with that of expert-created metadata. Gartner concurs that expert-created and user-generated metadata are complementary. He cites several examples of crowd-sourcing projects such Galaxy Zoo, also called a citizen science project, where hundreds of thousands of volunteers from around the world contributed in the identification and classification of planets and astronomical objects from a deluge of satellite images; and Ancient Lives, where online volunteers help transcribe ancient Greek papyri. As Gartner argues, "metadata never stands still" and it has helped shape "knowledge from antiquity" to the present day of the Semantic Web as "it represents our desire to make sense of the world". As Gruber (2007; 1993) indicates, metadata is a shared conceptualisation or understanding. It is not set in stone – meaning and semantics change across time and across cultures. This is where metadata sharing and interoperability come to play so as to bridge the syntactic and semantic gap.

Metadata is ubiquitous - it is all around us. Your name, age, title, height, weight, hobbies and where you live constitute personal metadata to describe you. Likewise, the attributes of a book such as title, author, format, genre, location and description are book metadata. The US National Information Standards Organisation (NISO, 2004) defines metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource". By providing descriptive (author, title and subject), administrative (identity, provenance, rights, contextual and technical) and structural information (relations with other information objects), metadata plays an important role in digital libraries to support the findability and discoverability of information objects by users and also librarians. According to Berners-Lee (1997), metadata is a systematic, consistent, named, address-able (using web URL, URI) and stored data to make assertions about things. Metadata can be understood by humans and/or computers.

As collections grow, the need for structured, scalable, interoperable and open metadata standards rise (Alemu, 2014). For example, the Library of Congress, the largest library in the world, owns more than 164 million information objects (print and digital), followed by the British library, which provides access to more than 150 million items. A more recent digital undertaking is Europeana.eu, which is a gateway to more than 51 million artworks, artifacts, books, videos and sounds from across Europe. Other initiatives include the Digital Public Library of America (which provides access to more than 20 million items from libraries, archives, and museums), Project Gutenberg (over 56,000 free and public domain e-books), World Digital Library (more than 19,147 items about 193 countries between 8000 B.C and 2000 A.D) and the Internet Archive (more than 15 petabytes of webpages) – all facilitated, mediated and enabled by an ever changing set of computer and web technologies (Cameron & Kenderdine, 2007; Kalay, Kvan, & Affleck, 2008). Digital technology brought a significant shift in how cultural heritage information objects can be represented, mediated, disseminated, preserved and accessed (Cameron & Kenderdine, 2007; Hedstrom et al., 2003; Kalay et al., 2008; Tammaro, 2016). New media and technologies provide opportunities for preserving and expanding access to cultural heritage. It encompasses both born digital and digitised representations of physical information objects (Kalay et al., 2008). Metadata and ontologies prove to be central for systematic organisation and discoverability. Metadata is also central for the preservation and access of cultural heritage objects. Information discovery including search, relevancy ranking, faceted refinement and grouping related resources all rely on metadata (Varnum, 2016). As Zeng and Qin (2008, p. 3) note, metadata is "the invisible hand" that enables effective information organisation. Information organisation is powered by metadata. As Weinberger opines, "metadata liberates knowledge". In recognition to this, significant investments have been made to design and develop metadata standards by a number of national, multinational and international initiatives in order to describe collections and to enhance the findability and discoverability of information objects (Alemneh, 2009; Chan & Zeng, 2006; Gartner, 2008; Gartner, 2016; Lagoze, 2000; Lagoze, 2010; Zeng & Qin, 2016).

As the size, format, genre and diversity of data and information expands with the development of new technologies (Blair, 2010; Weinberger, 2012), the need for scalable metadata approaches arise. It is especially important to develop overarching principles which would help explain both past and present practices as well as help guide the future so that scalable and interoperable systems could be developed (see Berners-Lee, 1997; 1998). In MTSR 2017, the importance of URIs, RDF, OWL, Linked Data were underlined, hence the need to move from document-based, human-readable metadata to machine-readable, structured, uniquely identified and interlinked metadata (metadata linking), from metadata silos to metadata openness, metadata sharing and re-use, and as regards to discovery, from a single interface to user-led, re-configurable interface (metadata filtering) (Alemu, 2014).

There is now a new focus on data. In connection with this, Christine Borgman, a renowned author and commentator on scholarly communication, open data, digital libraries and information infrastructure, argued in her book "Big data, little data, no data: scholarship in a networked world", (MIT press, 2015), that data is the new "gold" for organisations to stay competitive and relevant, hence its management, presentation, utilisation and preservation becomes critical. She argues that Big Data is a rather broad conceptualisation of an emerging field in Data Science which encompasses developments in data creation, storage, citation, re-purposing and using. The recent emphasis on data goes beyond publications – having significant implications for universities and research institutions to re-think their strategies and practices to data. She accentuates the need for metadata (technical, provenance and descriptive) so as to make sense and provide context for research data. Metadata provides provenance and description relevant for its discovery, re-use and curation (Borgman, 2015). Borgman (2015, p.31) states that "data has taken on a life of their own, or so it would seem from the popular press, independent of the scholarly context in which they are used as evidence of some phenomena". Data infrastructures such as standards, ontologies and metadata as well as international coordination for data curation, data sharing and data-reusing, Borgman (2015) emphasises, are all important concerns that warrant serious scientific and research consideration. MTSR papers focus on how data and metadata are defined, represented, structured and shared using emerging technologies such Linked Data, ontologies, open data, big data, RDF, OWL, SKOS, schema.org and BIBFRAME.

The concepts of metadata, ontology and semantics are not always clear cut. Metadata is a broader rubric which encompasses schemas (elements), taxonomies, thesauri and ontologies. According to Gartner (2016) "the crucial difference is the way in which an ontology brings its components together. Gone are the rigid rungs of the enumerative or even facetted ladder: now concepts can be related to each other like the strands of a spider's web, a network of ideas and linkages that can be as flexible as the subject being modelled requires". Ontologies are by definition extensible, flexible and scalable. They allow various interpretations (meanings) to co-exist. Semantics is the meaning behind the various interacting components of an ontology.

Peeter Normak, Director of the School of Digital Technologies of Tallinn University, noted the timeliness and relevance of topics such as metadata, open data, social data, linked data, ontologies and semantics for the emerging field of Data Science. He foresees a future where university research is not only data-led or data intensive but that data itself becomes the focus of research, hence the call for Data Science. Whilst taxonomies and thesauri tended to be hierarchical and siloed, ontologies flourish in a deeply networked and open environment (Schreur & Lorimer, 2017). Similarly, contemporary library standards and models carried forward some of the constraints inherited from the traditional card catalogue system. The metadata generated using conventional metadata standards is mainly attuned to human consumption rather than machine processing. To this end, Linked Data is considered to be a viable solution to enable data-centric and machine process-able metadata. Dr Aalberg keynoted that "libraries are on the move to re-usable data with rich and well-defined semantics" and he foresees them using FRBR bibliographic data coded in RDF format and guided by RDA vocabularies. Structuring, enriching, linking, crowdsourcing (annotating) and filtering appear to be the keywords de Boer at al. (2017) adopted in their award winning paper entitled "Enriching Media Collections for Event-Based Exploration".

In connection with this, the principle of metadata enriching becomes an emerging theme of MTSR 2017. This coheres with the theory of metadata enriching and filtering (Alemu, 2014). No institution could claim exclusivity of metadata ownership. In other words, a single institution, of whatever size and resources, cannot optimally enrich metadata on its own. The principle of metadata linking is a powerful means for creating seamless connections between disparate sources of data, which supports enriching by making library metadata share-able and re-usable. For this to happen, institutions need to use open and scalable metadata formats. In this regard, the principles which underpin Linked Data are considered important (Alemu, 2014). According to Allemnag and

Hendler (2008, pp. 7-13), Linked Data offers a distributed data model whereby “anyone can say anything about any topic” resulting in ‘variations and even 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” (Allemnag & Hendler, 2008, p. 11). Linked Data principles enable and operate in an open, dynamic and interactive system – allowing diversity and enabling richness. However, traditionally, library standards have mostly operated in a closed and static environment detached from the general web information landscape. The new and emerging principles cater towards metadata structure, machine process-ability, unique identification, granularity, interoperability, discoverability, metadata openness and metadata sharing (Tapscott & Williams, 2010). These are also some of the overarching topics presented at MTSR 2017.

Methodology

A total of 31 papers, one workshop and one keynote speech were presented at MTSR 2017, of which most are covered in this summary (see articles: Akbar, Fensel, & Fensel, 2017; Bunakov & Matthews, 2017; Can, Unalir, Sezer, Bursa, & Erdogdu, 2017; Cancellieri, Pontika, Pearce, Anastasiou, & Knoth, 2017; see Chen, 2017; Daif et al., 2017; de Boer et al., 2017; Dutta, Toulet, Emonet, & Jonquet, 2017; Eito-Brun, 2017; Fiorelli, Pazienza, Stellato, & Turbati, 2017; García-Barriocanal, Sánchez-Alonso, & Sicilia, 2017; Georgiadis et al., 2017; Grabus & Greenberg, 2017; Guerrini, ; Hedayati & Laanpere, 2017; Inan & Dikenelli, 2017; Jin, Hahn, & Croll, ; Jug & Žumer, 2017; Karimova, Castro, da Silva, Pereira, & Ribeiro, 2017; Lara-Clares, Garcia-Serrano, & Rodrigo, 2017; Opalek & Greenberg, 2017; Özacar et al., 2017; Parinov, 2017; Samuel, 2017; Sartori, 2017; Schreur & Lorimer, 2017; Simas, Barros, Salvador, Weber, & Amorim, 2017; Siragusa, Di Caro, & Tosalli, 2017; Sugimoto, 2017; Thiéblin, Amarger, Hernandez, Roussey, & Trojahn, 2017), thematically grouped into three topical conceptual categories, namely:

- Ontology and metadata enriching,
- Ontology and metadata interoperability and alignment, and
- Ontology and metadata discoverability and filtering.

Whilst the papers presented at MTSR-2017 are discussed under the above broad categories, it is not easy to delineate clear thematic differentiation between them. For instance, many papers cover more than one category, i.e. a paper discussing ontology creation discusses ontology alignment and discoverability as well. Put another way, these categories are non-exclusive.

Ontology and metadata enriching

Metadata enriching encompasses the design of schemas and ontologies as well as the creation, enhancement, storage and management of metadata values. Enriching is a continuous and non-deterministic process (Alemu & Stevens, 2015). Whilst at present standards-based metadata approaches, such as taxonomies and controlled vocabularies, attempt to predetermine (anticipate) users’ terminologies, it is nonetheless an untenable task to fully and richly describe the ever growing digital collections with metadata that caters to individual needs (Alemu, Stevens, Ross, & Chandler, 2012; Shirky, 2005; Weinberger, 2007). Following recommendations made by various authors (Alemu & Stevens, 2015; Gartner, 2016; Howe, 2009; Surowiecki, 2004; Tammaro, 2016; Weinberger, 2007; Weinberger, 2012), it is noted that once an ontology is designed, the metadata creation and enhancement can happen post-hoc by users and it can be continually enhanced through collective metadata intelligence. Metadata and ontology development can thus be looked at as a continuous rather than a deterministic process. Both the ontology and resultant metadata can be continually improved. The principle of enriching takes the view that metadata can change through time, new RDF nodes can be added, similar or even contending values can co-exist within the same ontology (Allemnag & Hendler, 2008). For example, OWL allows equivalency, inverse and cardinality relations and constraints to be added in an ontology. This has greater implications for metadata sharing and reuse. The Linked Data model offers the ability to uniquely identify resources and metadata using URIs, alleviating duplication of data as well avoiding naming and identification conflicts (Berners-Lee, 1997; Berners-Lee, 1998; Berners-Lee, Hendler, & Lassila, 2001).

Unlike traditional taxonomies which are relatively rigid, ontologies allow an open-world assumption, whereby various assertions could co-exist. In this regard, the design of ontologies and metadata follows an interpretivist rather than an objectivist ontological view point. Benefiting from a network effect, where each contribution slowly aggregates, enriching aims to collect as much metadata as possible, thus facilitating metadata diversity. All interpretations are valid, thus any user can add anything at any time, provided that such metadata assertions are not malicious. As new users join the network, it is possible that the metadata becomes further enriched. However, with new conversations taking place regarding a given information object, the ontology and the metadata have the possibility to continuously evolve (change). Enriching is characterised by metadata diversity where it implies the inclusion of a multitude of potentially conflicting metadata ascribed to information objects by users. Metadata diversity increases the likelihood of conforming to the multitude of perspectives and interpretations of various groups of potential users. In relation to idiosyncratic (personal) metadata entries, it is important to recognise that a given metadata entry that might be considered trivial for a general user might be important for the one who created it, since it is likely that the latter will search with those keywords. Such idiosyncrasies can be managed by providing personalised presentation, which can be managed through appropriate metadata filtering. The new metadata paradigm may thus permit the inclusion of metadata descriptions (interpretations) of information objects that may seem in opposition. In such instances, it is important to maintain the diversity of various interpretations. Nevertheless, such a metadata paradigm should not include random entries; instead, it should cater towards semantic and meaningful metadata whilst at the same time maintain the diversity of interpretations (Alemu & Stevens, 2015).

The current metadata landscape

As part of MTSR-2017, new and emerging metadata principles, namely metadata enriching, linking, openness and filtering were presented by Getaneh Alemu. The aim of this three-hour workshop was to provide an overview of contemporary metadata principles such as the principle of sufficiency and necessity, the principle of user convenience, the principle of representation and the principle of standardisation (Svenonius, 2000). A brief history of library cataloguing was presented followed by a quick overview of standards such as MARC, Dublin Core, AACR-2, RDA, FRBR, Linked Data and BIBFRAME (see also Alemu & Stevens, 2015; Gartner, 2016; Wright, 2007; Wright, 2014).

Metadata: does it matter?

Without metadata, the library's print and digital collections remain invisible to users (Zeng & Qin, 2016). Not only is metadata essential for library resources discoverability but it also supports circulation, acquisition and interlibrary loan functions. This explains why libraries continue to invest in cataloguing and metadata for staff, discovery tools, LMS software and bibliographic services. According to Zeng and Qin (2016) metadata is "the invisible hand" that serves users to find information. For these purposes, metadata experts use metadata standards, systems and tools. They also ensure information resources are properly displayed on discovery services. They measure their success on how much metadata enhances discovery and usage of information resources. Some of the metadata indicators include:

- Completeness;
- Up-to-date-ness;
- Quality, and
- Usefulness.

The metadata creation process happens through various stages of the information resource life cycle. In theory, metadata creation and enhancement (metadata enriching) is a continuous process and it involves authors, publishers, suppliers, librarians and users. When a library ignores its metadata and discovery function, it is at the same time ignoring its purpose for existence, which is information organisation and provision.

Metadata is a language with an aim to communicate the about-ness, location, context, meaning and provenance of an information object such as a book, a journal article, data or information. For instance, the statement "Animal Farm is written by George Orwell" is metadata. A Linked Data expression includes triples, namely Subject (Animal Farm) - Predicate (written by) - Object (George Orwell). The MARC format takes the following form:

100 a Orwell, George, d 1903-1950, e author.
245 a Animal farm / c George Orwell.

Of course, one could also use the language of FRBR: the original idea of Animal Farm being the WORK; the 1946 publication by Brace & Company with omission of its sub-title – the EXPRESSION; the paperback edition – the MANIFESTATION and the copy catalogued by Library of Congress with a call number PZ3.O793 An2 PR6029.R8 – the ITEM.

By applying FRBRisation rules, one can bring related editions and versions together. Now, consider the whole art and language of classification. Without it, you have a chaotic library which is not a library at all. Furthermore, the whole idea of serendipitous discovery, faceted navigation and browsing of related works depend on good quality metadata which is a building block for discovery which in turn is crucial for information organisation. But information organisation would be haphazard if it is not built on foundational principles.

Metadata is therefore an essential component of what librarians do. Metadata describes and signposts to the actual information resource – the pointer could be a name (label, keyword, statement). The statement "Herman Melville's Moby Dick is a fascinating read" is metadata. It makes assertions about the book. Likewise, likes, recommendations, tags and reviews are also metadata. Metadata helps to make sense of large sets of data, commonly referred to as Big Data. A given image (photograph) can have the following metadata - name, item type, date created, date modified, owner, location, size, caption, dimension, width, height, resolution, bit depth, compression, colour representation, camera maker, camera model, ISO speed, focal length, max aperture, metering mode, flash mode, lens maker, lens model, flash maker, flash model, contrast, brightness, light source, saturation, sharpness, digital zoom, subject, rating, tags, comments, programme name and copyright.

Metadata helps to answer the 'what, by whom, why, when and where' questions. In addition, tags, annotations, the table of contents or the index of a book are all metadata. In simple terms metadata is the combination of keywords you use to find a specific information resource such as a book, a journal article or just a piece of data (information). The richer the information resource is described with relevant, accurate, complete and user-centred metadata, the more efficient and effective your search could be. Keywords are though just one type of metadata. Reviews, annotations, comments, browsing/clicks habits, ratings, tags, number of downloads/popularity, sources, hardware/software used to create the information resource, provenance metadata such as who, where, for what purpose created/updated/shared it. Metadata is also about connections (links): metadata helps you create connections between digital resources. It brings together various versions and formats of an information object and related resources. Hence, good metadata is one enriched with links. This is because metadata that is enriched with links would give endless possibilities to explore and discover information objects. Potentially, every metadata value can be linked as words and phrases in an online dictionary can be linked, thus users can select any word and retrieve the meaning of it. Likewise, users can select any metadata link retrieving information objects associated to that metadata value.

A mixed metadata approach

Both librarian created and user-created (socially-constructed) metadata can be combined and mixed to maximise the findability (discoverability) and usage of information resources. In such a mixed metadata paradigm, users are considered as pro-active contributors. Hence, authors (content creators), librarians, publishers and most importantly, users can co-create, contribute, enhance, enrich and manage metadata. User-generated metadata improves the search-ability, browse-ability, find-ability and management of information resources. As Weinberger (2005) notes "when it comes to searching, what a work means to the searcher is far more important than the author's intentions", thus arguing for a more user-focused metadata.

It was highlighted that metadata is ubiquitous – it is the raison d'être for libraries as it constitutes one of their core functions, i.e., ensuring print and electronic resources are easily discovered and retrieved by users– hence justifying the return on investment. The same holds true for other cultural heritage institutions such as museums, galleries and archives. Whether these institutions use one or more of existing standards such as Dublin Core (web), VRA Core (museum and visual), EAD (archives), Marc21 (libraries), MODS (simplified MARC), IPTC Core (photo), CDWA (Categories for Description of Works of Art), METS (structure), PREMIS (preservation), ISO2709 (MARC), XML, HTML5, microdata, RDFa, RDF/XML, JSON, RDA, AACR2, CCO, DACS, OAIS, Linked Data, FRBR, LCSH (Library of Congress Subject Headings), AAT (Art & Architecture Thesaurus), TGN (Getty Thesaurus of Geographic Names), DDC, LCNAF or ISO 639-2 (ISO code for languages), the chief business of these institutions is to describe the great diversity of information resources (print or electronic books, articles, reports, datasets, cultural heritage objects or artifacts) with metadata that reflects the heterogeneity inherent in users (Zeng & Qin, 2016).

In the workshop, it was indicated that metadata has come a long way from paper listings to card catalogues to the OPAC and now discovery systems (Alemu & Stevens, 2015; Zeng & Qin, 2016). However, as Paul Otlet argued "the catalogue guides the reader as far as the location of the book but not to the contents within and also relationships between documents" – where the social space of the book (user metadata, annotations and reviews) is missing. The breadth and size of information resources grows by the day, "with an estimated 4 billion pages on the Web, which would take 57,000 years to read" (Boulton, 2014), making navigating and filtering all the more challenging. However, as David Weinberger argues "metadata liberates knowledge". Metadata describes, annotates, provides context and answers the "what, by whom, why, for whom and when" questions, thus helping the user to find, identify, make sense, contextualise, filter through and use information resources such as print and e-books, journal articles and data. Whilst there exists a plethora of standards in use in libraries and cultural heritage institutions, they are not however without challenges. Some of these challenges arise due to the size and diversity of collections, ever changing technologies and changing users' expectations.

Library standards coalescing around FRBR, RDA and BIBFRAME

In the workshop, participants were given a trainer's user name and password to access the RDA Toolkit. It is important to note that, with relationship designators, 336, 337, 338 (also called GMD in AACR2), the rules of three being omitted, abbreviations being spelled out and cataloguers being given the prerogative (flexibility) to apply their expertise/judgment, RDA results in a relatively complete description of resources that could enhance findability and discoverability (see also Anhalt & Stewart, 2012; Brenndorfer, 2016; El-Sherbini, 2012; Maxwell, 2013; Tillett, ; Welsh & Batley, 2012). RDA's flexibility does not necessarily imply anything that goes to a record is of good and acceptable quality. Flexibility does not mean inconsistent/less quality metadata. AACR2 was borne in a time when space on the card catalogue was a major issue, hence bibliographic description was simple, brief and abbreviated (such as ed., rev., vol., s.l., s.n., n.d. and et al), whereas now most of these limitations can be avoided with the use of computer storage and displays, and therefore RDA aims to reduce these restrictions as it places emphasis on the expertise of cataloguers not on stringent rules.

In principle, RDA should enable cataloguers to create metadata in such a way that similar works by the same author or on a related subject can easily be brought together to the user. In comparison, AACR2 was very much focused on the item being catalogued rather than on similar other works. RDA better complies with current web technologies (in line with FRBR, WEMI, FRAD theoretical framework and the use of relations and relationship designators). RDA also enables the creation of metadata (a catalogue record) that better caters to finding, discovering, identifying, selecting and obtaining information resources. In addition, RDA is intuitive for cataloguers and helps to generate user-friendly bibliographic metadata; for example, it avoids the use of abbreviations. However, due to current limitations of Library Management systems (LMS), it is not always easy to demonstrate the full benefit of RDA especially in relation to supporting metadata enriching, optimal discoverability and metadata linking. RDA is a theoretical model for the creation of richer bibliographic description and it is generally platform independent.

Furthermore, in the MTSR 2017 workshop, Alemu presented about the similarities and differences between FRBR and BIBFRAME. FRBR is a conceptual model whereas BIBFRAME is a data model following Linked Data principles. RDA on the other hand is a set of guidelines and specifications to record bibliographic data. BIBFRAME is therefore one instance of a data model that aims to replace MARC. FRBR, in addition to being a conceptual data model, helps guide the display as well. It helps to group related editions and versions together. FRBR display requires the use of authority headings and consistent entry of the title field.

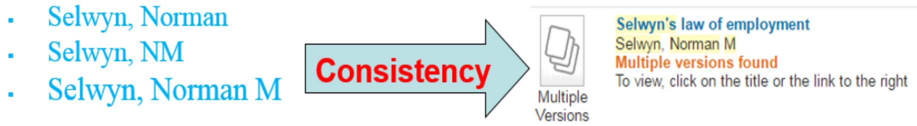

Figure 1: FRBR display

In the workshop, various national and institutional initiatives and projects were cited to indicate that libraries are actively working towards new models and approaches: notable ones include:

- *http://www.loc.gov/bibframe/docs/index.html*
- *http://bnb.data.bl.uk/*
- *http://www.bnf.fr/en/professionals/bnf_data_sets.html*
- *http://www.bne.es/en/Inicio/Perfiles/Bibliotecarios/DatosEnlazados*
- *http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata.html*
- *https://www.nls.uk/about-us/open-data*

In comparison to FRBR, BIBFRAME uses the following classes (entities): Work, Instance, Item, Agents, Subject, Event and Annotation.

Source: http://bibframe.org/tools/editor/#

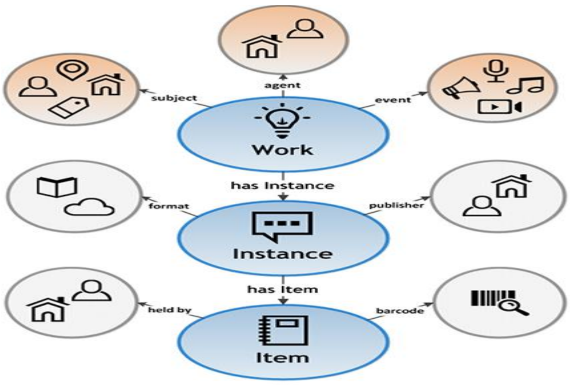

Figure 2: BIBFRAME model (http://swib.org/swib14/slides/miller_swib14_57.pdf)

BIBFRAME is still a work in progress. However, the logic behind the data model follows the FRBR conceptual data model. It also uses RDA as a guide for data entry. In addition to theoretical concepts and principles, the workshop also covered two practical demonstrations on the RDA Toolkit and BIBFRAME editor.

Metadata enriching can happen at various stages of the metadata life cycle: modelling, ontology/schema design, metadata creation, utilisation, enhancement, management and preservation. Enriching can happen when technologies (formats) change. With the introduction of Linked Data principles and the emergence of RDF, legacy library bibliographic metadata need to be transformed to new formats and specifications. Chen (2017) notes that in its current state, the MARC bibliographic record format locks libraries from the Linked Data world as it was singularly designed to work in a library environment. This echoes arguments raised elsewhere by metadata experts (Alemu, Stevens, Ross, & Chandler, 2014; Alemu & Stevens, 2015; Coyle, 2015; Dunsire, Harper, Hillmann, & Phipps, 2012; Godby, Wang, & Mixter, 2015; Kroeger, 2013; Lampert & Southwick, 2013; Pesch & Miller, 2016). It thus requires reconceptualization and reformatting in accordance with Linked Data requirements. Chen (2017) suggests metadata conversion stages, namely data identification, data modelling, naming with URIs, reusing existing terms, publishing human and machine-readable descriptions, RDF conversion, license, download, host and announcement. The role of consistent URIs is underlined (Chen, 2017). Chen highlighted that the granularity of metadata becomes an issue for a 1-1 mapping hence “a contextual mapping” is required which involves re-modelling of MARC data into a Linked data model. This shows that there is a need for manual intervention of metadata and ontology experts. Citing Europeana LOD and British Library’s BNB Linked data initiatives as exemplary ongoing projects, Chen (2017) is optimistic that legacy library metadata can be better harnessed by re-modelling and converting it to a Linked Data format.

MTSR’s keynote speaker Dr. Trond Aalberg, Associate Professor at the Norwegian University of Science and Technology, noted that libraries are embracing linked, re-usable data with rich and well-defined semantics. He claimed that libraries have a long-standing tradition of developing cataloguing standards, bibliographic record formats and tools. Libraries have a long tradition of creating and maintaining bibliographic information to support discovery and retrieval of library resources and documenting intellectual (what?) work? He reiterated that amidst ever dynamic user expectations and technological changes, libraries, though they were early adaptors of standards, now face a growing challenge to use new web-based, linked and open data models. He described various initiatives such as FRBR, FRBRoo, CIDOC CRM, FRAD, FRASAD, BIBFRAME and SCHEMA.ORG. With large volumes of data (e.g. some national libraries own more than 30 million bibliographic records) arise data quality issues such as accuracy, completeness and conciseness (non-redundancy). By demonstrating an experimental tool called BIBSURF: Discover Bibliographic Entities, developed at NTNU, he highlighted the need for more research to try to re-use legacy bibliographic data.

Eito-Brun (2017) notes the existence of a plethora of standards in the cultural heritage domain. Some of the standards in use by archival expects include Encoding Archival Description (EAD), Metadata Object Description Standard (MODS), PREMIS, MARCXML, VRA-Core, TextMD, MIX, AudioMD and ALTO. He demonstrated a model which integrates various standards, mainly the archival description standard EAD and METS that offers an integrated approach to digitise and describe archival records. It is indicated that there is no single standard that meets the metadata requirements of archivists, hence the need to mix and match various solutions from the various standards (Eito-Brun, 2017). Similarly, Schreur and Lorimer (2017) describe why libraries should place new attention on Linked Data, and why they move away from what is considered as an outdated MARC format. The authors argue that MARC was not designed to be interoperable for data that resides outside the library environment. Schreur and Lorimer are involved in a six-university collaborative project called "Linked Data for Production (LD4P)" which is aimed at developing practical workflows to convert MARC data into Linked Data formats such as BIBFRAME. The authors acknowledged an apparent duplication of effort by various institutions who are currently trying to devise strategies and techniques to convert MARC data to Linked Data formats. The key challenge, as Schreur & Lorimer indicated, is to move from "library-centric to a web-centric" data formats – for which BIBFRAME could serve as a high-level data model (Schreur & Lorimer, 2017).

In the absence of a semantically structured, enriched and linked ontology, researchers in the digital humanities had to manually triangulate information from heterogeneous media collections to inform their research (de Boer et al., 2017). To address these challenges, a Linked Data model is designed with entities such as sem:Event, sem:Place, sem:Actor, sem:Time, oa:Annotation, dive:MediaObject and skos:Concept. As a result, a knowledge graph is created linking media objects, people and places with historical events which in turn lends itself for navigation and serendipitous browsing. Collections are enriched with structured metadata which also creates semantic links across the various digital objects. Firstly, a semantic data model is designed. Secondly, through a crowdsourcing service, users (scholars) were allowed to create links between entities and events. "Based on extensive requirements analysis done with historians and media scholars, we present a methodology to publish, represent, enrich, and link heritage collections so that they can be explored by domain expert users" (de Boer et al., 2017).

Based on Eugene Garfield's basic assumption that "citing and cited references have a strong link through semantics", Parinov (2017) conducted a content-based citation analysis where the full-text content of both cited and citing papers is analysed. As Parinov (2007, p.298) describes, content-based citation analysis offers benefits such as "citation semantic attributes summarization, information retrieval based on the citation context, citation recommendation and prediction services enhanced knowledge graph and conceptual networks". As demonstrated from the papers presented at MTSR-2017, metadata is pervasive and affects every aspect of professional activity and personal life (Gartner, 2016). As Jug and Žumer (2017) argue, metadata is key for discoverability not just in libraries but also in online book stores. Hence, if book publishers and dealers are serious about selling books, they need to reconsider how they utilise metadata (Jug & Žumer, 2017). The breadth, diversity and quality of metadata are equally important. In addition, the authors call for interoperability and cooperation among publishers and libraries. As Gruber defines "an ontology is [an explicit] specification of a conceptualisation" (Gruber, 2007; Gruber, 1993; Gruber, 2008). This highlights that all domains or fields including curriculum design can be modelled using an ontology. In MTSR 2017, Hedayati & Laanpere (2017) demonstrated the importance of ontologies in "representing, organising, formalising and standardising the knowledge in ICT domain so that ontology-driven curriculum modelling (ODC) can be built, shared and reused in the curriculum development, implementation and evaluation process (Hedayati & Laanpere, 2017). Ontologies can help foster a shared understanding of a domain and help obviate inconsistencies and duplication of effort.

According to Gartner (2016), an ontology has three components: semantics - the meanings of the fields or elements; syntax - the way in which the metadata is encoded; and content rules - the rules that form and content of the metadata. Whilst humans may be adept at tacit knowledge, computers are not. Hence, the need for an explicit specification of a domain. Glossaries are often used to determine the definition of a term, its context and relations with other terms. As Scheidegger, Campos and Cavalcanti (2017) argue "the success of computer systems interoperability and data integration greatly depends on precisely capturing the conceptualizations and human interpretations of the reality being represented". Through a process called ontological analysis relevant similarities and distinctions of terminologies (entities) can be identified. This helps to bridge the gap between business concepts and its representations which in turn obviates inconsistencies, ambiguities and implementation issues. The process helps to bring precision and consistency in definitions and representations (Scheidegger et al., 2017). The authors used the Unified Foundational Ontology (UFO) as a framework to identify the relationship of a certain concept in terms of generalization, specialization and dependence with other concepts. If you type Apple in Wikipedia, it offers you at least three options – Apple Inc., an apple tree or Apple Records. The process is called disambiguation. Inan and Dikenelli (2017) indicate that enriched ontology structures support disambiguation. The authors reported that "if the ontology has an instance and class rich structure", it results in better disambiguation and better entity linking. Ontology design does not always fit into formal logical structures which can be expressed explicitly. One such problematic area is representing temporal changes, also called tenses relations which are not binary, for example settlement (geographical) changes, changes in marriage status where ternary relation of the form R(x, y, t), where one of its arguments, say the third one, refers to times (i.e., moments or periods) (Garbacz & Trypuz, 2017).

Interoperability - ontology and metadata schema alignment

There exists a plethora of metadata standards/schemas and ontologies across domains, indicating the increasing need to make semantic, structural and syntactic alignments across ontologies. Such alignments (agreements) could be implemented through cross-mapping, federation, harvesting (Arms et al., 2002; Blanchi & Petrone, 2001; EC, 2010; Haslhofer & Klas, 2010; Hunter, 2003; Lim & Liew, 2010; Nagamori & Sugimoto, 2006; Nilsson, 2010; Office, 2004; Rothenberg, 2008). According to Dutta, Toulet, Emonet and Jonquet (2017) there is no single ontology (metadata schema) that meets the requirements of all information professionals. To address such challenges, metadata standards are analysed and a metadata vocabulary for Ontology Description and publication, referred as MOD 1.2, is developed. MOD 1.2 is derived from a total of 23 existing metadata vocabularies (e.g., Dublin Core, OMV, DCAT, VoID). A metadata usage analytics within ontologies and ontology repositories is conducted. As a result, a total of 88 properties were identified.

Dutta et al.(2017) acknowledges the existence of mostly disparate initiatives across the world to build ontologies and metadata schemas. Given that ontology construction is time consuming and expensive, sharing and re-use is suggested. A knowledge base is designed using Protégé where users can filter through various metadata properties from the ontologies included in the knowledge base – hence mixing, matching and reconfiguring a new ontology (Dutta et al., 2017). Cultural heritage collections are by their very nature heterogeneous and they cross boundaries. A case in point presented by Özacar et al. (2017) is a museum in Ephesus. The authors argue that the cultural heritage domain benefits by integrating ontologies and metadata from diverse and disparate sources. The paper presents a framework for intelligent integration of ontologies from heterogeneous data. A web application is developed which aggregates content from Ephesus (Turkey) and British Museum (London). The web application provides two main functionalities: (a) querying the Ephesus Museum knowledge base (b) finding the most relevant objects in the British Museum with the selected object from the Ephesus Museum (Özacar et al, 2017). Lara-Clares et al (2017) discussed a semantic data model that crawls the web for data and information based on a given profile and creates a new narrative by using semantic graphs – something Ted Nelson envisioned, i.e. to automatically re-create new documents out of existing data (Wright, 2007).

Analogous to working simultaneously on a single document by more than one author using track changes in Microsoft Word, as Fiorelli, Pazienza, Stellato, & Turbati (2017) demonstrate, the same can be applied to an ontology so that various versions of the same ontology can be asynchronously modified and later merged together. Such workflows can be developed, changes can be tracked back to a specific state, and triples can be added to nodes or merged following rules set to the ontology, hence enabling a collaborative and iterative development process for ontologies. Since one of the key defining characteristics of an ontology that distinguishes it from a taxonomy or thesauri is its scalability to include various statements (triples) - using such constructs as owl:sameAs, equivalentProperty, owl:differentFrom, owl:AllDifferent, owl:equivalentClass or owl:complementOf – version control is an important area of investigation under ontology alignment/merging in a collaborative environment. The paper notes the importance of version control and change validation for RDF datasets. After reviewing existing RDF version control systems, the authors suggested that two data sets (d1, d2) can be compared and computed automatically to ascertain the addition of new nodes, which is computed as $\Delta add = d2 \setminus d1$. As the authors argued "by shifting the focus of the management from versions to changes, we observed that it is possible to implement this synchronous workflow in a clearer manner" (Fiorelli et al., 2017). One of the key features of ontologies is its scalability and openness. Digital libraries are likely to operate in a heterogeneous and distributed environment, thus the possibility of working on different versions of the same ontology and the ability to automatically compare and merge or identify a particular version of a dataset is appealing (Fiorelli et al., 2017).

Metadata and ontology discoverability and filtering

Once an ontology is designed and is enriched with metadata, the layer of discoverability is designed for the user. Georgiadis et al.(2017) demonstrated a specialised discovery interface called Searchculture.gr where digital cultural heritage artifacts (objects) are enriched with metadata attributes such as Creator, Institution, Type, Historical period, Chronology, Europeana type, Rights, Language, Repository/Collection and Subject. The Historical period and Chronology chosen covers from early Bronze Age (3000 B.C.) to modern Greece (1900 A.D.). It uses an ontology-design method called "chronological normalization" which is used to enhance discoverability and navigation using time/period facets and visual navigation. One can filter search results using a particular period, for example filter results by a historical timeline from 526 B.C. to 5 B.C. (Georgiadis et al., 2017).

Linked data and big data developments are often looked at through the lenses of the principle of openness. However, as Grabus and Greenberg (2017) argue, data could be sensitive, confidential and closed, hence requiring a long and protracted process of data sharing agreements. Some of the issues include anonymising data, safeguarding sensitive information and health and financial data. The authors contend that "data sharing of sensitive and private information introduces a set of challenges, and requires the development of most robust agreements, […] to expedite the process of developing data sharing agreements". The research looks at the possibility of developing a framework to help the automatic generation of data sharing agreements in "restrictive environments" (Grabus & Greenberg, 2017, p.309). The development and adoption of Linked Data and ontologies does not make metadata standards obsolete. Opalek and Greenberg (2017) noted the role of metadata standardisation in regulating and credentialing health professionals. Such metadata includes demographic data, work experience, qualification, criminal history, licence information and malpractice (criminal) history. In the United States, nearly a fourth of medical professionals earned their medical degrees abroad, thus metadata is found crucial to authenticate, regulate and license these professionals – hence the rationale for a national project involving several universities.

Thiéblin et al (2017) present a novel but challenging approach of querying distributed and heterogeneous Linked Data sources through automatic re-writing of queries, based on entity and value relations (such as owl:sameAs, owl:equivalentClass, or owl:equivalentProperty) which automatically reformulate a query and apply to these heterogeneous datasets, providing a single point of entry to the user. Such systems hide the complexity from users who do not need to go through the painstaking process of searching (querying) across all ontologies and data sources. In such an inter-linked and open environment, complex query rewriting and alignment on data sources happens backstage without presenting any complexity to the user (Thiéblin et al., 2017). Similarly, Daif et al. (2017) presented a Linked Data ontology based on CIDOC Conceptual Reference Model (CRM) which allows users to navigate through semantic associations between selected entities such as heritage items, characters, events or locations. Edmond and Nugent Folan (2017) reiterate the notion that the metadata of an object depends "on the eye of the beholder" – noting the importance of data narratives – often referred to as documentation, provenance or metadata. Such narratives, the authors indicated, are important for "serendipitous discovery of connections between otherwise dissimilar documents". The same data could be used by various people. "We must be more rigorous about documenting and sharing information about the transformations applied to data, so that we can access not just the data as it is now, but retrace its journey to its current state via a sort of "data passport." Third, we must work toward a state where cultural heritage institutional finding aids are able to converge with the secondary literature that discusses the collections represented there. The historic separation between scientific publishers and cultural heritage institutions is a huge barrier to an obvious opportunity to enhance the big data of the catalogues with the rich data of scholarly production" (Edmond & Nugent Folan, 2017).

MTSR-2017 also saw new and emerging topics on the role and application of metadata for block chain, nanotechnology and host intrusion (Bunakov & Matthews, 2017; Can et al., 2017; García-Barriocanal et al., 2017). Block chain technology contains chains of blocks (nodes) which are located in a distributed environment where at each block transaction (provenance) metadata is stored. Block chain is a new, novel and distributed approach to managing databases and transactions. According to García-Barriocanal et al. (2017) block chain technology helps to obviate existing challenges to digital archives and helps to offer "decentralized identification, deferencing, proof of statement and (separately) indexing". Simas, Barros, Salvador, Weber and Amorim (2017) presented a Linked Data solution to collect and disseminate emergency and disaster data to various destinations.

Among the papers presented was a paper on "Creative Knowledge Environments Promotion through Case-Based Knowledge Artifacts" (Sartori, 2017) which aims at developing a framework to retrieve and reuse previous cases (decisions) and revise solutions when varying cases are presented – solving thus new problems based on past solutions. The framework aims to support creative decision-making processes not only by retrieving similar cases from memory but also by learning from previous solutions hence new ones based on the problem at hand. This framework fits with the visions of the Semantic Web (Berners-Lee et al., 2001). Bunakov and Matthews (2017) argue that the creation and application of metadata is ubiquitous – affecting most organisations. New research areas such as nanotechnology generate huge amounts of metadata which needs methodical design and harnessing. They noted that "nanotechnology research and innovation deal with substantial amounts of data of all sorts. It may be data about nano-materials (either physical or computer-simulated) or it may be contextual data that is important for the research or industrial management, for complying with health and safety regulations, as well as for keeping records about the provenance of materials and products. Managing these various data requires good metadata which presents challenges both from a metadata design perspective and from an operational perspective of metadata quality and its semantic interoperability" (Bunakov & Matthews, 2017, p.247).

As computer security threats such as viruses and malwares increase by the day, a novel approach to detect network security breaches is all the more prudent (Can et al., 2017). Can et al. (2017) developed an Intrusion Detection System (IDS) Ontology "that examines the processes and services of a device and the devices connected to the network, and also the packets on the network to which these devices are connected". The ontology defines classes, object and data type properties. "In order to detect intrusions by using IDS ontology, first we parsed malwares from Symantec's website to a csv file. Then, by using Jena, these malwares are written as individuals to IDS ontology's Malware class. Then using Facebook's osquery, individual Process and Service classes which belong to the working computer system were created. The implementation compares the individuals of Malware class with the individuals of Process and Service classes to detect intrusions. If it finds a match between individuals in these classes of the IDS ontology, then an intrusion is detected and a warning message is shown" (Can et al., 2017, p.82).

Scheidegger et al. (2017) noted that data or information without its semantic context could be "too amorphous, too ambiguous, too subjective, to slippery and elusive", making it difficult to make sense, hence pointing out the need to develop ontologies that help to discover similarities and distinctions between terminologies used in various domains. This ensures semantic richness without losing precision. According to Karimova, Castro, da Silva, Pereira and Ribeiro (2017), research data management is an emergent field that needs systematic metadata and semantic annotation to ensure data is findable, accessible, interoperable and reusable. In this regard, Cancellieri, Pontika, Pearce, Anastasiou and Knoth (2017) identified limitations of existing metadata harvesting systems and suggested a scalable model. The new model, called CORE, adopts a systematic metadata ingestion system with components such as metadata download (extracting and cleaning), full-text download (correction), information extraction (citation extraction), enrichments and indexing (duplicate detection, related content identification). The CORE architecture is scalable, easy to maintain and recoverable - currently enabling seamless access to about 82 million open access research papers (Cancellieri et al., 2017). In addition, Sugimoto (2017) underlined the need to comply with the FAIR data principles i.e., Findable, Accessible, Interoperable, and Re-usable. To liberate data for re-use, sharing without losing its context and provenance requires standardization of APIs, data structure and user-friendly GUI tool to use APIs more efficiently so that researchers could use data without having high level technical expertise of these systems (Sugimoto, 2017).

Future directions

In books such as "Too much to know" (Blair, 2011), "Too big to know" (Weinberger, 2012) and "Little data, big data, no data" (Borgman, 2015), "Shaping knowledge from antiquity to the Semantic Web" (Gartner, 2016) and "An emergent theory of digital library metadata" (Alemu & Stevens, 2015), the role of metadata in particular and information management in general as well as the need to underpin systems design and development (the how) with foundational theoretical principles and theories (the why) is emphasised (Svenonius, 2000). Metadata, ontologies and semantics are essential components of any scalable, open, inclusive, functional and future proof information system. Conceptual and physical data models provide the framework to enrich, link, interoperate, cross-walk, merge, compare, revise and ameliorate it. Open world assumptions and open formats are also important. This is where technical specifications and formats such as HTTP, URI, RDF, RDF-S, OWL, JSON-DL and SPARQL could be used to design and develop open, linked and scalable ontologies and metadata schemas. In libraries, the focus is slowly gearing towards RDA, FRBR, RDF, Linked Open Data, BIBFRAME and Schema.org. As Alemu (2014) noted "Linked Data is a viable opportunity for libraries to link their metadata with external data providers, which in turn would make library metadata connected to external information sources. However, libraries are not easily going to abandon their principles, which they have developed over hundreds of years. Linked Data principles however do not require libraries to abandon their existing principles and legacy databases".

At MTSR 2017, the concepts of metadata and ontology enriching, linking and interoperability (ontology/schema mappings and alignments) were discussed. For instance, De Boer et al. (2017) discussed metadata enrichment through a hybrid of techniques including Named Entity Recognition (NER) tools, Natural Language Processing (NLP), crowdsourcing and nichesourcing. The general themes of De Boer et al. (2017) could be summed as enrichment, linking and serendipitous browsing. MTSR proves to be an engaging and vibrant conference focusing on metadata, semantics and ontologies. As Borgman (2015) argues a new set of knowledge infrastructure (tools, metadata, policy, strategies and expertise) is required for creating, storing, representing, identifying, structuring, linking, citing, curating, accessing, discovering, sharing, using and reusing data to solve problems, support research and scholarship. With regard to openness to research data, the FAIR data principles of findable, accessible, interoperable, and re-usable data are the way forward. However, when data is sensitive (such private, confidential, health data), the systems need to make informed exceptions so that privacy and confidentiality are maintained.

It is important that MTSR technical papers and demonstrations are undergirded with foundational and theoretical concepts. It is revealing to learn that libraries and archives begin to open their metadata for re-use and sharing by institutions outside their domains. It is important MTSR pays attention to the potential of user-generated metadata and find technical solutions to structure, make sense and harness it effectively. Overall, MTSR 2017 was a successful conference on emergent theories and applications of metadata and ontologies. Recurring key concepts and standards from the conference include: Linked data, open data, big data, RDF, OWL, SKOS, schema.org, BIBFRAME, ontology design, discovery, search and information retrieval. The future of metadata is enriched, linked, open and filtered.

References

Akbar, Z., Fensel, A., & Fensel, D. (2017). Enabling analysis of user engagements across multiple online communication channels. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 147-159). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_14

Allemnag, D., & Hendler, J. (2008). *Semantic Web for the working ontologist: effective modeling in RDFS and OWL.* Amsterdam: Morgan Kaufmann.

Alemneh, D. G. (2009). An examination of the adoption of preservation metadata in cultural heritage institutions: an exploratory study using diffusion of innovations theory (PhD), University of North Texas. Retrieved from: http://digital.library.unt.edu/ark:/67531/metadc9937/?q=gelaw

Alemu, G., & Stevens, B. (2015). An emergent theory of digital *library metadata: Enrich then filter*. Waltham, Massachusetts: Chandos Publishing.

Alemu, G. (2014). A theory of digital library metadata: the emergence of enriching and filtering. Portsmouth: University of Portsmouth PhD thesis. Retrieved from: https://researchportal.port.ac.uk/portal/files/5315012/GetanehAlemu_PhD_Thesis_A_Theory_of_Digital_Library_Metadata_Bound_Copy_FinalVersion.pdf

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). The social space of metadata: Perspectives of LIS academics and postgraduates on standards-based and socially constructed metadata approaches. *Journal of Library Metadata*, *12*(4), 311-344. doi:10.1080/19386389.2012.735523

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2014). Toward an emerging principle of linking socially-constructed metadata. *Journal of Library Metadata, 14*(2), 103-129. doi:10.1080/19386389.2014.914775

Allemnag, D., & Hendler, J. (2008). *Semantic web for the working ontologist: Effective modeling in RDFS and OWL*. Amsterdam: Morgan Kaufmann.

Anhalt, J., & Stewart, R. A. (2012). RDA simplified. *Cataloging & Classification Quarterly, 50*(1), 33-42. doi:10.1080/01639374.2011.615378

Arms, W., Hillmann, D., Lagoze, C., Krafft, D., Marisa, R., Saylor, J., Van, d. S. (2002). A spectrum of interoperability: The site for science prototype for the NSDL. *D-Lib Magazine, 8*(1).

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

Berners-Lee, T. (1998). What the semantic web can represent. Retrieved from: https://www.w3.org/DesignIssues/RDFnot.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.

Blair, A. (2010). *Too much to know: managing scholarly information before the modern age*. New Haven, Conn.: Yale University Press.

Blanchi, C., & Petrone, J. (2001). Distributed interoperable metadata registry. *D-Lib Magazine, 7*(12)

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, Massachusetts: The MIT Press.

Brenndorfer, T. (2016). *RDA essentials*. London: Facet Publishing.

Bunakov, V., & Matthews, B. (2017). Metadata for nanotechnology: Interoperability aspects. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017,* *Proceedings* (pp. 247-252). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_24

Cameron, F., & Kenderdine, S. (2007). Theorizing digital cultural heritage: A critical discourse. Cambridge, Mass.

Can, O., Unalir, M. O., Sezer, E., Bursa, O., & Erdogdu, B. (2017). An ontology based approach for host intrusion detection systems. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 80-86). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_8

Cancellieri, M., Pontika, N., Pearce, S., Anastasiou, L., & Knoth, P. (2017). Building scalable digital library ingestion pipelines using microservices. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 275-285). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_27

Chan, L. M., & Zeng, M. L. (2006). Metadata interoperability and standardization – A study of methodology part I: achieving interoperability at the schema level. D-Lib Magazine, 12(6).

Chen, Y. (2017). (2017). A review of practices for transforming library legacy records into linked open data. Paper presented at the *Metadata and Semantic Research*, 123-133.

Coyle, K. (2015). Linked data: An evolution. *JLIS.It*, *4*(1), 53. doi://dx.doi.org/10.4403/jlis.it-5443

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

Daif, A., Dahroug, A., López-Nores, M., Gil-Solla, A., Ramos-Cabrer, M., Pazos-Arias, J., & Blanco-Fernández, Y. (2017). Developing quiz games linked to networks of semantic connections among cultural venues. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 239-246). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_23 de Boer, V., Melgar, L., Inel, O., Ortiz, C. M., Aroyo, L., & Oomen, J. (2017). Enriching media collections for event-based exploration. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 189-201). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_18

Dunsire, G., Harper, C., Hillmann, D., & Phipps, J. (2012). Linked data vocabulary management: Infrastructure support, data integration, and interoperability. *Information Standards Quarterly, 24*(2-3), 4. Retrieved from https://search.proquest.com/docview/1835567160?accountid=13969

Dutta, B., Toulet, A., Emonet, V., & Jonquet, C. (2017). New generation metadata vocabulary for ontology description and publication. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 173-185). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_17

EC. (2010). *European interoperability framework for pan-european e-government services*. Luxembourg: Office for Official Publications of the European Communities.

Edmond, J., & Nugent Folan, G. (2017). (2017). Data, metadata, narrative. barriers to the reuse of cultural sources. Paper presented at the *Metadata and Semantic Research*, 253-260.

Eito-Brun, R. (2017). The combined use of EAD and METS for archival material. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 230-236). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_22

El-Sherbini, M. (2012). *RDA: Strategies for implementation*. London: Facet.

Fiorelli, M., Pazienza, M. T., Stellato, A., & Turbati, A. (2017). Version control and change validation for RDF datasets. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia,* *November 28 – December 1, 2017, Proceedings* (pp. 3-14). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_1

Garbacz, P., & Trypuz, R. (2017). Representation of tensed relations in OWL. Paper presented at the *Metadata and Semantic Research,* 62-73.

García-Barriocanal, E., Sánchez-Alonso, S., & Sicilia, M. (2017). Deploying metadata on blockchain technologies. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 38-49). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_4

Gartner, R. (2008). Metadata for digital libraries: State of the art and future directions. Bristol: JISC Technology & Standards Watch. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.183.3762&rep=rep1&type=pdf

Gartner, R. (2016). *Metadata: Shaping knowledge from antiquity to the semantic web*. Cham, Switzerland: Springer.

Georgiadis, H., Papanoti, A., Paschou, M., Roubani, A., Hardouveli, D., & Sachini, E. (2017). The semantic enrichment strategy for types, chronologies and historical periods in searchculture.gr. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 211-223). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_20

Godby, C. J., Wang, S., & Mixter, J. K. (2015). *Library linked data in the cloud: OCLC's experiments with new models of resource description*. San Rafael, California: Morgan & Claypool Publishers. Retrieved from http://ieeexplore.ieee.org/xpl/ebooks/bookPdfWithBanner.jsp?fileName=7111834.pdf&bkn=7111833&pdfType=book; http://eresources.solent.ac.uk/V/?func=find-db-info&doc_num=000002992

Grabus, S., & Greenberg, J. (2017). Toward a metadata framework for sharing sensitive and closed data: An analysis of data sharing agreement attributes. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 300-311). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_29

Gruber, T. (2007). Ontology of folsonomy: A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems, 3*(2).

Gruber, T. (1993). Toward principles for the design of ontologies used for knowledge sharing. *International Journal Human-Computer Studies, 43*, 907-928.

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.

Guerrini, M.Introduction to the seminar: Global interoperability and linked data in libraries. *JLIS.it: Italian Journal of Library and Information Science, 4*(1), 21-24. doi://dx.doi.org/10.4403/jlis.it-6307

Haslhofer, B., & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. *Acm, 42*(2)

Hedayati, M. H., & Laanpere, M. (2017). Validating the ontology-driven reference model for the vocational ICT curriculum development. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference*, *MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 261-272). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_26

Hedstrom, M., Ross, S., Ashley, K., Christensen-Dalsgaard, B., Duff, W., Gladney, H., Neuhold, E. (2003). Invest to save: Report and recommendations of the NSF-DELOS working group on digital archiving and preservation.

Howe, J. (2009). *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: Three Rivers Press.

Hunter, J. (2003). Enhancing the semantic interoperability of multimedia through a core ontology. *EEE Transactions on Circuits and Systems for Video Technology, 13*(1), 49-58.

Inan, E., & Dikenelli, O. (2017). Effect of enriched ontology structures on RDF embedding-based entity linking. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 15-24). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_2

Jin, Q., Hahn, J., & Croll, G. (2016). BIBFRAME transformation for enhanced discovery. *Library Resources & Technical Services, 60*(4), 223-235.

Jug, T., & Žumer, M. (2017). Lost identity – metadata presence in online bookstores. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 160-166). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_15

Kalay, Y. E., Kvan, T., & Affleck, J. (2008). New heritage: New media and cultural heritage. London: Routledge. Retrieved from http://lib.myilibrary.com?id=106295&entityid=https://idp1.solent.ac.uk/idp/shibboleth; http://portal.solent.ac.uk/library/help/eresources/ebooks-help.aspx

Karimova, Y., Castro, J. A., da Silva, J. R., Pereira, N., & Ribeiro, C. (2017). Promoting semantic annotation of research data by their creators: A use case with B2NOTE at the end of the RDM workflow. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 112-122). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_11

Kroeger, A. (2013). The road to BIBFRAME: The evolution of the idea of bibliographic transition into a post-MARC future. *Cataloging & Classification Quarterly, 51*(8), 873-890. doi:10.1080/01639374.2013.823584

Lagoze, C. (2000) Accommodating simplicity and complexity in metadata: lessons from the Dublin Core experience. Retrieved from https://pdfs.semanticscholar.org/b539/161d3e09db6d18fb8120a06858495226e6ee.pdf

Lagoze, C. (2010). Lost identity: The assimilation of digital libraries into the web Available from Lost Identity: the Assimilation of Digital Libraries into the Web.

Lampert, C. K., & Southwick, S. B. (2013). Leading to linking: Introducing linked data to academic library digital collections. *Journal of Library Metadata, 13*(2-3), 230-253. doi:10.1080/19386389.2013.826095

Lara-Clares, A., Garcia-Serrano, A., & Rodrigo, C. (2017). Enrichment of accessible LD and visualization for humanities: MPOC model and prototype. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 327-332). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_31

Lim, S., & Liew, C. L. (2010). (2010). GLAM metadata interoperability. Paper presented at the *The Role of Digital Libraries in a Time of Global Change,* 140-143.

Maxwell, R. L. (2013). *Maxwell's handbook for RDA*. London: Facet.

Nagamori, M., & Sugimoto, S. (2006). A metadata schema registry as a tool to enhance metadata interoperability. *TCDL Bulletin, 3*(1)

Nilsson, M. (2010). *From interoperability to harmonization in metadata standardization: Designing an evolvable framework for metadata harmonization* Available from From interoperability to harmonization in metadata standardization: designing an evolvable framework for metadata harmonization.

NISO. (2004). Understanding metadata. Retrieved from https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf

Office, C. (2004). *E-government interoperability framework.* ().

Opalek, A., & Greenberg, J. (2017). The representation of agents as resources for the purpose of professional regulation and global health workforce planning. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 103-111). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_10

Özacar, T., Öztürk, Ö, Salloutah, L., Yüksel, F., Abdülbaki, B., & Bilici, E. (2017). A semantic web case study: Representing the ephesus museum collection using erlangen CRM ontology. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 202-210). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_19

Parinov, S. (2017). Semantic attributes for citation relationships: Creation and visualization. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 286-299). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_28

Pesch, O., & Miller, E. (2016). Using BIBFRAME and library linked data to solve real problems: An interview with eric miller of zepheira. *The Serials Librarian, 71*(1), 1-8. doi:10.1080/0361526X.2016.1183159

Rothenberg, J. (2008). Interoperability as a semantic cross-cutting concern. *Interoperabiliteit: Eerlijk zullen we alles delen*. Den Haag:

Samuel, J. (2017). Collaborative approach to developing a multilingual ontology: A case study of wikidata. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and*

Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings (pp. 167-172). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_16

Sartori, F. (2017). Creative knowledge environments promotion through case-based knowledge artifacts. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 50-61). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_5

Scheidegger, P. M. L., Campos, M. L. M., & Cavalcanti, M. C. (2017). An approach for systematic definitions construction based on ontological analysis. Paper presented at the *Metadata and Semantic Research*, 87-99.

Schreur, P. E., & Lorimer, N. (2017). Linked data in libraries' technical services workflows. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 224-229). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_21

Shirky, C. (2005). Ontology is overrated: Categories, links, and tags. *Clay Shirky's Writings about the Internet,*

Simas, F., Barros, R., Salvador, L., Weber, M., & Amorim, S. (2017). A data exchange tool based on ontology for Emergency response systems. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 74-79). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_7

Siragusa, G., Di Caro, L., & Tosalli, M. (2017). Automatic extraction of correction patterns from expert-revised corpora. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia,* *November 28 – December 1, 2017, Proceedings* (pp. 134-146). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_13

Sugimoto, G. (2017). Battle without FAIR and easy data in digital humanities. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 315-326). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_30

Surowiecki, J. (2004). *The wisdom of crowds : Why the many are smarter than the few*. London: Abacus.

Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, Mass.: MIT Press.

Tammaro, A. M. (2016). Heritage curation in the digital age: Professional challenges and opportunities. *International Information & Library Review, 48*(2), 122-128. doi:10.1080/10572317.2016.1176454

Tapscott, D., & Williams, A. D. (2010). *Wikinomics: how mass collaboration changes everything* (Expanded ed.). London: Penguin Books.

Thiéblin, E., Amarger, F., Hernandez, N., Roussey, C., & Trojahn, D. S. (2017). Cross-querying LOD datasets using complex alignments: An application to agronomic taxa. In E. Garoufallou, S. Virkus, R. Siatri & D. Koutsomiha (Eds.), *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28 – December 1, 2017, Proceedings* (pp. 25-37). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_3

Tillett, B.RDA and the semantic web, linked data environment. *JLIS.It, 4*(1), 139. doi://dx.doi.org/10.4403/jlis.it-6303

Varnum, K. (2016). Exploring discovery: The front door to a library's licensed and digitized content. London: Facet Publishing.

Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. New York, N.Y.: Henry Holt.

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.

Welsh, A., & Batley, S. (2012). *Practical cataloguing: AACR, RDA and MARC 21*. London: Facet.

Wright, A. (2007). *Glut: Mastering information through the ages*. Washington, D.C.: Joseph Henry Press.

Wright, A. (2014). *Cataloging the world: Paul otlet and the birth of the information age*. New York: Oxford University Press.

Zeng, M. L., & Qin, J. (2016). *Metadata* (2nd ed.). London: Facet Publishing.