Repositório ISCTE-IUL

From the web of bibliographic data to the web of bibliographic meaning: structuring, interlinking and validating ontologies on the semantic web

Helena Simões Patrício; Maria Inês Cordeiro; Pedro Nogueira Ramos

Biblioteca Nacional de Portugal, Lisboa, Portugal; Information Science, Technologies and Architecture Research Center (ISTAR-IUL); University Institute of Lisbon (ISCTE-IUL)

Emails: hpatricio@bnportugal.gov.pt; icordeiro@bnportugal.gov.pt; pedro.ramos@iscte-iul.pt

Abstract: Bibliographic datasets have revealed good levels of technical interoperability observing the principles and good practices of linked data. However, they have a low level of quality from the semantic point of view, due to many factors: lack of a common conceptual framework for a diversity of standards often used together, reduced number of links between the ontologies underlying datasets, proliferation of heterogeneous vocabularies, underuse of semantic mechanisms in data structures, "ontology hijacking" (Feeney et al., 2018), point-to-point mappings, as well as limitations of Semantic Web languages for the requirements of bibliographic data interoperability. After reviewing such issues, a research direction is proposed to overcome the misalignments found by means of a reference model and a superontology, using SHACL (Shapes Constraint Language) to solve current limitations of RDF languages.

Keywords: Linked Open Data; Bibliographic data; Semantic Web; SHACL; LOD (Linked Open Data) validation; Ontologies; Reference model; Bibliographic standards

Biographical notes: Helena Patrício is Director of Special Collections Services at the National Library of Portugal (NLP) since 2012, in charge of the National Digital Library and the NLP management in national and European projects. She was Director of Information Systems at the NLP (2007-2012). She graduated in Law (1996), has a post-graduation in Library and Information Science (1998) and a master in Information Studies and Digital Libraries (2004) with the thesis "Comparative analysis of the relational model and the RDF applied to the conceptual modelling of legislative information". She is a member of LIBER Linked Open Data Working Group and a PhD student, researching in the field of semantic web and linked open data.

Maria Inês Cordeiro is Director-General of the NLP. She holds a PhD in Library and Information Studies from University College London, a post-graduation in Librarianship and a degree in History. She held several management positions at the NLP and at the Art Library of the Calouste Gulbenkian Foundation. For almost thirty years she served in various IFLA committees and working groups as well as Chairman, Editor-in-chief and advisor at the UDC Consortium. Her research interests include knowledge organization, information systems and library technologies.

Pedro Ramos is Associate Professor and researcher at ISCTE-IUL, Lisbon. He holds a PhD in Science and Information Technology, Master in Management Information Systems and a BSc in Management. He coordinates several IT courses, masters and research activities in Information Systems at ISCTE-IUL and has large experience in the development of computer applications to the industry. His scientific research interests and publications are in Formal Organization Modelling, Data Modelling and Big Data Management.

1 Introduction

Literature review shows that bibliographic data publication in the Semantic Web has been carried out according to linked data technologies and principles, assuring good levels of technical interoperability (Jain et al., 2010b; Hogan et al., 2012; Dutta, 2017; Talleras, 2018). Problems do not usually arise in the publication of isolated sets of linked data and in the exchange of information regarding data instances. However, they appear in the sharing of their intended meaning, i.e., at the semantic interoperability level, understood as the capacity of two or more systems to exchange ontologies. As Dutta (2017) points out, there is a great emphasis on data publication as linked data, but little attention is given to the description of such data in terms of concepts, properties and relationships between datasets. Linked datasets have thus an expressivity problem, being mere triple collections that do not use the power of ontology languages such as RDFS and OWL.

Problems with the semantic interoperability of bibliographic datasets are deepened by syntactic issues of RDFS/OWL languages, whose permissiveness does not allow validation of data. This requires the use of other languages, aligned with the Open World paradigm of the Semantic Web, that are able to detect errors and inconsistencies in bibliographic data structures.

This paper is structured as follows. In section 2 - "Data quality in the Semantic Web" the major background concepts related to metadata sharing and data quality are reviewed, in order to establish a framework for linked data quality evaluation. Subsequently, in section 3 - "Bibliographic ontologies evaluation", this paper updates the literature review about the major interoperability limitations of bibliographic ontologies undertaken by Patrício, Cordeiro and Ramos (2019) and published in MTSR 2019 Proceedings

(Garoufallou et al., 2019), both at the *top-down* level of bibliographic standard vocabularies and at the *bottom-up* level of linked data library implementations. A research project is presented in Chapter 4 – "A reference model and a SHACL-based superontology as solutions" to investigate solutions involving the creation of a reference model that could conceptually frame the heterogeneous initiatives of bibliographic data transformation and the development of a superontology, that could function for high level semantic interlinking and data validating, using SHACL, thus differing from traditional bibliographic control systems and XML/database mapping solutions such as crosswalks and application profiles.

2   Data quality in the Semantic Web: background concepts and evaluation framework

Several authors have discussed frameworks for data quality concepts adapted from the database "closed world" to the semantic web context (Hogan, 2012; Zaveri, 2012; Kontokostas, 2014; Schmachtenberg, 2014; Farber, 2016; Dutta, 2017). Before presenting these frameworks it is important to review some core concepts in metadata interoperability.

Interoperability is the ability of two or more systems to share information. This involves two basic components: the syntactic aspects of "gaining access to the shared data", finding a common communication medium; and the semantic issues of "incorporating that information into the data structures of the consuming system", choosing a shared language (Hebeler et al., 2009, p. 65).

In the information domain, interoperability implies not only the transport and communication of data, but also the understanding and reuse of what is communicated. For this purpose, interoperability relies on the existence of metadata that supports data coherence, consistency, quality and integration capabilities, by defining and documenting data structure, syntax, semantics and behaviour (Cordeiro, 2005).

Technical interoperability refers to data exchange protocols, Web protocols in the case of linked data. Syntactic interoperability is related to data exchange formats, which in the Semantic Web are languages such as RDF, RDFS and OWL. Semantic interoperability consists not just in the simple exchange of information but in the sharing of an intended meaning. To be so, the information exchanged is described in an unambiguous and machine interpretable way, using ontologies. In short, semantic interoperability consists on the ability of two or more systems to exchange and use ontologies. Therefore semantic issues can only be effectively addressed by semantic web languages (Hebeler et al., 2009) that can create ontologies that provide metadata vocabularies for the exchange of data semantics (Antoniou et al., 2012).In this paper, we will use the term "ontology" in the TBox sense (Talleras, 2018), as a high level representation comprising concepts, properties and constraints. We will not use the term "metadata scheme" because ontologies are more complex systems that provide rules of inference and descriptive logic for computational reasoning (Talleras, 2018). We will also apply the expressions "element set" or element vocabulary as

synonyms of ontology (NISO, 2017). As for the so-called ABox ontologies (Talleras, 2018), which consist of data or instances generated according to the TBox, we prefer the term "datasets".

Given the scope of our research, we are interested only in analyzing quality dimensions that relate most directly to technical, syntactic and semantic interoperability issues, leaving aside factors such as accessibility, contextual circumstances that depend on the execution of specific tasks, versatility, etc.

In respect to data quality, we share the view that quality is not an absolute measure, but rather an assessment of "fitness for use" or suitability of data against a particular use-case (Kontokostas, 2014), i.e., corresponding to user's needs rather than conforming to a specification (Juran and Godfrey, 1998). In the context of the Semantic Web, "fitness for use" means the capacity to relate to other communities in the Web of Data, encompassing functions such as integrated access to data; enrichment of data to support their inter-relationship and contextualization; and improved visibility and reuse of data (Hogan, 2012; Dutta, 2017; Talleras, 2018). In addition, metadata cannot be studied without considering the domain or discipline that they relate to (Greenberg and Garoufallou, 2013).

In order to fulfil the tasks of semantic navigation and exploitation, at the instance level bibliographic data must have the quality that allows applications to locate, parse, retrieve, discover and consume it (Hogan, 2012). At the level of ontologies, the quality that enables navigation and semantic search relies on the interlinking, reuse and reasoning allowed by their semantic constructs (Hogan, 2012).

In this context, the semantic limitations of bibliographic ontologies identified in the literature will be analysed in three levels: i) technical (compliance with the principles of linked data); ii) syntactic (use of RDF/RDFS/OWL languages) and semantic (conceptual framework, data structures and interconnection of ontologies), as shown in Fig. 1.

Figure 1 Quality dimensions per interoperability level
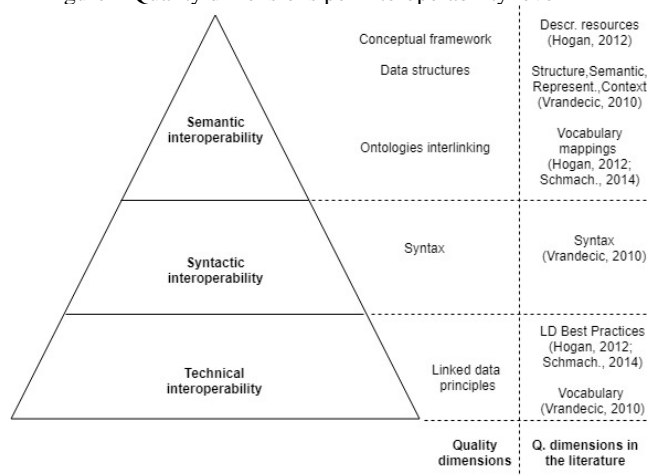
3   Bibliographic ontologies evaluation

The following sections summarize the bibliographic ontologies interoperability problems already reviewed in

Patrício, Cordeiro and Ramos (2019), adding quality evaluations of general linked data implementations undertaken by Hogan (2012), Schmachtenberg (2014) and Feeney et al. (2018). And, specific to the library domain, the evaluation by Papadakis, et al. (2015) of linked datasets from 7 national libraries; the analysis by Jett et al. (2016) regarding linked data representations of normative bibliographic ontologies; the Talleras (2018) comparative evaluation of bibliographic datasets published as linked data by France (BNF), Germany (DNB), UK (BL-BNB) and Spain (BNE) national libraries. Finally, the results of the OCLC survey conducted from April 2018 to 81 institutions, including 13 national libraries (Smith-Yoshimura, 2018) are also taken into account.

3.1 Technical interoperability: conformance with linked data principles

Linked data principles and best practices correspond generally to the five-star rule defined by Tim Berners-Lee (2006) and more recently developed in two W3C recommendations and technical notes (Hyland et al., 2014, Lóscio and Burle, 2017), aimed at ensuring that linked datasets comply with Web protocols and are, therefore, technically interoperable. The majority of authors conclude that bibliographic datasets are generally in line with the essential principles of linked data. However, as far as bibliographic standards are concerned, linked data principles are not fully observed as their elements are not completely described in machine-processable form, being oriented for human consumption mostly, rather than for automated reasoning (Talleras, 2018).

Despite the general observance of good data practices and the problem mentioned by Talleras, there are other specific problems identified in the literature about bibliographic datasets, which we will group in the following categories, according to principles stated by Hogan (2012): Naming; Dereferencing and RDF use; and URI outlinks.

Naming

According to this principle, entities of a dataset must be identified using URIs, i.e., unambiguous references. According to Talleras (2018), there are some failures, because in most datasets libraries use literal values for the identification of entities, only assigning URIs to a minority of entities (39% in BNB, 12% in BNE, 36% in BNF and 20% in DNB). Other non-conformities, such as the use of blank nodes by BNE and DNB, make external links to these resources impossible, data merging difficult and prevent indexing by crawlers (Hogan, 2012).

Dereferencing and RDF use

Dereferenceable URIs means that HTTP protocol is used in naming entities, so that they can return information about a resource, "looking up" its name. In effect, an HTTP URI is a Web address that can be accessed to retrieve information about the identified entity. Using Web technology, this data can be easily accessed by people and applications.

Unlike in most domains where datasets are highly compliant with this recommendation (Hogan, 2012), in the case of linked bibliographic datasets problems occur. As verified by Papadakis (2015) the majority (BL, BNE, LIBRIS and National Széchény Library) of analyzed datasets were not dereferenceable.

On the other hand, when someone "looks up" a URI, the information must be provided in RDF. HTTP URIs must return RDF representations, at least in RDF/XML format. Hogan (2012) reports that 30% of the analyzed datasets are not conformant with this principle.

URI outlinks

This principle recommends using URIs to identify entities in external datasets, that is, the object of an outlink must be a URI in an external dataset and that URI must be dereferenceable. In fact, to find entities from other systems, the links to those datasets must also return remote documents and not mere literal values.

Of the datasets analyzed by Schmachtenberg (2014), 56% had at least one link to another dataset. The category with more links is that of social networks; and publications (including libraries) is the category with the least outlinks.

In respect to links between data instances, where the subject is a local entity and the object is an entity of an external dataset, Schmachtenberg (2014) reports that, in general domains, there is a low number of external links (8% on average). This low percentage of outlinks also occurs in bibliographic datasets, where links to entities of the same communities prevail (Papadakis, 2015). Accordingly, Talleras (2018) reports that the 4 bibliographic datasets analyzed link to 28 external datasets only, of which 11 are bibliographic and only 4 are clearly from other communities, with emphasis on DBPedia and Geonames. That is, the links to datasets with large expression are very few, prevailing the links to bibliographic datasets as it is evident from the type of properties most used: dct:language (for language codes) and rdfs:seeAlso (for Dewey notations) (Schmachtenberg, 2014).

Regarding links between ontologies, Jett et al. (2016), in the analysis of normative ontologies, verified that all of them, with the exception of BIBFRAME, use literal values instead of URIs in the links to elements of other ontologies. The use of literals to describe information that exists elsewhere is highly redundant and inhibits the effect of automatically aggregating additional information about a given element.

3.2 Syntactic limitations

In what concerns the use of semantic web languages such as RDF and OWL, several issues arise: the low adequacy of RDF granular and atomized elements for the purposes of bibliographic data (Yee, 2009); the need to avoid complex RDF constructs like reification, collections and containers, because of their lack of semantics or unclear usage (Hogan, 2012; Farber, 2016); and some incompatibilities on the combination of OWL and RDF/RDFS (Feeney et al., 2018).

However, the major syntactic problem is the "permissivity of linked data languages" (Feeney et al., 2018) or the missing capability of OWL and RDFS for defining

constraints to impose data structures validation. In fact, the use of RDF restrictions such as "range" and "domain" to constrain the use of a property to a certain class allows the inference of new information only, not its validation. This is because reasoners do not detect errors and may infer RDFS/OWL descriptions that are formally correct despite containing errors (Feeney et al., 2018). On the other hand, due to the Semantic Web "AAA Principle" that states that Anyone can say Anything about Anything, in RDF the use of a given property cannot be limited to a given class.

This limitation has a major impact on linked data implementations of bibliographic ontologies, because the specification of data structures that can be validated against certain constraints is a requirement of multi-entity models such as FRBR (Functional Requirements for Bibliographic Records): as each entity has its own attributes or properties, their restrictions will require the use of languages other than RDF (Baker et al., 2014).

In addition, it is not possible to use RDF to define a hierarchy like the FRBR WEMI (Work, Expression, Manifestation, Item) entities relation, because RDF has a graph rather than a hierarchical or tree structure. Therefore RDF can connect to virtually everything in any direction (Yee, 2009) and expresses transitive properties and classes for inference only. That is, RDF does not solve the historical problems of lack of transitivity of bibliographic data models.

3.3 Semantic interoperability issues: conceptual, structural and vocabulary limitations

Issues of semantic interoperability or potential integration of data from different sets referred to in the literature will be approached in three dimensions of analysis: adequacy of conceptual models to the Semantic Web paradigm; semantic mechanisms in data structures; and reuse of external vocabularies.

3.3.1 Adequacy of conceptual models

From this perspective, it is very important to understand how well aligned bibliographic conceptual models are with the new Semantic Web paradigm, since modelling languages are built with a certain paradigm in mind, which constrains its applicability (Cordeiro, 2005). The most relevant misalignments occur because the FRBR model and its RDF representations do not fit well the Semantic Web and because there are conceptual problems arising in bibliographic ontologies as they lack a reference model.

a) Limitations of the FRBR model

According to some authors (e.g., Murray, 2008; Willer and Dunsire, 2013), FRBR is not aligned with the Semantic Web paradigm because its elements derived from standards prior to it

In fact, FRBR model is based on requirements defined for legacy systems such as card catalogues or MARC (Machine Readable Cataloging) formats, resulting in the creation of entities, attributes and relationships from pre-existing standards. Besides, as explained in Patrício, Cordeiro and

Ramos (2019), FRBR entities are not data structures designed to be connected, making it difficult for RDF descriptions to co-exist with data from other communities (Murray, 2008).

Patrício, Cordeiro and Ramos (2019) summarised the most relevant misalignments of RDF representations of FRBR (such as FRBRer, FRBR Core and FRBRoo ontologies) with the techniques of linked data, as reported by several authors (Murray and Tillet, 2011; Peponakis, 2012; Baker et al., 2014; Martin and Mundle, 2014; Coyle, 2015; Godby et al., 2015; Coyle, 2016; Zapoudinou et al., 2016). First aspect to note is the inability to express class hierarchy, thus not enabling transitivity and basic mechanisms of inference. Consequently, the entities below the WEMI sequence are unable to use the attributes of the higher entities, resulting in a sequence of instantiations rather than a hierarchy. Other limitations of RDF implementations derive from their inadequacy to the perspective of a multi-entity model of entities as points of view, because FRBR ontologies make a strict demarcation of WEMI entities and specify with little clarity the relations between them.

The IFLA Library Reference Model (IFLA-LRM) approved in August 2017, is a editorial consolidation of the FRBR family of models, intended as a single and coherent model better adapted to the Semantic Web (Peponakis, 2016; Riva et al., 2017). Despite Riva (2016) conviction that FRBR semantic issues are overcome with IFLA-LRM, it seems relevant to ask whether convergence with the Semantic Web has improved with the new model and to analyse the transformation initiatives that, meanwhile, will appear.

b) Lack of a common conceptual framework

The multiplicity of bibliographic ontologies, understood as both vocabularies of standard elements and local LOD implementations, reveals the need for a common conceptual model that could prevent contradictions in their combined application (Yee, 2009). As Sprochi (2016) points out, a reference model is needed to frame the different levels of bibliographic standards that are closely related and strongly dependent on one another for implementation.
The main criticisms about the RDF publications of bibliographic standards, such as ISBD (International Standard Bibliographic Description), RDA (Resource Description and Access), MARC and BIBFRAME, were reviewed in our previous paper (Patrício, Cordeiro and Ramos, 2019). Regarding ISBD and MARC LOD representations, conceptual problems identified in the bibliography can be briefly summarised as the lack of a model based on entities and relationships, as it is typical of the Semantic Web, contrasting with the flat model underlying the bibliographic record as a text (Svensson, 2013; Willer and Dunsire, 2013; Szeto, 2013).

These limitations motivated the development of new standard bibliographic ontologies born in the Semantic Web context such as RDA and BIBFRAME. According to several authors (e.g., Szeto, 2013; Coyle, 2016), RDA is completely compatible with the Semantic Web because it implements FRBR as a multi-entity conceptual model. BIBFRAME also

appears among the ontologies most compatible with the open Web because, unlike FRBR, it uses class hierarchy and does not define disjunctions between classes (Coyle, 2016). Some authors (e.g., Peponakis, 2016) point out significant differences between the FRBR model and RDA which may justify a deeper analysis of this bibliographic standard.

In the context of bibliographic linked data there is a tension between *top-down* approaches – like RDA, BIBFRAME and other bibliographic standards consisting of holistic ontologies, with unique names for classes and properties (Talleras, 2018) aiming at exclusive instantiations (Vrandecic, 2010) - and *bottom-up* approaches - carried out by data transformation initiatives led by libraries that apply different ontologies by mixing local elements with elements from external vocabularies - that can lead to conceptual inconsistency. The existence of a framing conceptual model would alleviate this tension, contributing to a more consistent relationship between the different kinds of bibliographic ontologies.

An example of this type of conflict is presented by Talleras (2018) concerning the FRBR model: despite being considered as a standard for bibliographic data in the Semantic Web, it is implemented in very different ways by the four national libraries he analyzed.

For example, BNE and BNF ontologies implement *Work, Expression* and *Manifestation* entities while DNB and BNB ontologies only have *Manifestation* type entities. Relations between entities are also very different, since the BNE establishes inverse relationships between all entities and BNF neither establishes a relationship between *Expression* and *Work* nor between *Expression* and *Manifestation*. In respect to the responsibility relationship, there are also significant differences, since the BNF describes in detail the attributes of responsibility (470 properties), specifying relationships of both *creator* and *contributor* with both *Work* and *Expression*; while the BNE establishes relationships of *creator* with the *Work* entity only and of *contributor* with the *Manifestation* entity only (Talleras, 2018).

Figure 2 - WEMI implementations (Talleras, 2018)
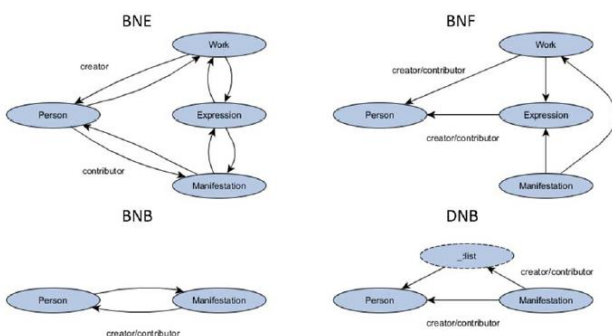
Talleras (2018) notes that BNF and BNE risked more in the FRBRization of data than DNB and BNB, which are more oriented to the *Manifestation* entity and use slightly different models, much influenced by their legacy data. In the relationship with the *creator*, the DNB system relies on RDF containers to list the creators in an orderly fashion; the BNB establishes inverse relationships with *Manifestation,* in a way

similar to that of BNE, but admitting relationships of the *creator* and *contributor* type.

It can be concluded that the lack of a common conceptual model causes the development of ontologies based on very different models and a mix of standards with different abstraction levels. For these reasons, Suominen and Hyvonen (2017) declare that libraries are risking to abandon "silos" of MARC data to adopt "silos" of linked data.

3.3.2 Semantic mechanisms in data structures

In this dimension of analysis we focus on problems regarding the way resources are described (Hogan, 2012), which relate to both the declared structure of data and the inference of new statements from explicitly declared triples (Vrandecic, 2010). In this section two aspects are highlighted: the poor use of semantic mechanisms and the proliferation of vocabularies.

a) Poor use of semantic mechanisms

Bibliographic ontologies are not taking advantage of all the potentialities of linked data technologies, by not using basic mechanisms like classification or class-level relationships. For example, many local ontologies make direct use of external classes, applying them at the instance level only, and not at the vocabulary level. This prevents inference of all instances of a given local class as instances of the external class, imposing the classification of each instance at the data level.

Another example is the little use by RDA of class hierarchy (Coyle, 2016); for this reason the addition of a new class obliges to define relationships at the instance level because it is not possible to infer new relationships from already established relationships with any superclass to which the new class would belong.

Finally, data constraints in bibliographic standards are often declared with textual notes, not formalized with inference languages' constructs.

b) Proliferation of vocabularies

The proliferation of bibliographic vocabularies and the absence of good practice in vocabulary development and management is causing many problems in library implementations (Hanneman and Kett, 2010; Dunsire et al., 2012a; Hallo et al., 2016, Suominen and Hyvonen, 2017).

At both the levels of standards and local data transformation initiatives, there is an excessive number of bibliographic ontologies whose heterogeneity, overlap and lack of interconnection make data search, integration and reuse difficult (Jain et al., 2010a, Willer and Dunsire, 2013, Talleras, 2018).

In *top-down* approaches, bibliographic standards in RDF replicate their original and structural heterogeneity, leading to inconsistency or even incompatibilities, as exemplified by Zapounidou et al. (2016) regarding FRBR and BIBFRAME.

As for *bottom-up* LOD implementations, each library selects the external ontologies to be used for publishing datasets and/or defines a local vocabulary of elements. This mixture of ontologies and the creation of new elements/properties may not fit the data to be modelled

(Hanneman and Kett, 2010) or can even hamper the combination and use of those elements together (Suominen and Hyvonen, 2017).

The proliferation of local bibliographic ontologies is well evident from both the results of the OCLC survey, where 22% of organizations reported using local vocabularies (Smith-Yoshimura, 2018) and the analysis conducted by Talleras (2018), concluding that the use of local vocabularies in the four sets of bibliographic data analysed was 70.4%., on average, with each of the datasets using different and exclusive elements to express the same FRBR entities.

3.3.3   Reuse of external vocabularies

In the Semantic Web, conceptual vocabularies for information sharing between systems are specified by ontologies, facilitating interoperability between multiple and diverse systems (Gruber, 2009). As vocabularies of elements that provide the correct interpretation for linked data elements, making them self-descriptive (Hawtin, 2011), ontologies are fundamental for semantic interoperability. In addition, the reuse of external ontologies' properties and classes enables data interpretation and processing by applications (Talleras, 2018) and potentiates reasoning (Jain et al., 2010a).

In this context it is important to analyse bibliographic ontologies, identifying their major semantic limitations: outlinking problems, ontology hijacking and point-to-point mappings.

a) Outlinking problems

In the development of bibliographic ontologies, a *cherry-picking* methodology (Godby, 2016) has been followed, meaning the use of elements from external vocabularies mixed with local classes and properties. As a good practice to reduce the heterogeneity of datasets and increase their visibility for external communities, it is better to *cherry-pick* than to create "island" terms without any link to external ontologies (Hogan, 2012).

However, *cherry-picking* methodology is only useful if links to the external elements are included; but unfortunately, this is not the practice followed by bibliographic ontologies, as pointed out by Patrício, Cordeiro and Ramos (2019) regarding the low outlinking in FRBRer, ISBD and RDA ontologies. As for bottom-level ontologies developed by libraries, in the four sets of bibliographic data analyzed by Talleras (2018) only 28 target ontologies were identified, of which only 8 are shared by at least two datasets. On the other hand, for data element vocabularies, only 3 properties (owl: sameAs; rdf: type and dct: language) are shared by the four datasets (from a global universe of 1,141 properties). In their analysis, Jett et al. (2016) also noted that ontologies do not contain explicit declarations of equivalence between classes, a feature in developing ontologies that can be due to the semantic uncertainty caused by poorly documented ontologies.

While bibliographic standards are not the most frequently referenced by bibliographic datasets (Schmachtenberg, 2014), BIBFRAME is the most reused ontology (Smith-Yoshimura, 2018). Talleras (2018) reveals that none of the 38 external vocabularies used by the four datasets he analysed are important bibliographic normative ontologies such as BIBFRAME and FRBR. As for the reuse of local ontologies, the OCLC survey shows that British Library Terms (BLTerms) is referenced by one dataset only (Smith-Yoshimura, 2018).

In short, because of the small number of outlinks, bibliographic ontologies do not benefit from the reuse of known vocabularies that can support interoperability and increase usability by third parties.

Another problem is the direct use of external elements at the level of data instances, that is, without explicit alignment with other vocabularies, through declarations of equivalence. The absence of such links does not allow bibliographic ontologies to benefit from advantages such as creating very precise ontological structuring, linking to more generic vocabularies (Jett et al., 2016) with the consequent search engine optimization. The absence of external elements duly formalized may be caused by the experimental nature of many bibliographic data transformation initiatives or, as Godby (2016) noted, by the lack of time for discussion and integration of elements of pre-existing ontologies; and also, as reported by 28% of the 2018 OCLC survey respondents, by difficulties in data alignment, matching and disambiguation (Smith-Yoshimura, 2018).

The most serious lack of interlinking between ontologies occur with the so-called "proprietary ontologies" (Schmachtenberg, 2014), which are those not reused by any other external vocabulary, i.e., that are used by a single dataset only. The analysis of datasets led by Schmachtenberg (2014) showed that 59% of the vocabularies represented in the LODCloud are proprietary ontologies, a percentage that reduces to 34% in the category of Publications.

b) Ontology hijacking

The formal and explicit definition of elements by ontologies does not mean that these definitions are followed when used "in the wild", and errors may occur in links to elements from other external ontologies (Hogan, 2012). In fact, it is difficult to ensure consistency in the reuse of ontologies, especially when they are developed independently and their components are later combined. It is in this context that Feeney, Brennan and Gleansong (2018) refer to "ontology hijacking" problems.

Ontology hijacking can cause "uncoordinated interoperability" errors (Feeney et al., 2018) when each ontology makes external references according to a perspective and scope of its own. Such references are not modular, therefore they may be inconsistent when combined. It may also happen that statements link two different ontologies wrongly. For example, when an external class is referenced as being a property in a local ontology (Feeney et al., 2018).

Another problem of hijacking relates to the ontologies' lifecycle, consisting on "ontology degeneration" (Feeney et al., 2018) that results in "orphan" or "zombi" vocabularies (NISO, 2017). In this case, referenced ontologies may become unavailable or changed thus becoming incompatible

with the local ontology (Feeney et al., 2018). Regarding the unavailability of external ontologies, these authors conclude that 12% of the analyzed ontologies suffer from this. RDA ontologies, for example, use terms from two ontologies that no longer exist (http://metadataregistry.org/uri/profile/regap and http://metadataregistry.org/uri/profile/rdakit); and in FRBR Core, the frbr:Work and frbr:Event classes are defined as subclasses of non-existent external ontology elements.

But the most serious problem of "hijacking" occurs when ontologies explicitly alter other ontologies, which happens in most cases through the use of equivalence relations. To avoid such situations, it is always preferable to use hierarchical relationships or simply to directly apply the external ontology element. In the analysis conducted by Feeney et al. (2018), the ontologies that violate third ontologies mostly are FRBR Core (32 detected violations, changes in the ontologies foaf, dc, cc, geo, rdf, among other) and BIBO (50 violations in dc, foad, rdf and rdfs).

Regarding violations of third party ontologies, Kontokostas et al. (2014) performed tests against all the ontologies referenced by the BNE dataset, with 11 thousand errors reported in relation to FRBRer, 37 thousand in relation to DCTerms and 28 million regarding ISBD. Most of these errors are violation of the rdfs:range and rdfs:domain properties of the external ontologies, as well as disjunction properties.

All this reinforces the urgent need for a validation mechanism able to ensure the correctness of datasets regarding the ontologies that they apply.

c) Point-to-point mappings

Since in bibliographic ontologies there is a low usage of links to external vocabularies and the reuse of bibliographic standards is even scarcer, mappings between elements of different ontologies become even more relevant for interoperability. Although there are several alignments between bibliographic ontologies, we are not aware of any standard ontology created at a higher level to express semantic relations between them.

Mappings have been made trough point-to-point connections between elements of different ontologies, that work for 1:1 relationships, but do not ensure semantic interoperability in 1:* or *: 1 relationships and do not solve situations of mismatch as well (Howarth, 2012). An example of these distributed mappings is the alignments made by the IFLA ISBD Working Group, relating ISBD with the external vocabulary RDA. As the alignments are made unidirectionally from ISBD to RDA, the reverse mapping, from RDA to ISBD, is also needed (Escolano Rodriguez, 2016). Creating a central ontology to represent the semantic connections between these vocabularies at a higher level of abstraction would prevent situations like these.

In respect to local library ontologies, interoperability problems of the datasets analyzed by Talleras (2018) result from the application of linked data principles following a methodology based on application profiles, mixing elements from different standards. The use of application profiles and other database and XML technologies, such as "crosswalks" or schema-to-schema mappings, facilitates the exchange of data between different schemas but does not solve semantic compatibility issues (Howarth, 2012). Usually, *crosswalks* are neither available as separate resources, nor used beyond the organizations that create them, thus not being suitable for an open, global and shared environment as the Semantic Web (NISO, 2017). Database and XML mapping concepts differ from semantic mapping mechanisms (Doerr et al., 2012) that are based on ontologies' connections. In a context of open data multidimensional perspectives, unique/central maps do not exist and data transformation is a process distinct from mapping, with different semantic maps being shareable as independent resources (Dunsire et al., 2012b; NISO, 2017). This is the mapping paradigm that will frame our research.

4. A reference model and a SCHACL-based superontology as solutions

The right context for inscribing the solution we propose to investigate seems to be the creation of a reference model, using Semantic Web standards as a *lingua franca* to solve problems of heterogeneity between domains and datasets (Talleras, 2018), because RDF "open world" mechanisms enable the combination of multiple data sources in bibliographic descriptions, aggregating multiple viewpoints about a resource.

The first step is to investigate if higher abstraction mechanisms can potentiate semantic interoperability, at two levels: the creation of a reference model capable of encompassing the different models existing in the bibliographic and similar domains; and the specification of a superontology based on the reference model, i.e., a reference ontology with a level of abstraction higher than existing standard and local ontologies, in the sense defined by Brinkley (2006).

The reference model would function as a high level conceptual framework for bibliographic information that could be used in an unambiguous and consistent way by the various specific implementations, besides being able to relate them. Because the scattered nature of traditional bibliographic standards has been replicated in their RDF publications, a reference model to frame bibliographic standards could provide for better consistency among them and the quality of their inter-relationship. This model would take into account not only bibliographic models, but also experiences, standards, ontologies and reference models in related domains like museums, institutional repositories, digital libraries (Garoufallo and Papatheodorou, 2014) or in the Learning Object Metadata domain (Balatsoukas et al., 2011; Balatsoukas et al., 2012). As stated by Willer and Dunsire (2013), the need is for rethinking the more abstract models, rather than just define a new framework for old data elements.

Following the understanding of authors such as Jain et al. (2010a), Jett et al. (2016) and Feeney et al. (2018), we consider that the problems arising from the lack of semantic links between ontologies can be solved with an upper level ontology, to be further integrated by domain-specific ontologies, in order to improve discovery of knowledge, increase reasoning capabilities and enable consistency checks (Jain et al., 2010a). In this same sense, Jett et al.

(2016) advocate the development of ontologies that overlap other ontologies, to bridge among ontology classes. Feeney et al. (2018) suggest the creation of a unified model of ontologies able to combine heterogeneous linked data vocabularies in a consistent logical model.

The superontology would be an instrument for relating semantically the elements of standard bibliographic vocabularies and for specifying mechanisms for restricting or constraining bibliographic data. As already explained, RDF languages have limitations that make them inadequate for certain interoperability requirements that imply restrictions and constraints. Despite the multiplicity of points of view of an "open world", the need still remains for a solution to "close the world" when needed to constrain and validate data structures.

SHACL (Shapes Constraint Language), a high-level vocabulary for the expression of data constraints which is simultaneously a language for ontologies, approved as a W3C Recommendation in July 2017, emerged from this need. It allows the specification of constraints (called "shapes") for the validation of RDF graphs (Knublauch and Kontokostas, 2017), being also more powerful than OWL in inference mechanisms because it can be used for rule-based inferences (Knublauch, 2017).

As a formal and standard syntax for constraints implementation, SHACL enables them to be processable by machines. In addition, another formal language, the W3C RDF vocabulary PROF (The Profiles Ontology) (Atkinson and Car, 2019) can help to describe our superontology as a SHACL resource that defines and implements specifications and sets of constraints on the use of more general vocabularies such as bibliographic standards or library ontologies, to increase semantic interoperability. PROF can enable the superontology specification using a formal, machine-readable language.

The use of SHACL and PROF vocabulary in our superontology solution differs from the usage of text or "platform specific" languages in the development of metadata schema, application profiles (Atkinson and Car, 2019). It also differs from RDFS/OWL ontology extensions that use constraints for inference only. Besides, our superontology is intended to be more than a constraint profile, differing from SHACL profiles like ARM SHACL (Art and Rare Materials BIBFRAME profile) (Kovary et al., 2018).

Contrasting with the idea that SCHACL only applies to instance data (Debattista, 2018) and that there is no way to define metadata schema with constraints in a standard language (Coyle, 2019), SHACL is a validation language applicable both to instances and to other RDFS/OWL ontologies, e.g. the DBPedia ontology validation example referred to in Gayo et al. (2018).

Our research goal is to test and demonstrate the possibility of specifying a high-level ontology for the description and validation of bibliographic vocabularies, using SHACL as a standard language and PROF as a formal profile vocabulary.

5. Conclusions

This paper has presented a review of interoperability problems concerning bibliographic ontologies published in the Semantic Web. Issues were identified by evaluating bibliographic ontologies at different levels of interoperability, highlighting especially the semantic level. From the literature analysed it was shown that, despite existing general conformance with the linked data principles and best practices, there are problems difficult to overcome deriving from the underuse of semantic mechanisms, such as little assignment of URIs and URI outlinks, proliferation of vocabularies and poor ontology reuse or ontology "hijacking", among other. Two major problems made evident are the absence of a common framework for integrating and interlinking different ontologies in a consistent manner; and limitations of linked data languages such as RDF and OWL, especially in what concerns the enabling of constraint mechanisms, capable of ensuring the semantic validity and quality of data.

In order to contribute solutions for these various problems and improve the quality and exploitation of bibliographic ontologies and datasets in the open web, our research proposes to define a reference model to frame bibliographic standards and to further specify a superontology capable of relating elements from different ontologies. For the specification of the superontology, the use of SHACL as a standard language and PROF as a formal profile vocabulary will be studied as means to overcome limitations of the RDF family of languages.

Immediate steps include further research in the analysis of existing reference models and ontologies, in order to exemplify the reviewed limitation and demonstrate the effectiveness of a SHACL solution, and the study of the best methodology for analysing data quality in the process of reusing heterogeneous datasets, such as the investigation of metadata quality issues in research data repositories undertaken by Rousidis et al. (2014) and Balatsoukas et al. (2018).

References

All URL valid as of 20 April 2019

Antoniou, G. et al. (2012) *A Semantic Web primer*. 3ª ed. MIT press, Cambridge (MA).

Atkinson, R. and Car, N.J. (Eds.), (2019) *The Profiles Ontology*. W3C Working Draft 2 April 2019, W3C. https://www.w3.org/TR/dx-prof/

Baker, T., Coyle, K. and Petiya, S. (2014) 'Multi-entity models of resource description in the Semantic Web: a comparison of FRBR, RDA and BIBFRAME', *Library Hi Tech*, 32(4), pp. 562-582. doi: 10.1108/LHT-08-2014-0081.

Balatsoukas, P. et al. (2011) 'A survey on the importance of learning object metadata for relevance judgment' in García-Barriocanal, E. et al. (Eds) *Metadata and Semantic Research. MTSR 2011*. Communications in Computer and Information Science, vol 240. Springer, Cham, pp. 300-311. doi: 10.1007/978-3-642-24731-6_31

....

Balatsoukas, P. et al. (2012) 'Learners' perceptions on the importance of learning object metadata for relevance judgement', *International Journal of Metadata, Semantics and Ontologies,* vol 7 (4), pp. 283-294.

Balatsoukas, P., Rousidis, D. and Garoufallou E. (2018) 'A method for examining metadata quality in open research datasets using OAI-PMH and SQL queries: the case of Dublin Core 'Subject'element and suggestions for user-centred metadata annotation design', *International Journal of Metadata, Semantics and Ontologies,* vol 1 (1), pp. 1-8.

Berners-Lee, T. (2006) *Linked Data*. W3C. https://www.w3.org/DesignIssues/LinkedData.html.

Brinkley, J.F., et al. (2006) 'A framework for using reference ontologies as a foundation for the semantic web' in *AMYA Symposium Proceedings 2006*, pp. 96-100. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839690/.

Coyle, K., Silvello, G. and Tammaro, A.M. (2014) 'Comparing methodologies: Linked Open Data and Digital Libraries', in *AIUCD'14 - Proceedings of the Third AIUCD Annual Conference - Humanities and Their Methods in the Digital Ecosystem*. doi: 10.1145/2802612.2802615.

Coyle, K. (2015) 'FRBR, twenty years on', *Cataloging and Classification Quarterly*, 53(3-4), pp. 265-285. doi 10.1080/01639374.2014.943446.

Coyle, K. (2016) *FRBR, before and after: a look at our bibliographic models*. American Library Association, Chicago.

Coyle, K. (2019) 'I'd like to buy a VOWEL', *Coyle's InFormation.* http://kcoyle.blogspot.com/2019/03/id-like-to-buy-vowel.html.

Cordeiro, M.I. (2005*) Information technology frameworks in LIS: exploring IT constructs as sources of conceptual alignment.* PhD thesis, University of London, London.

Debattista, J. et al. (2018) 'Evaluating the Quality of the LOD Cloud: An Empirical Investigation', *Semantic Web,* 9(6), pp. 859-901.http://www.semantic-web-journal.net/system/files/swj1757.pdf.

Doerr, M., Riva, P. and Žumer, M. (2012) 'FRBR entities: identity and identification', *Cataloging & Classification Quarterly,* 50(5-7), pp. 517-541. doi: 10.1080/01639374.2012.681252

Dunsire, G. et al (2012a) 'Linked data vocabulary management: infrastructure, data integration, and interoperability', *Information Standards Quarterly*, 24(2-3), pp. 4-13.

Dunsire, G., Hillman, D. and Phipps, J. (2012b) 'Reconsidering Universal Bibliographic Control in light of the Semantic Web', *Journal of Library Metadata,* 12(2-3), pp. 164-176.

Dutta, B. (2017) 'Examining the interrelatedness between ontologies and Linked Data', *Library High Tech,* 35(2), pp. 312-329.

Escolano Rodriguez, E. (2016) 'RDA and ISBD: history of a relationship', *JLIS.it*, 7(2), pp. 49-81.

Farber, M. et al. (2016) 'Linked Data quality of DBPedia, Freebase. OpenCyc, Wikidata, and YAGO', *Semantic Web,* 1(1-5), pp. 1-45.

Feeney, K.C., Brennan, R. and Gleason, G.M. (2018) 'Linked data schemata: fixing unsound foundations', *Semantic Web,* 9(1), pp. 53-75.

Garoufallou E. and Papatheodorou, Christos (2014) 'A critical introduction to metadata for e-science and e-research', *International Journal of Metadata, Semantics and Ontologies,* vol 9 (1), pp. 1-8.

Garoufallou E., Sartori F., Siatri R. and Zervas M. (Eds) (2019) *Metadata and Semantic Research. MTSR 2018.* Communications in Computer and Information Science, vol 846. Springer, Cham. doi: 10.1007/978-3-030-14401-2

Gayo, J.E.L. et al. (2018) *Validating RDF Data.* Morgan & Claypool, San Rafael, CA.

Godby, C.J., Wang, S. and Mixter, J.K. (2015) *Library linked data in the cloud: OCLC's experiments with new models of resource description.* Morgan & Claypool, San Rafael, CA. doi: 10.2200/S00620ED1V01Y201412WBE012.

Godby, C.J. (2016) 'A division of labor: the role of Schema.org in a Semantic Web model of library resources', in Seikel, J. and Seikel, M. (Eds.), *Linked Data for the cultural heritage*, ALCTS, Chicago, pp.73-101.

Greenberg, J. and Garoufallou, E. (2013) 'Change and a future for metadata' in Garoufallou E. and Greenberg, J. (Eds) *Metadata and Semantic Research. MTSR 2013.* Communications in Computer and Information Science, vol 390. Springer, Cham, pp. 1-5. doi: 10.1007/978-3-319-03437-9_1

Gruber, T. (2009) 'Ontology', in Liu, L. and Oszu, M. (Eds.), *Encyclopedia of Database Systems*. Springer. doi:10.1007/978-0-387-39940-9_1318.

Hallo, M. et al. (2016) 'Current state of linked data in digital libraries', *Journal of Information Science*, 42(2), pp. 117-127. doi:10.1177/0165551515594729.

Hanneman, J. and Kett, J. (2010), 'Linked data for libraries', *World Library and Information Congress: 76th IFLA General Conference*, Gothenburg. https://www.ifla.org/past-wlic/2010/149-hannemann-en.pdf.

Hawtin, R. et al. (2011) *Review of the evidence for the value of the "linked data" approach: final report to JISC*. JISC, Curtis+Cartwright. http://repository.jisc.ac.uk/559/1/JISC_Linked_Data_Review_Oct2011.pdf.

Hebeler, J., et al. (2009). *Semantic Web programming*. Willey, Indianapolis,

Hyland, Bernardette et al. (2014) *Best practices for publishing Linked Data. W3C Working Group Note 9 January 2014*. W3C. https://www.w3.org/TR/ld-bp.

Hogan, A. et al. (2012) 'An empirical survey of Linked Data conformance', *Journal of Web Semantics*, vol. 14, pp. 14-44. https://doi.org/10.1016/j.websem.2012.02.001.

Howarth, L.C. (2012) 'FRBR and linked data: connecting FRBR and linked data'*, Cataloging & Classification Quarterly,* 50(5-7), pp. 763-776. doi: 10.1080/01639374.2012.680835.

Jain, P. et al. (2010a) *Linked data is merely more data.* Association for the Advancement of Artificial Inteligence, Stanford, CA.

Jain, P. et al., (2010b) *Ontology alignment for Linked Open Data*. Shangai, Springer.

Juran, J.M. and Godfrey, A.B. (1998) *Juran's quality handbook.* 5th ed. McGraw-Hill, New York.

Jett, J. et al. (2016) 'Enhancing scholarly use of digital libraries: a comparative survey and review of bibliographic metadata ontologies' in *2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*, pp. 35-44.

Kontokostas, D. et al. (2014) 'Test-driven evaluation of Linked Data quality' in *WWW '14 Proceedings of the 23rd international conference on World wide web*. ACM, New York, pp. 747-757.

Kovari, J. et al. (2018), 'Modeling application profiles in the Art and Rare Materials BIBFRAME Ontology Extension' in *Proceedings International Conference on Dublin Core and Metadata Applications*. http://dcpapers.dublincore.org/pubs/article/download/3979/2181.

Knublauch, H. and Kontokostas, D. (Eds.), (2017) *Shapes Constraint Language (SHACL). W3C Recommendation 20 July 2017*, W3C. https://www.w3.org/TR/shacl/

Knublauch, H. (2017) *SHACL and OWL compared.* https://spinrdf.org/shacl-and-owl.html.

Lóscio, B.F. and Burle, C. (2017) *Data on the Web Best Practices. W3C Recommendation 31 January 2017.* W3C. https://www.w3.org/TR/dwbp/#dataFormats.

Martin, K.E. and Mundle, K. (2014) 'Positioning libraries for a new bibliographic Universe: a review of cataloging and classification literature 2011-12', *Library Resources & Technical Service*, 58(4), pp. 233-249.

Murray, R.J. (2008), 'The FRBR-Theoretic library: the role of conceptual data modeling in cultural heritage information system design' in *iPRES 2008: Proceedings of the Fifth International Conference on Preservation of Digital Objects*. British Library, London, pp. 163–168.

Murray, R.J. and Tillett, B. (2011) 'Cataloging theory in search of graph theory and other ivory towers', *Information Technology and Libraries,* 30(4), pp. 170-184.

NISO (2017) *Issues in vocabulary management: a technical report of the National Information Standards Organization*. NISO TR-06-2017, NISO, Baltimore.

Papadakis, I., Kyprianos, K. and Stefanidakis, M. (2015) 'Linked Data URIs and libraries: the story so far', *D-Lib Magazine,* 21(5/6).http://www.dlib.org/dlib/may15/papadakis/05papadakis.html.

Patrício, H.S., Cordeiro, M.I. and Ramos, P.N. (2019) 'Formalizing Enrichment Mechanisms for Bibliographic Ontologies in the Semantic Web' in Garoufallou E., Sartori F., Siatri R. and Zervas M. (Eds) *Metadata and Semantic Research. MTSR 2018*. Communications in Computer and Information Science, vol 846. Springer, Cham, pp. 147-158. doi: 10.1007/978-3-030-14401-2_14

Peponakis, M. (2012) 'Conceptualizations of the cataloging object: a critique on current perceptions of FRBR Group 1 entities', *Cataloging & Classification Quarterly,* 50(5-7), pp. 587-602.

Peponakis, M. (2016) 'In the name of the name: RDF literals, ER attributes, and the potential to rethink the structures and visualizations of catalogs', *Information Technology and Libraries*, 35(2), pp. 19-38. https://ejournals.bc.edu/index.php/ital/article/view/8749.

Riva, P., Le Boeuf, P. and Žumer, M. (2017) *IFLA Library Reference Model: a conceptual model for bibliographic information*, IFLA, Den Haag. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf.

Riva, P. (2016) 'Il nuovo modello concettualle dell' universo bibliografico: FRBR Library Reference Model', *AIB Studi*, 56(2), pp. 265-275. https://aibstudi.aib.it/article/download/11480/10740.

Rousidis, D. et al. (2014) 'Metadata for Big Data: a preliminary investigation of metadata quality issues in research data repositories', *Information Services & Use*, 34, pp. 279-286. doi: 10.3233/ISU--140746

Schmachtenberg, M., Bizer, C. and Paulheim, H. (2014) 'Adoption of the Linked Data best practices in different topical domains', in Mika P. et al. (eds) *The Semantic Web – ISWC 2014. ISWC 2014.* Lecture Notes in Computer Science, vol 8796. Cham, Springer, pp. 245-260.

Smith-Yoshimura, K. (2018) 'Analysis of 2018 International Linked Data Survey for Implementers', *Code4Lib Journal,* 42. https://journal.code4lib.org/articles/13867.

Sprochi, A. (2016) 'Where are we headed? Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model', *International Information & Library Review*, 48(2), pp. 129-136. doi: 10.1080/10572317.2016.1176455

Suominen, O. and Hyvonen, N. (2017) 'From MARC silos to Linked Data silos?', *o-bib. Das offene Bibliotheksjournal*, 4(2). https://www.o-bib.de/article/view/2017H2S1-13/5883.

Svensson, L.G. (2013) 'Are current bibliographic models suitable for integration with the Web?', *Information Standards Quarterly*, 25(4), pp. 7-13. https://groups.niso.org/apps/group_public/download.php/11931/FE_Svensson_bibliographic_models_isqv25no4.pdf.

Szeto, K. (2013) 'Positioning library data for the Semantic Web: recent developments in resource description', *Journal of Web Librarianship*, 7(3), pp. 305-321. doi: 10.1080/19322909.2013.802584.

Talleras, K. (2018) *Metadata structures of the Bibliographic Universe: transformation. interoperability, conceptualizations and quality.* PhD thesis, Oslo Metropolitan University, Oslo. https://oda-hioa.archive.knowledgearc.net/handle/10642/6622.

Vrandecic, D. (2010) *Ontology evaluation.* PhD thesis, Karlsruher Instituts für Technologie, Karlsruhe, Germany. http://simia.net/download/ontology_evaluation.pdf.

Willer, M. and Dunsire, G. (2013) *Bibliographic information organization in the Semantic Web*. Chandos, Oxford.

Yee, M.M. (2009) 'Can bibliographic data be put directly onto the Semantic Web? *Information Technology and Libraries,* 28(2), pp. 55-80. https://ejournals.bc.edu/index.php/ital/article/view/3175.

Zapounidou, S., Sfakakis, M. and Papatheodorou, C. (2016) 'Representing and integrating bibliographic information into the Semantic Web: a comparison of four conceptual models',

....

Journal of Information Science, 43(4), pp.525-553.
https://doi.org/10.1177/0165551516650410.

Zaveri, A. et al. (2012) *Quality assessment methodologies for Linked Open Data: a systematic literature review and conceptual framework.* SWJ, IOS Press.
https://pdfs.semanticscholar.org/916a/ef4b7cb2e650628ef48b364850d209aff041.pdf.