A Constructivist Grounded Theory Approach to Semantic Metadata Interoperability in Digital Libraries: Preliminary Reflections

Getaneh Alemu, *M.Sc.*, **Brett Stevens**, *Ph.D*, **Penny Ross**, *Ph.D*
Getaneh.Alemu@port.ac.uk Brett.Stevens@port.ac.uk Penny.Ross@port.ac.uk
University of Portsmouth, Portsmouth, United Kingdom

Abstract
This paper highlights the problems of semantic metadata interoperability in digital libraries. The prevalence of the plethora of standards and the lack of semantic interoperability is partly attributed to the absence of theoretical foundations to underpin current metadata approaches and solutions. Present metadata standards and interoperability approaches are mainly top-down and hierarchical, which fail to take into account the diversity of cultural, linguistic and local perspectives. In this regard, it is suggested that, a social constructivist approach should be adopted as libraries and other cultural heritage institutions house information objects that need to be enriched with metadata, which reflects the diversity of views and perspectives of their users. Following on Charmaz (2006), a constructivist grounded theory method is adopted to investigate how library professionals and library users view metadata standards, collaborative metadata approaches and Semantic Web technologies in relation to semantic metadata interoperability. The method allows an active interplay between the researcher and the participants who are library and information science researchers, librarians and library users. From the first phase of data collection, preliminary reflections are presented on how library and information science professionals view current metadata practises especially as used in academic library contexts. However, as the study is ongoing, it is too early to generate theoretical categories and conclusions.

Keywords: metadata, digital libraries, semantic interoperability, constructivist grounded theory, social constructivism

1. Introduction
Metadata is an important component of any digital library and repository system (Chan & Zeng, 2006; Day, 2003a, 2003b; Duval, Hodgins, Sutton, & Weibel, 2002; Nilsson, 2010; NISO, 2004). While the term metadata is a recent concept, the notion of describing books and other information resources is as old as the existence of libraries (Day, 2005; Wright, 2007). Metadata is defined as data about data (CCSDS, 2002). However, Lavoie and Gartner (2005) and Day (2005) argue that this definition is less helpful and they suggest that metadata should be defined in relation to its functions. Such definition is given by the US National Information Standards Organization (NISO, 2004, p. 1) which defines metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource". Significant investment has been made to define metadata schemas by a number of national and international initiatives across the world (Anderson, Delve , Pinchbeck, & Alemu, 2009; Gartner, 2008; OCLC/RLG, 2001). These sit alongside local standards many of which are almost unknown by the wider community but which have been adopted for use by individual institutions. On the current trend, it is reasonable to expect that the situation will continue to become more complex over time. Each of these 'standards' requires implementers to follow some naming scheme, identification mechanism, controlled vocabulary, authority control, encoding scheme, format and technology. However, on closer examination, it is apparent that there is frequently some internal inconsistency in what the standards require of implementers. Common problems include: imprecise definition of terms, ambiguous characterisation of metadata elements, incomplete or otherwise incorrect identification protocols, conventions and encoding schemes (Haslhofer & Klas, 2010). In practice, these deficiencies give rise to serious difficulties for librarians and archivists. The diversity of metadata standards, the existence of local schemas and the heterogeneity in metadata use and implementation has significant implications for institutions to provide seamless and integrated access to information resources and share and exchange content and metadata records across heterogeneous digital libraries. With the growing trend towards establishing institutional, regional and international cooperation, such as the formation of the European Commission and the African Union, the quest for information sharing and exchange makes interoperability an important concern.

Interoperability is a broad term which encompasses the ability of separately developed systems to work together without the end user exerting significant effort. Today, interoperability has become a catch phrase in many regional bodies that need to collaborate. The interest for interoperability emanates for various reasons including information exchange, saving costs, benefiting tax payers from unnecessary bureaucracy, and facilitating business transactions (EC, 2010). In the context of digital libraries, interoperability refers to the ability to cross-search and integrate information resources from "multiple autonomous and heterogeneous information systems" (Haslhofer & Klas, 2010, p. 1). It also refers to the ability of bridging between information silos, re-using information and understanding the exchanged information (Rothenberg, 2008, Miller, 2000). However, according to Rothenberg (2008) one of the challenges to ensure interoperability is that when systems are designed, it is difficult to decisively determine what other systems will require from the system being designed.

Interoperability can be considered at various levels. Ouksel and Sheth (1999, p. 5) categorise it as system interoperability (compatibility between hardware and operating systems), syntactic interoperability (similarity in encoding and representation), structural interoperability (unified data-models, data structures and schemas) and semantic interoperability (consistent terminology and meanings). Taking it into a broader perspective, Miller (2000, pp. 2-4) classified interoperability into technical, semantic and added political/human (decisions to make resources widely available), inter-community (sharing interdisciplinary information across boundaries), legal (issues related to freedom of information, data protection regulations, and intellectual property rights) and international (languages) interoperability. Similarly, the European Commission in its Interoperability Framework action plan (EC, 2010) emphasises on the need for political will, mutual agreement between regional governments and stakeholders to streamline business functions and institutional activities. Interoperability is also a major national concern. For example, the UK's e-Government Interoperability Framework (e-GIF) focuses on technical aspects of interoperability such as interconnectivity, data integration, e-services and content management. The framework aims at setting and aligning to standards and specifications such as the XML and Dublin Core (UK Cabinet Office, 2005). The framework specifies that "the ultimate test for interoperability is the coherent exchange of information and services between systems" (Cabinet Office, 2005, p.29). The various types and levels of interoperability indicate that it is a cross-cutting concern (Rothenberg, 2008, p. 346) and achieving it depends on several overarching factors.

According to Miller (2000, p. 2) to be interoperable "one should actively be engaged in the ongoing process of ensuring that the systems, procedures and culture of an organisation are managed in such a way as to maximise opportunities for exchange and re-use of information, whether internally or externally". This, according to Miller, is more than using compatible hardware and software. Rothenberg also shares a similar view. He indicates that interoperability "implies far more than simply getting ICT systems to communicate with each other" as it also implies compatible interpretations, policies, and procedures to make sense of the exchanged information (Rothenberg, 2008, p. 347). In order for systems to be interoperable, therefore, organisations who design and maintain these systems should take into consideration not only technical aspects of the system but also the semantic, organisational, cultural and legal issues.

For digital libraries, achieving metadata interoperability is a big challenge (Chan & Zeng, 2006, p. 2). An ideal solution to metadata interoperability would be for all the digital libraries to implement and use a single standard (Chan & Zeng, 2006, p. 3). Even though such an approach has for so long been pursued by libraries with the adoption of the Dewey Decimal Classification, Anglo-American Cataloguing rules (AACR2), MAchine-Readable Cataloguing (MARC) and currently Dublin Core standards, such efforts have had problems. The prevalence of several metadata standards and "in-house" schemas has become evident. In such situations, achieving metadata interoperability using a single standard becomes difficult (Haslhofer & Klas, 2010, p.2). In situations where various metadata standards exist, some of the approaches of metadata interoperability include the use of metadata derivation, application profiles, metadata-cross walks (metadata matching), metadata registries and the use of Semantic Web technologies (Chan & Zeng, 2006; Day, 2003b; Nagamori & Sugimoto, 2006; NISO, 2004). However, not all these methods provide the required semantic interoperability for effective cross-searching, content sharing, and information integration. Hence, metadata interoperability remains a challenge.

Among the above approaches, metadata derivation involves developing a new schema from an existing one (Chan & Zeng, 2006). Examples are MARC-XML, MARC-Lite, and Metadata Object Description Schema (MODS) which are derived from the MARC standard. MARC is considered to be very cumbersome and complex hence schemas which are considered light to implement are derived (Chan & Zeng, 2006; Guenther & McCallum, 2003). For example Day argues that "MARC formats may not be the best 'fit' for the dynamic and fugitive resources that inhabit the Web environment" (Day, 2000). Guenther & McCallum (2003, p. 12) argue that the shift from the complex MARC format to a flexible and versatile XML encoding is a timely and important adaptation. The underlying problem in this approach is that since the problem of metadata interoperability is related to each metadata element, the way it is defined, labelled, represented, related (to other elements), content values (controlled vocabularies), and constraints, making the schema light does not enable semantic interoperability as there is a need to ensure those fields (in the light schema) and their corresponding values are properly understood by the end user or system.

A second approach to interoperability is the use of application profiles (Bake, Dekkers, Heery, Patel, & Salokhe, 2008; Chan & Zeng, 2006; Heery & Patel, 2000; Hillmann & Phipps, 2007). This is also called a 'mix-and-match' solution as it aims to bring several elements from different schemas (Dekkers, 2002; Duval, et al., 2002; Heery & Patel, 2000). The idea of developing and using application profiles looks a promising approach. However, the problem of metadata interoperability is rooted to the way each metadata element and associated

values are semantically defined and used. However current standards fail to address these underlying issues. Plus, as indicated by Haslhofer and Klas (2010) metadata has different levels of abstraction: meta-model, metadata schema and metadata instance. Application profiles enable sharing of best metadata practises and enable re-use of metadata elements. They help to avoid unnecessary duplication of effort. However, they are a schema level solution. So while exposing metadata schemas forms part of the solution towards interoperability, it does not indicate how the metadata records (content values) are exchanged and used. As Nilsson (2010, p. 136) argues "the problem with defining meta-data application profiles using XML schema is that each application profile defines precisely which schemas you are allowed to use. Therefore, for each new meta-data vocabulary you need to support, you will need to define a new application profile. This automatically puts a stop to the use of alternative meta-data descriptors, and results in an authoritarian limit on meta-data expressions". In this paper, it is argued that rigid and authoritarian specifications need to be addressed in order for such solutions to scale.

The third approach to interoperability is metadata cross walking (Chan & Zeng, 2006). A metadata cross walk "is a set of transformations applied to the content of elements in a source metadata standard that result in the storage of appropriately modified content in the analogous elements of a target metadata standard" (St. Pierre & LaPlant, 1998). For instance, a metadata cross-walk can be performed between Dublin Core and MARC and the common elements can help to merge various records of information objects defined using these different schemas. For instance, the element "245 $a" in MARC is equivalent to the "Title" element in Dublin Core. However, such equivalency mapping is very cumbersome and resource. Besides, by mapping a richer metadata schema to a simple schema such as MARC to Dublin Core, the fields that do not correspond are lost. Nilsson indicates that metadata cross-walks/mappings are only helpful as short-term solutions to interoperate between standards. The problems of cross-walks include the need for manual construction of the mapping, the problem of maintenance of the mapping schema, the disparity in terminology leading to incomplete mapping, lack of scalability as the number of standards grow, and problematic nature of mapping the semantics (Nilsson, 2010, p. 70).

A fourth approach to interoperability is the use of metadata registries (Bake, et al., 2008; Chan & Zeng, 2006; Day, 2003a; Dekkers, 2002; Heery, Johnson, Fülöp, & Micsik, 2003). Metadata registries make various metadata specifications explicit, which implementers may choose and apply in their own contexts, which may lead to the development of application profiles (Bailer, 2007). That being an important service, the problem still remains that these registries do not hold metadata content values. It is important to note that these metadata interoperability solutions do offer some level of interoperability at the schema level. However, as it stands now they do not deal interoperability at the semantic (content) level.

In contrast to the solutions suggested by Chan, Zeng, Nagamori and Sugimoto (Chan & Zeng, 2006; Nagamori & Sugimoto, 2006; Zeng & Chan, 2006), Nilsson indicates that current metadata interoperability techniques and methodologies such as metadata cross-walks, application profiles and metadata registries play either a marginal role or are severely limited. One of the problems identified is the limitation of XML to provide semantic mark up to metadata schema and content. Standards such as MODS, MARC-XML and METS use XML as their data encoding structure. However, as Decker et al and Nilsson (Decker et al., 2000; Nilsson, 2010) indicate XML is ineffective for semantic interoperability. This is because XML "aims at document structure and imposes no common interpretation of the data contained in the document" (Decker, et al., 2000, p. 6) hence does not embed semantics in its schema. Day, Nilsson, and Rothenberg (Day, 2000, 2003a, 2003b; Nilsson, 2010; Rothenberg, 2008) argue that semantic interoperability can be achieved through the use of Semantic Web technologies such as RDF, RDFS, and OWL. It is indicated that RDF's simple data model enables to create links between information resources. RDF schema adds vocabularies such as Class, SubClass, Domain, and Range to enable more meaningful representation of resources. By extending RDFS with yet additional vocabularies, the Web Ontology Language (OWL) allows to define additional semantic constructs such as equivalency, inverse and cardinality relations and constraints (Allemnag & Hendler, 2008; W3C, 2004). One of the features of the RDF model is the ability to uniquely and globally identify resources and metadata attributes (relations) using Uniform Resource Identifiers (URIs). The use of URIs for metadata element names, labels, and relations, according to Nilsson (2010), helps to avoid naming and identification conflicts in the use of elements. This is also suggested by Day and Rothenberg (Day, 2000, 2003a, 2003b; Rothenberg, 2008). Though there exist several academic papers and technical specifications on RDF, RSDFS, SPRQL, and OWL, up until now there are no viable semantic web related metadata solutions.

Semantic interoperability surpasses mere exchange of information and focuses on how the exchanged information can be meaningfully and semantically interpreted. This makes semantic interoperability an important issue for institutions. It involves, among other things, language, culture, values, and policies, and even politics. This also means the issues underlying semantic interoperability should be addressed at different levels: mainly at philosophical, theoretical, methodological as well as technological levels. This paper highlights the

need to define the philosophical perspective in defining standards and metadata interoperability solutions. For instance, practices in library standards such as MARC seem to imply an objectivist philosophical perspective, whereas in reality, libraries and the interpretation of their information objects (metadata) tends to be disparate, perhaps suggesting the need for an interpretive perspective. The design and deployment of Online Public Access Catalogues (OPAC) seem to favour an objectivist perspective, whilst the proliferation of Web 2.0 applications such as social tagging (collaborative metadata) seems to follow a social constructivist philosophical perspective. Thus, the philosophical assumptions that underline the decisions of metadata standards agencies greatly affect interoperability approaches and solutions.

Overall, review to existing works on metadata interoperability show that the authors start from addressing the 'how' instead of the 'why' of interoperability. While answering the 'how' questions is critical, to finding syntactic and structural interoperability, it says little about semantics. One major problem is that semantic metadata interoperability solutions lack theoretical underpinnings. It is however crucial that such theories are developed. The theories should be grounded in data and it is important that such data is obtained from practising librarians and metadata experts in the library and information science field as well as library users.

2. Philosophical and Theoretical Foundations of Semantic Metadata Interoperability Solutions

2.1. Philosophical Perspectives

The research will consider how a social constructivist approach can be adopted in order to achieve semantic metadata interoperability. As recommended by Guba and Lincoln (1994), Grix (2004), Creswell (2003) and Charmaz (2006), scholarly investigation should lay its foundation on the building blocks of research. According to Guba and Lincoln (1994, p. 109) the inquiry paradigms such as positivism, post-positivism, critical theory, and constructivism have three major questions to answer: what is to be known (ontology)? What is the relationship between the inquirer and the thing to be known (epistemology)? And how should the inquirer pursue his/her inquiry (methodology)? Each one of these philosophical assumptions should be addressed and their implications clearly understood by the investigator right at the outset of the research process. Furthermore, the assumptions should be guided by the nature of the research problem at hand, the investigator's experiences and the intended audience of the findings (Creswell, 2003, pp. 21-23). Such philosophical perspectives as to whether the investigator has adopted a positivist or interpretive paradigm should also be explicitly stated at the same stage.

In accordance with the above, it is felt that a thorough examination, and in-depth understanding, and a clear statement of the underlying ontological and epistemological perspectives will help re-evaluate the existing metadata standards and metadata interoperability solutions. For the purposes of this paper, an interpretive ontological perspective and a social constructivist epistemological approach are deemed appropriate. The paper's main contention is that current metadata practises are mainly top-down, hierarchical and stem from a foundationalism (objectivist) ontological viewpoint. Such a position as this, ontologically speaking, can only advocate a single solution to problems. It is worth noting that, though not explicitly stated in their policies, metadata agencies such as MARC and Dublin Core can be considered as examples of such a top-down approach.

2.2. Adopting a Social Constructivist Perspective in Semantic Metadata Interoperability

According to Crotty (1998, p. 42) constructivism "posits that all meaningful reality is contingent upon human practises, being constructed in and out of interaction between human beings and their world, and developed and transmitted within an essentially social context". The underlying assumption is that meaning is constructed and shaped from objects with the active engagement of the observer/researcher. According to Duffy and Jonassen (1992, p. 3), social constructivism posits that "meaning is imposed on the world by us, rather than existing in the world independently of us. There are many ways to structure the world, and there are many meanings or perspectives for any event or concept". This is contrary to the objectivist view that "truth and meaning reside in their objects independently of any consciousness" (Crotty, 1998, p. 42).

One may question the relevance of social constructivism for semantic metadata interoperability. Semantic interoperability, by definition, deals with problems associated to information sharing and exchange. The goal of semantic metadata interoperability is to enabling information sharing and exchange through negotiated meanings of the terms and expressions (Veltman, 2001, p. 7). The nature of knowledge in social constructivism focuses on "individual reconstructions coalescing around consensus" thus promoting shared and negotiated meaning (Guba & Lincoln, 1994, p. 112). Social constructivists assert that "realities are apprehendable in the form of multiple, intangible mental constructions, socially and experientially based, local and specific in nature, and dependent for their form and content on the individual persons or groups holding the constructions" (Guba & Lincoln, 1994, p. 110). Recent developments such as the shift towards web-based publishing media such as Wikipedia, the spread of social tagging, and the adoption of social networking applications, an overwhelming move towards the acceptance of disparate points of views and negotiated meanings, as well as a general,

implicit, tendency to arrive at a neutral point of view, all point to a need for embracing a social constructivist perspective. Recognising and accepting the existence of multiple interpretations of an object obviously has a bearing on semantic metadata interoperability as it implies and accounts for differences in the interpretations of digital objects (information resources) among individuals, groups, countries and geographic regions.

However, an examination of present practises of libraries and archives tends to demonstrate a concentration of their efforts at finding a singular solution to their information organisation problems. The Anglo-American Cataloguing Rules (AACR), MARC and Dublin Core are notable examples of such attempts. The underlying assumption, in all of these three standards and similar ones, has been that cultural heritage institutions would eventually coalesce around a single metadata standard, hence clearing the way to achieving interoperability among various information systems. Major proponents of such a top-down solution include Melville Dewey (Veltman, 2001, p. 3). Similar views regarding the organisation of digital information systems and the establishment of standards that govern their operations are still being propagated. Veltman, on the other hand, argues that the search for the single, ontologically true, metadata solution does not reflect the pragmatic reality that prevails at different institutions. As Veltman (2001, p. 7) correctly contends many of the international metadata initiatives focus "more on the universal meaning of the basic fields or elements (containers) than on the local and regional contents in those fields or elements". The question as to why all libraries do not just use a single standard might arise. The problem is related to the fact that libraries are cultural heritage institutions and culture is a fluid phenomenon. The latter's fluidity makes it difficult, if not impossible, to provide objective definitions and explanations to the objects housed in the former. Libraries and archives provide lodgings to cultural artefacts such as paintings, photographs, writings, as well as physical artefacts (eg. the Rosetta stone at the British Museum). By their very nature these objects convey different meanings for diverse user groups, and hence, can be interpreted variously. Put simply, human beings are highly unlikely to agree on a singular, top-down and hierarchical classification of objects. This assertion is likely to assume increasing importance when it comes to how museum objects, such as paintings, are depicted and interpreted. Thus, knowledge representation systems such as metadata standards should be able to reflect the various interpretations of reality. Unfortunately, most current standards tend to adhere to what is known as the ontologically and objectively true viewpoint which substantially fails to capture and represent local and/or regional perspectives and interpretations.

An attempt to overcome these shortcomings should be cognisant of the existence of a multitude of metadata standards, the prolificacy of metadata interoperability solutions, and the ubiquitous nature of digital libraries and repositories. Though these facts make the task appear daunting, one can safely assert that the problem of semantic interoperability is best addressed through collaborative approaches in which the web is considered as enabler and facilitator of such collaboration. An inherent advantage of the web is the virtual social space that it creates for fostering bottom-up collaboration. The web, especially, what Gruber (2008, p. 1) calls the 'social web' creates an "ecosystem of participation, where value is created by the aggregation of many individual user contributions". Gruber argues that such a web of collected intelligence can be combined with the potential of the Semantic Web, an approach that attaches meaning to data and integrates structured data from several sources, thereby creating new value from the data itself (Gruber, 2008, pp. 4-7). While reviewing recent developments, Shirky (2005) and Weinberger (2007) assert that collaborative tagging (folksonomy) is an enabler for implementing an effective information organisation system. On the other hand, standardisation agencies such as the Library of Congress, Online Computer Library Centre (OCLC), and Dublin Core Metadata Initiative (DCMI) tend to favour a common metadata framework that facilitates interoperability. In the middle ground are to be found the likes of Gruber (2008) who point out that both ontologies and folksonomies can be mashed up in the attempt to establish a more efficient system of information organisation.

To conclude, what is evident in the design and structure of present day metadata approaches is the lack of a theory that substantiates any one of the solutions. Since metadata constitutes a central part of digital libraries, it is of paramount importance that the choice of metadata approaches is underpinned by a theoretical framework. Considering the disparity in the nature of digital libraries, their collections and the varying user needs, a social constructivist philosophical approach should be adopted to address the issues of semantic metadata interoperability.

3. Grounded Theory Method in Metadata

3.1. The Grounded Theory Method

The grounded theory method was developed by Barney Glaser and Anselm Strauss in 1967 (for more on its origins: Bryant & Charmaz, 2007; Charmaz, 2006; Glaser, 2001; Mills, Bonner, & Francis, 2006; Strauss & Corbin, 1998). It is a well suited method for qualitative research. The basic tenet of the grounded theory method is the concept of developing a theory that is grounded in data through simultaneous data collection and analysis techniques (Bryant & Charmaz, 2007). Other characteristics of the method include avoidance of preconceived theories, pre-formulated hypothesis and the reflective and critical analysis of situations and context of a research problem or phenomena (Charmaz, 2006; Strauss & Corbin, 1998).

Currently, there are three main approaches in the implementation of grounded theory method. The first approach is called the Glaserian approach (after the originator Barney Glaser), compels the researcher to postpone the process of literature review until such time as data analysis is completed and the theory is generated. The second approach came into existence when Anslem Strauss, who was also the co-author of the method, holds views different from that of Glaser. Glaser strongly opposed to Strauss and Corbin's detail procedures of data analysis (Glaser, 1992; Strauss & Corbin, 1990). While Glaser wants to adhere to the original tenets of "The Discovery of Grounded Theory" (Glaser and Strauss, 1967), Strauss along with his colleague Juliet Corbin argues the method should be evolving in accordance with pragmatic situations (Strauss & Corbin, 1998). These differences led to a split in grounded theory methodology. The second approach also called the Straussian grounded theory method. Yet, another approach, the third flavour, is attributed to Kathy Charmaz (Charmaz, 2006) who argues that both Glaser and Strauss tend to be positivists in their treatment of the researcher as a distant and objective observer in data collection and analysis. Charmaz's approach is called the constructivist grounded theory method which follows a constructivist philosophical approach wherein both the researcher and participants mutually co-construct meaning during data collection and analysis.

Classic grounded theorists, such as Glaser, espouse the view that the researcher should keep some distance in the research process so as not to inject bias and preconceived ideas into it. As opposed to this objectivist approach, later grounded theorists especially Charmaz (2006) and Mills, Bonner and Francis (2006) adopt a constructivist approach to grounded theory, emphasise the view that the interaction between the investigator and the participant such in interviews cannot be neutral as such. Mills, Bonner and Francis (2006, p. 9) argue that through active engagements during the interview process, ideas are raised, discussed and knowledge is mutually constructed. According to this view, the researcher and the participant co-construct data, in the process known as data generation. Like Charmaz (2006), Mills, Bonner and Francis (2006, p.10) advocate for non-hierarchical intimacy, reciprocity, open interchange of ideas and negotiation (includes agreeing on the location and time of interview). The researcher also has the opportunity to reflect on his/her viewpoints and perspectives (Mills, Bonner and Francis, 2006, p.12), in a way similar to what happens during other conversations and academic discussions. By acting thus, the interviewer has the opportunity to voice his view points and perspectives as well as allowing the voices of the interviewees to be heard.

3.2. Constructivist grounded theory method for semantic metadata interoperability

As Lehmann (2010, p. 1) explains grounded theory is an appropriate method for information systems, as the domain deals with overarching components such as technology, data, procedures, and people. The patterns of behaviour, views and perspectives of users is considered as the core component hence grounded theory fits with the study of these patterns. Allan (2007, p. 1) also contends that grounded theory is a systematic and rigorous method in information systems research. He outlines how the procedures such as open coding, constant comparison, memo writing, and theoretical coding can be used to conceptualise actual problems in information science research and help to generate theory to explain patterns in behaviour, users' satisfaction and other relevant research issues. The method is especially relevant in areas where there is a scarcity of theories. The sub-category of information systems that deals with digital libraries is such a domain, as it is one where the generation and use of theories is scant (Andersen & Skouvig, 2006; Floridi, 2000; Hjorland, 2000; Lehmann, 2010). Andersen and Skouvig (2006, p. 318) argue that "for knowledge organization to uphold significance recognizable by society, it needs to engage in and be informed by theories and understandings that locate and analyze society and its historically developed forms of organization". There is therefore a need to develop theories.

As semantic interoperability is of a qualitative concern (Haslhofer & Klas, 2010, p.17), grounded theory, as a qualitative data analysis method, is a fitting methodology to explore and understand the issues as it studies actualities instead of potential applications of a solution or standard. The conceptualisation inductively generates concepts, categories and theory from users' actual experiences in using library systems and resources. A grounded theory will then emerge from the conceptualisation.

3.3. The Research Question

There are contending views in grounded theory, as to whether the research question needs to be formulated before data collection begins. On the one hand, Glaser (2001) argues that solely identifying a general research interest is adequate and the researcher should not formulate any specific research question at all. However, on the other hand, Strauss and Corbin (1998) and Charmaz (2006) contend that it is impractical to expect the researcher to delve into the research 'field' without some sort of pre-conceived research questions. This research takes the latter approach because it is argued that the research question should be first understood and stated so as to ring-fence the scope and delimit the issues that need to be addressed. According to Strauss & Corbin (1998, p. 40) "it is impossible for any investigator to cover all aspects of a problem. The research question helps to narrow the problem down to a workable size". In addition, it is also argued that that the research problem should

guide the choice of methodology (Creswell, 2003; Grix, 2004; Strauss & Corbin, 1998). In light of this argument, the following broad research questions are formulated for this research:

- What are the experiences of librarians and users in using metadata while accessing information from websites, digital libraries and information repositories?
- What kinds of solutions, in relation to semantic metadata interoperability, do librarians and users consider practical for facilitating information exchange, information sharing, and data integration?
- How much useful do librarians and users consider the Semantic Web and Web 2.0 technologies in relation to semantic metadata interoperability?
- How do librarians and users compare the value of the top-down, hierarchical approach and the bottom-up user driven approach to metadata development, in relation to semantic metadata interoperability?

3.4. Data Collection in Constructivist Grounded Theory

3.4.1. Participants

In the current study, three categories of participants are involved: academicians in the field of library and information science (including lecturers and post graduate students), librarians, and general library end-users. The selection of these participants is essentially purposive. One of the features of a grounded theory methodology is the fact that the number of participants (sample size) cannot be predetermined beforehand. Instead, the concepts and categories that emerge from the analysis of the first phase of data collection will be used to plan and implement the next phase of data collection until such time as theoretical saturation is reached. This happens when additional data fails to provide insights regarding the emergent concepts and categories (Coleman & O'Connor, 2007).

3.4.2. Choice of Research Site

The initial site selected for the study is the School of Information Studies, Tallinn University, Estonia. Beginning in 2008, the university has hosted a number of MSc students in Digital Library Learning (DILL), under the EU-funded Erasmus Mundus programme. In the initial Phase-I data collection exercise, a total of 11 participants have been interviewed, from February 21st through 25th, 2011. The interviewees consisted of 2 lecturers, 1 PhD researcher and 8 MSc students. The countries of origins of these participants include Bangladesh, China, Denmark, Estonia, Ethiopia, Italy, Turkey, Uganda, Venezuela and Zimbabwe.

3.4.3. In-Depth Interviews

In a constructivist grounded theory methodology, both the interviewer and interviewee are actively engaged in conversations. Intensive interviews are conducted mostly with open-ended questions. According to Charmaz (2006, p.25) "the in-depth nature of an intensive interview fosters eliciting each participant's interpretation of his or her experience". As recommended by Charmaz, intensive interviews should be contextual and negotiated. As part of the Phase-I data collection endeavour, introductory contacts were made via email in order to ascertain the willingness of each participant as well as reach an agreement as to the timing and venue of the interview. Before the start of each interview, a Research Background Information Sheet and a Consent-to-be-Interviewed Form were distributed. Interviews were voice recorded. Interviews were made purposefully informal so as to encourage dynamic participation in the discussions on the part of interviewees. Furthermore, rather than following a scripted question and answer approach, a more engaging approach was followed, using open-ended questions.

As pointed out by Charmaz (2006, p. 26, interviews in constructivist grounded theory enable the researcher to ask for more detail, to delve into an issue, to go back and forth among important points and request more explanation. Finally, while utilising this approach, it is also important to summarise the participant's views and reflections so that the interviewer confirms that they have been properly understood. Putting it in another way, it is essential that the participant receives "affirmation and understanding" (Charmaz, 2006, p.27).

4. Preliminary Reflections

This research is still on-going. However, in grounded theory it is permissible to reflect on issues that are discernible from participants' responses. For sure, the full details of concepts and categories are expected to emerge as an output of the data analysis process. In what follows, an attempt will be made to convey some of the tentative reflections, based on the data that has been collected so far.

4.1. Prolificacy of Standards

All participants acknowledge the existence of very many standards. Some even repeated the often cited adage that "the good thing about standards is that there are so many you can choose from", making the selection process a daunting task. One participant expressed the opinion that "libraries should base their [selection] decisions on the type of resources and the subjects they are describing". It is also pointed out that interoperability is a much sought after issue, even if it is a complicated one. Participants have alluded to the

complexity of MARC as well as the simplicity of Dublin Core, while noting that simplicity comes at the cost of metadata richness.

4.2. The Open Public Access Catalogue (OPAC)

Most participants are unanimous in that they find OPAC old-fashioned, especially in comparison to popular search engines such as Google. For example, most OPACs do not seem to have alternative spelling options. The lack of such seemingly simple features makes OPAC less useful. In addition, most OPACs do not allow users to rate, comment, review and share resources with other users. As found out from the interviews, the participants rarely go to the library in person. This is mainly because they could access the information resources from electronic information services including library databases and e-journals. One participant even mentioned the fact that he has never gone to the library during the past two years. Some respondents view the library as a place that is not important to them. Most asserted that they rarely use the library's OPAC. One participant sees the OPAC as a tool that was born to replace the card catalogue. He ironically stated that OPAC is the "biggest innovation for libraries that ever happened" believing that libraries are changing too slowly to trying to cope with users' novel needs and expectations. Another participant said that the OPAC is made for books and fits the physical attributes of the books and less to other genres of information. He cited MARC as an example of such an attempt to reutilising the descriptive standard that had been designed for books to other genres of resources such as e-journals, CDROM, music, and posters. Hence, he argues that we have now a different information landscape but a standard that is anachronistic. According to participants, most OPACs lack interactively and are mainly static.

4.3. Top-Down Hierarchies versus Bottom-Up Approaches

When it comes to classification systems (standards) and collaborative (non-standardised) approaches, the views of the participants were very diverse. However, there is a consensus among the responses that the existing classification systems and the new bottom-up approaches of tagging can be utilised together and should not be considered as opposing methods. Some are however, wary of the lack of control and structure in Web 2.0 applications such as tagging. One participant reflected on how some Web 2.0 technologies come and go. She stressed the need to answer why we use a specific technology before starting to use it in library functions. She cited the example of Second Life and how libraries jumped into the bandwagon of just being part of Second Life, while librarians creating their avatars without answering the why of such technologies. She said that, currently, the use of Second Life in libraries has diminished. She added that she does not see Web 2.0 technologies replacing the old systems of information organisation. Another participant stressed the need for libraries to provide richer description of library collections. In order to be able do this, he recommended that librarians should collaborate worldwide. According to him, librarians should be permitted, by their institutions, to catalogue collections of other institutions and vice versa, instead of relying on metadata records from proprietary companies. He pointed out that "the way Asians describe Asian art is quite different from the way a Westerner does".

Another participant stated that the issue of using standardised approaches (such as hierarchies and categories) versus Web 2.0 technologies is more of a philosophical nature than technological. In support of this, he cites the "Divine Comedy", where the organisation of the poem reflects the theoretical (philosophical) framework of Italian society at that time. According to him, the work is a complete summary of all the medieval beliefs and church teachings. Furthermore, the division of the poems is well thought out, each category having 33 divisions, which along with the introduction brings the total number of categories to 100. He then compared this with the Dewey Decimal Classification system. He indicated that both Dewey and Dante represented cultural frameworks of their societies and that they were right in their own ways. He noted that the situation now is different "because there are too many traditions altogether and we don't believe any more in a rigid, [monolithic], structure. We [do] believe in change." This change, the participant believed, has brought yet another challenge: which of the systems (standards, frameworks, systems) should libraries use in such an ever changing tradition? He also said that the existence of structures and rules in bottom-up approaches should be acknowledged. He advocated thus: "I believe that when we talk about Wikipedia, the crowd sourcing, the power of the crowd, and the bottom-up approach, we always think about democracy. It is a very beautiful world but there is always the risk of it being an empty world. Because there is an assumption that, in a democracy everyone can do what he/she wants". He continued stating that, in such a freedom there is always an obligation to act within the strictures of the community and within its accepted bounds. The limits can be as strong as a hierarchy, in which one cannot go beyond it, or they can consist of more flexible limits. It is not complete anarchy. Hence there are always laws, bounds, and limits- there is always a structure. The important question is how much does this structure allow one to accommodate the large amount of useful information?

5. Conclusions

There is no lack of metadata standards. However, the main challenge in today's digital libraries is for institutions to provide seamless access to information resources and for the users to make sense of the information they have accessed. The existence of several standards poses a technical and semantic challenge of interoperation between various digital libraries and repositories. The approaches to metadata interoperability currently focus on providing technical solutions. However, not all these methods provide the required semantic interoperability for effective cross-searching, content sharing, and information integration. Hence, semantic metadata interoperability remains to be a big challenge.

It is argued that there is a lack of theoretical framework to underpin metadata approaches and semantic interoperability solutions. The current interoperability solutions such as metadata-mapping, metadata registries and application profiles focus solely at a syntactic level, hence failing to address the semantics aspect of the problem. It is also argued that present metadata approaches are mainly top-down and the actual users are not involved. Therefore, rather than trying to force interoperability solutions around a single standard, fostering an approach that promotes and encourages diversity seems prudent, as the latter approach is more attuned to human nature and the operations of its institutions. The focus should therefore be on bridging the semantics of the elements and metadata values that are being employed in various standards and digital libraries. Cultural artefacts very often lend themselves to various interpretations and contexts. As a result most are described in varying metadata schemas, which in turn are developed at local, national, regional and international levels. Respecting and accommodating such differences, while pursing semantic consistency through a diversified approach would accrue meaningful results in the endeavour to achieve semantic interoperability. The paper focuses on solutions that respect diversity for a simple reason that a single solution or meaning cannot be enforced amidst cultural differences. It is argued that semantic interoperability does not, in any way, imply a singular understanding of a phenomenon. Instead it is mainly about allowing divergent groups to understand the intentions of each other when assigning meaning to a specific information object. Due to the very nature of the diversity inherent in institutional and cultural interpretations as well as differences in the usage of terms in metadata vocabularies, semantic metadata interoperability issues can better be addressed by adopting a social constructivist philosophical approach and by utilising a constructivist grounded theory methodology.

References

Allan, G. (2007). *The use of the grounded theory methodology in investigating practitioners integration of COTS components in information systems.* Paper presented at the Twenty Eighth International Conference on Information Systems, Montréal, Canada.

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

Andersen, J., & Skouvig, L. (2006). Knowledge organization: a sociohistorical analysis and critique. *Library Quarterly, 76*(3), 300-322.

Anderson, D., Delve , J., Pinchbeck, D., & Alemu, G. A. (2009). Preservation metadata standards for emulation access platforms: KEEP Project. Retrieved from http://www.keep-project.eu/ezpub2/index.php?/eng/Download/Public-deliverables/Deliverable-3.1

Bailer, W. (2007). *Interoperability of multimedia metadata: from digital cinema to cultural heritage*. Paper presented at the CHORUS Workshop on Metadata in Audio-Visual/Multimedia Productions and Archiving. Retrieved from http://www.ist-chorus.org/_events_RTF/documents/WSChorus_Bailer.pdf

Bake, T., Dekkers, M., Heery, R., Patel, M., & Salokhe, G. (2008). *What terms does your metadata use? Application profiles as machine-understandable narratives.* Paper presented at the International Conference on Dublin Core and Metadata Applications, Berlin.

Bryant, A., & Charmaz, K. (2007). Introduction grounded theory research: methods and practises. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 1-28). Los Angeles: SAGE Publications.

CCSDS. (2002). *Reference Model for an Open Archival Information System (OAIS)* (Blue Book Issue 1 No. CCSDS 650.0-B-1). Washington, DC: Consultative Committee for Space Data Systems.

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).

Charmaz, K. (2006). *Constructing grounded theory: a practical guide through qualitative analysis*. London: SAGE Publications.

Coleman, G., & O'Connor, R. (2007). Using grounded theory to understand software process improvement: A study of Irish software product companies. *Information and Software Technology 49 (2007) 654–667, 49*, 654-667.

Creswell, J. W. (2003). *Research design : qualitative, quantitative, and mixed method approaches* (2nd ed.). Thousand Oaks, CA. ; London: SAGE Publications.

Crotty, M. (1998). *The foundations of social research: meaning and perspective in the research process*. London: SAGE Publications.

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
Day, M. (2005). *DCC Digital Curation Manual Instalment on Metadata*.
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.
Dekkers, M. (2002). Issues in cross-standard interoperability. Retrieved from www.cores-eu.net/interoperability/d31.pdf
Duffy, T. M., & Jonassen, D. H. (Eds.). (1992). *Constructivism and the technology of instruction: a conversation*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine, 8*(4).
EC. (2010). European Interoperability Framework for Pan-European e-Government services. Retrieved 24 March 2011, from http://ec.europa.eu/idabc/servlets/Docd552.pdf?id=19529
Floridi, L. (2000). On defining library and information science as applied philosophy of information. *Social Epistemology, 16*(1), 37–49.
Gartner, R. (2008). Metadata for digital libraries: state of the art and future directions. Bristol: Technology & Standards Watch.
Glaser, B. (1992). *Basics of grounded theory analysis: emergence vs forcing*. Mill Valley, CA: Sociology Press.
Glaser, B. (2001). *The grounded thoery perspective: conceptualisation contrasted with description*. Mill Valey, CA: The Sociology Press.
Grix, J. (2004). *The foundations of research* New York: Palgrave Macmillan.
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.
Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. Chapter 6 in N.K. Denzin & Y.S. Lincoln (Eds) SAGE. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: SAGE.
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., & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. *ACM, 42*(2).
Heery, R., Johnson, P., Fülöp, C., & Micsik, A. (2003, 28 September – 2 October 2003). *Metadata schema registries in the partially Semantic Web: the CORES experience.* Paper presented at the Dublin Core Conference: Supporting Communities of Discourse and Practice – Metadata Research & Application, Seattle, WA.
Heery, R., & Patel, M. (2000). Application profiles: mixing and matching metadata schemas. *Ariadne* (25).
Hillmann, D. I., & Phipps, J. (2007). *Application profiles: exposing and enforcing metadata quality.* Paper presented at the International Conference on Dublin Core and Metadata Applications.
Hjorland, B. (2000). Library and information science: practice, theory, and philosophical basis. *Information Processing and Management 36*, 501-531.
Lehmann, H. (2010). *Grounded theory and information systems: are we missing the point?* Paper presented at the The 43rd Hawaii International Conference on System Sciences, Koloa, Kauai, Hawaii.
Miller, P. (2000). Interoperability what is it and why should I want it? *Ariadne*(24).
Mills, J., Bonner, A., & Francis, K. (2006). Adopting a constructivist approach to grounded theory: Implications for research design. *International Journal of Nursing Practice, 12*, 8-13.
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.* KTH School of Computer Science and Communication, Stockholm.
NISO. (2004). Understanding metadata. National Information Standards Organization. Retrieved from http://www.niso.org/publications/press/UnderstandingMetadata.pdf
OCLC/RLG. (2001). *Preservation metadata for digital objects: a review of the state of the art*: OCLC/RLG.
Ouksel, A. M., & Sheth, A. (1999). Semantic interoperability in global information systems: a brief introduction to the research area and the special section. *ACM SIGMOD Record, 28*(1), 5-12.
Rothenberg, J. (2008). Interoperability as a semantic cross-cutting concern. *Interoperabiliteit: Eerlijk zullen we alles delen*. Den Haag.
Shirky, C. (2005). Ontology is overrated: categories, links, and tags. http://www.shirky.com/writings/ontology_overrated.html
Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks, CA: SAGE Publications.
Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks, CA: SAGE Publications.
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. (2004). OWL: Web Ontology Language overview. Retrieved from http://www.w3.org/TR/owl-features/
Weinberger, D. (2007). Everything is Miscellaneous. New York: Times books.
Wright, A. (2007). *Glut: mastering information through the age*. Ithaca: Corenell University Press.
Zeng, M. L., & Chan, L. M. (2006). Metadata interoperability and standardization – a study of methodology part II: achieving interoperability at the record and repository levels. *D-Lib Magazine, 12*(6).