



ISBD adaptation to SW of bibliographic data in linked data

Elena Escolano Rodríguez

Introduction

The achievement of any business model depends on the users' belief, the concept is known as "true economy";¹ belief in truth and in the quality of data will be the best investment in the future of ubiquity. Consequently, the emphasis will be on the quality of this information. Indeed it is very important it is controlled and truthful. In this context the source of information is a relevant value, given that it is the source that lends authority to data. If expressed as linked data, the information which has at length been selected, controlled, validated, recorded and structured in cultural institutions as libraries' databases, will certainly take on an important role. In this way, libraries, museums, and archives can retake a prominent position on the web; they have traditionally selected, structured and organized information and have at the same time contributed to convert information into quality data. Nowadays, the need to provide access to both integrated heterogeneous knowledge and distributed homogeneous knowledge in several domains is considerably grow-

¹<http://www.chiefmartec.com/2010/03/business-models-for-linked-data-and-web-30.html>.



ing. The objective is to use the web as a single global database, so the linked data tool will allow for this global web. It is necessary to recognize the ever increasing importance attached to the discovery of objects both digital and not. Therefore, the data that describes the objects should be available there where users are, integrated in a global web which means they should be open and as a result reusable. In the last years the semantic web cloud has increased in a significant way. This is confirmed by the growth occurred from 2007 to 2010 and up to September 2011;² the rise becomes evident observing the graphs in the green part, concerning the publications domain. This phenomenon underlines the importance for library community to have their structured and controlled data available on the web with this new tool. Consequently, for libraries this will represent:³

- the extensibility and the integration of their data with those of other institutions, with the consequent increase of information that users can retrieve; in this way it would be possible to complete, aggregate and link the library data with other structured information in different ways, in accordance with other standards. The integration can guarantee a greater and better service to the user, not only by virtue of the involvement of museums and archives but thanks to the collaboration with new centres and institutions, or with other products and information sources; this process would allow the widespread diffusion of information recorded by libraries and the integration with other web segments, such as Wikipedia, Geonames, etc.;

²http://richard.cyganiak.de/2007/10/lod/lod-datasets_2011-09-19_colored.html.

³Linked Data Incubator Group wiki: http://www.w3.org/2005/Incubator/lld/wiki/Draft_Benefits and <http://www.w3.org/2005/Incubator/lld/wiki/Benefits>.

- the compliance with requests for public administration transparency; the libraries of public institutions could take charge, upfront, of the planning and coordination of this political action, avoiding duplications of projects within the same sector which should be linked to each other. In this process the libraries could take on a real crucial role;
- the semantic understanding of library language, so far quite unclear for the users, can contribute to optimise the results obtained during the search and consequently the whole library services;
- the possibility of presenting the search results in the user's language. In the future, it will allow, or at least make it easier, to define large cooperation areas, thanks to the automatic conversion in the linguistic form accepted by a specific community. These applications will also concern and improve the multilingual cooperative cataloguing, that is the records could be created and presented in one language without the necessity of creating a new "record".

IFLA contribute

The work carried out by the IFLA ISBD Review Group has also been carried bearing in mind this scope: making available in the cloud the information stored in our database, or as explained in the consolidate ISBD edition at paragraph "A.1.2 Scope": "improving the portability of bibliographic data in the semantic web and consequently the interoperability of the ISBD standard in connection with other content standards (International Federation of Library Associations and Institutions. ISBD Review Group and International Federation of Library Associations and Institutions. Cataloguing

Section. Standing Committee p. 1). Linked data is necessary for participating in the web of data, but for taking part in the semantic web, putting data on the web and link them is not enough: there are other necessary requirements which, according to Berners-Lee (“Linked Data - Design Issues”), are:

1. using URI for identifying or referring to sources. The URI (Uniform Resource Identifier) is the characters set used to indicate univocally the names of the resources on the web and are expressed in a machine-readable form;
2. using HTTP URIs, so that the user can look for and locate resources through them (this is called dereferencing)
3. providing useful information about the resource when we search it with URI, using standards (for example RDF, SPARQL);
4. including links with other URIs for finding out linked information.

The studies on the semantic web are specifically focused on formal ontologies, that is, the logical structure in which the semantic of a particular domain is organized. Aiming at integrating and managing the knowledge of this dispersed information, the research has also contributed to facilitate relationships between ontologies, specifying their context clarify how widespread knowledge is related to several resources. Some information can be automatically captured and information related to the source can explain the context. So it can be reduced the presence of non-intentioned or unwanted meaning in the ontology, obtaining a greater clearness and facilitating the analysis and the search. In order the libraries participate in the semantic web, it was necessary to create the ontology that reflect the logical structure of the library domain, providing useful information to make it understandable. Particularly, in the library field there

was much work done on definition of a common well structured and standardized basis, represented by IFLA very much consolidated standards, which ensure quality, exchange capability and sustainability. This regulation encourages the right development of the semantic web, because standards are important components for linked data. IFLA's contribution and participation to this process is justified not only because the utility and the importance it has for libraries but also for additional reasons: it was considered essential to protect the own terminology, at the same time specifying the context and the origin of the metadata (that is a very important issue in linked data). In addition, for ISBD it was a main objective to reposition the IFLA standard and its value as important tool for the delivery and reuse of structured authorized bibliographic data in the Internet environment. IFLA has carried out several actions. First of all, it decided to declare its own models and standards in the Resource Description Framework (RDF). It was followed the recommendation, by the advisor Gordon Dunsire, to the FRBR Review Group in 2008 , and to the ISBD Review Group in 2009 . Its application was decided and authorized during that meeting of the ISBD Review Group, at the IFLA Conference held in Milan in 2009. Work started on the declaration of ISBD set of elements in RDF, in order to present and be submitted to IFLA Cataloguing Section for approval as part of the ISBD consolidated edition of 2011. For such purpose, it was necessary to create a namespace that would properly identify the URIs of RDF declarations by IFLA for its own models and standards, what was recommended in the 2008 report. The recommendation consisted in protecting elements, terms and definitions related to the IFLA models and standards, using a sort of brand, to save them from unlike interpretations from other standards. This action helped also to achieve what has been mentioned before related to the business model, in which basis to obtain results

or indirect benefits, it is necessary this quality brand. To carry out this work, study and elaboration a Namespace Group, coordinated since 2009 by Gordon Dunsire, was created within IFLA with the objective of management of the IFLA standards declarations. When establishing the namespace, several issues were taken into account. The namespace had to be clear, short, expanded and applicable to each model and standard. It was decided to adopt the URL form (which begins with `http://...`), that in the future may be dereferenced, in order to retrieve the RDF or the HTML file when the URI is processed as an ordinary URL. Once the decision was made, the focus was on the namespace structure: it was decided to adopt this quality mark: `http://iflstandards.info`, considering the potentiality of URL to be intelligible both for computers and humans. Following, the abbreviations of the standard referred would be identified, such as for instance: `http://iflstandards.info/ns/isbd/elements`. Then it should be considered how to identify the element in the URI. URIs can contain letters and numbers. It could be useful to remember that URI is specifically defined for machine understanding, it is not a label intended for the user even if it can guide him. In fact, the context of an element could be briefly identified with one word but at risk of misleading him to believe this textual information is similar to a label: the label in itself is not sufficient, the programmer, the human being, has to read the full declaration corresponding to the URI with its definition for the correct application. Due to these reasons debated at IFLA General Conference held in Gothenburg in 2010, after which it was decided that URIs would be opaque, without reference to a specific language, because IFLA has to recognize and encourage the multilingualism; therefore, in order to guarantee linguistic neutrality, a numerical solution should be adopted. An opaque URI would also extend its use to linguistic communities different from the English ones ensuring, at the same time, access to

these ontologies in other languages without the necessity of creating independent URIs. The declarations contain important information such as metadata name, label, definition, notes used for extending the information or its application, the filiation (whether it is property or sub-property), the state of acceptance, etc. The utility of translation affects definitions, notes and also the labels. Using an opaque URI and specifying the language in which you desire to obtain the information, it is possible to collect all declarations in different languages with the same URI. If an opaque URI had not been used, it would have been necessary to create one for each language to be afterwards linked to the others as "same as". The problems related to translation will be further developed below. The labels refer more to the comprehension of the programmer than to the machine; it was necessary to disambiguate and adapt them because the relations present in FRBR are coincident for several entities (in RDF classes) so it was necessary to specify the domain of the relationship; and also in some cases it was not clear the relationship orientation (the range in RDF). For example in Italian: "ha come forma variante" is a relationship (property or subproperty for RDF) which can be applied both to the entity/class "Person" and "Corporate Body". Therefore it was necessary to add information in brackets, to identify more specifically the classes which the properties belong and the direction of the relationship. Both FRBR models and ISBD standard include controlled vocabularies. In the former case it consists of the user's tasks while in the ISBD correspond to the terms used for Area 0: : Content form and media type. Vocabularies were identified by completing the URI with the expression "terms" <http://iflastandards.info/ns/isbd/terms/> and the notation which indicates the concrete term of the normalized vocabulary. It stands evident that the vocabulary, which is recorded in the language of the cataloguing agency, would be simply converted into another

language when the record or the information is captured by other agency, especially in a cooperative environment, in a controlled, normalized and automatic manner. In this way, it is also possible to map or create correspondences with similar vocabularies but not structured in the same way; as for example with the Resource Description & Access (RDA) vocabulary used to describe content and support: Content Type and Media Type. However this is not possible without problems for establishing an equivalence relationship of "same as" type, because, in some cases, there is not a total correspondence 1=1. Until now these declarations have been made in Open Metadata Registry, which a space created by the W3C is containing several ontologies about different domains, but in the future it will be possible to transfer these declaration to a specific section in the IFLA website where they can be hosted and managed. Regarding the sustainability and maintenance of the IFLA Namespace is sill and issue on course.

Multilingualism development

The basis for the semantic web is basically in English, which has worrying consequences about cultural and linguistic diversity. Even if English is recognized a IFLA working language, there are also other six official languages that require the development of multilingualism. The first issue of the ISBD/XML Working Group plan, approved in November 2011 , states the intention of promoting the translation of ISBD and the declarations in OMR, in addition of the definition of guidelines for translators. From my participation, on several occasions, in debates concerning the translation of IFLA declarations, I am going to highlight some issues that affect many Latin languages such as Spanish, Italian, Croatian, Slovenian etc. The significant topics discussed are the following:

Style issues

As far as labels are concerned, there is a good practice, drawn from some communities of the semantic web, to use capital letters for classes names in RDF. Moreover, in English, words are joined together, what is called CamelCase, for instance the ISBD subclass: `ParallelTitleCompoundEncodingScheme`, but this is not possible to apply to Spanish. In some cases the use of capital letters could be accepted, even for prepositions while the conjunction without space is not accepted. Therefore, in Spanish it was accepted the use of capital letters for the first letter of the first word or for every word, but without joining the words. Another issue regards the property labels: they are always verbal phrases. In fact their aim is to serve as predicate in the RDF triple RDF: Subject – Predicate – Object. With respect to Latin alphabets according to the best practices used for the semantic web community, it is recommended to write in lower case. From the beginning it was adopted the convention to avoid, as far as possible, to use the indefinite articles, when possible, with the aim of normalizing and reducing the length of the labels. Likewise, when having to choose between the singular and plural, it was preferred to use the singular, whenever possible. These decisions were also applied during the creation of the ISBD set of elements, since the standards were being revised at that time.

Sources of reference

In the RDF data model, the source of reference, the text of the standard, is essential for programmers and developers; indeed, they could use and consult it as an additional aid to make a better semantic contextualization of the property. From the start, a decision taken by the FRBR Review Group was that labels, definitions and scope notes of the RDF's framework would be kept aligned and

matched, as possible, with the text of FRBR; this would have the advantage of allowing natural language processing. In particular, labels would have match with the text accepted in the standard; concerning definitions, their alignment with the text is important, even if few modifications are required to adjust it to the context and making it understandable, that is, they will be as extracts; in the case of scope notes more flexibility in the compiling is also allowed. At the beginning, as there was no experience on which to base our work, the way we decided to follow for translating into Spanish these dispositions (labels, definitions, notes), was inevitably their literal translation from the English version. This decision presented lot of problems for the comprehension of the text in the language of translation, Spanish, and, at the same time, for the respect to the official standard text. In the case of labels, for example, it was necessary to add prepositions to help interpreting the properties, which were not in the English version of the FRBR report. We continued to work in this way for a while. However, after facing many situations, partially already cited, which were useful as forced the group to reconsider certain issues and decisions. As, for example, it was considered that the Spanish developers would have preferred to use the official Spanish translation of the FRBR report as reference source and, therefore, the declarations would have to align with the official Spanish version, instead of the English one. That revoked the initial decision to base the Spanish translation of the RDF declarations on that available in English, always avoiding the semantic ambiguity. Two solutions arised: If available, we would have to use the official translation of the reference source for the declaration in RDF, and if not present, it would be necessary to base the translations on the English declarations, concerning labels, definitions and scope notes. Obviously, if translations of reference sources (standard and models) are not updated represent other serious problem. Even

for this case, it was thought various possibilities: full translation of labels, definitions and scope notes (that in the Spanish panorama had already been made for the elements of ISBD); or, more simply, translation of labels (matching with the state of art of the translations in Spanish of the model of the FRBR family). During the meeting of the ISBD/XML Study Group, November 4, 2011 in Edinburgh (UK), Gordon Dunsire stressed the ambiguity of the term "statement" in the OMR as meaning "aggregated elements", whereas in ISBD the meaning for such term is "the information from the source" (eg., "1.4.5.10 parallel title and parallel statement of responsibility"). It has been recognized as necessary and urgent that the ISBD Review Group revise the current labels and definitions in OMR, and that the ISBD/XML Review Group provides a report on the possible need of change from the work on the ISBD application profile that is being prepared.

Qualifications

As previously mentioned, in the declarations concerning properties, especially in the case of the FRBR models relationships, it was necessary to use parentheses for the disambiguation, as there is homonymy depending on whether it applies the relation to a kind of entity or to another. As the name of the relation is the same even if it is applicable to different entities, it was necessary to do a disambiguation by adding consecutively, in brackets, the main class of the property (that is the domain); the necessity to disambiguate the second term of the relation has required the adoption of further brackets for the orientation of the relation (the range). The activity of translation was also useful as revision for these qualifications; it appeared that the use of parentheses to other languages than English was not clear nor systematic.

Vocabularies

OMR contains also the declarations of the model controlled vocabulary; in the case of ISBD, the vocabulary consist on the terms used in the Area 0 of the description Content form and Media type). Vocabularies are designed for their display to the user. It has been used Simplified Knowledge Organisation System (SKOS) for their representation. With regard to translations, we face the following problem: the grammatical flexion for masculine and feminine adjectives in Spanish and in other languages. Specially in the case of ISBD qualifiers for the content forms of sense and type of content - mainly the latter - that, in Spanish and other languages, have the same gender declination of the names from which they depend; that is, different declination when it is masculine or feminine adjective to be used. Thus we have:

- Imagen (cartográfica) and Objeto (cartográfico);
- Música (notada) and Movimiento (notado).

During the process of translating the vocabularies of ISBD Area 0 we, Spanish, opted for the compound form with slash, that is, "cartográfico/a", that could be used in this way, although if not matched with natural language, or by giving the option to libraries to use the most convenient type. This is the model that languages similar to Spanish have followed in their translation of the Area 0, but elsewhere, for example in the publication of ISBD's examples, we used the simple form. This solution is not applicable, however, for the disposition in SKOS, that provides the preferred label (prefLabel) and does not allow more than one prefLabel per language. The preferred label is the one we expect will be used for the friendly display and that contains the semantics. The case is still under study and debated within the Namespace Task Group. For the moment, the Grupo de Ingeniería Ontológica of the Universidad Politécnica de Madrid proposed a

possible solution with the ontology of LIR / Lemon. It will take time to explore and implement it. Therefore, at the moment, and to avoid the compound form that would not coincide with the natural language, we have declared the two forms of masculine and feminine as SKOS "alternate labels", excluding a preferred label. The solution seems to be consistent with SKOS, problems will arise during the applications that require the preferred label for display purposes. As a label in the SKOS model cannot be alternative and preferred at the same time, in the future it will be necessary to eliminate one or both. The subject is under study and affects many languages. There are rules that have different vocabularies from those accepted by IFLA, with this tool will be easier to map (or find matching) with those vocabularies, and simultaneously link with other languages. The recording in OMR allows the status of publishing at different levels, and each component of the triple can have its own status. There are no fixed rules for the status of the record, but the general use is that definitions should remain in the same way when the status is "published" and labels and scope notes can be changed.

Recent actions

It is not possible to say that all has been said in advance influenced the changes that can be perceived in the new consolidated ISBD edition, but indeed they have certainly a prominent impact on several ones and have also motivated some decisions. IFLA has always had the objective of updating standards to the technological innovation with the aim to support all different kind of libraries improvement, but without forgetting the cumulated experience over the year and the different status and resources libraries could have, so it is fundamental that IFLA developments are useful for any kind of library, ensuring the scalability. The approach from ISBD view comes from

the recognition that the human judgment and logic are essential as criteria to select the value data and record it in the description of the resource. This does not mean undervaluing or diminishing the technology that permits the automatic collection of data, but the rules have to be created with the goal of ensuring the quality; and the tools used could change according to the moment, the situations and the cases. On 31st January 2011 the IFLA cataloguing section approved the new consolidate edition published by De Gruyter Saur in July. The changes that could be noticed include those concerning primarily the review of a standard: variations of editorial style and changes in the introduction for better orienting its application, revision or addition of new definitions for removing ambiguity and other examples are included. Once the Review Group clarified what "data element" means, it was possible to modify the standard making it less repetitive, more consistent, easy and logic to apply. The main modifications in the final edition are due to reasons above mentioned, that have compelled to a careful analysis of the ISBD elements; to the search of a better consistence and quality of data, as well as the interoperability among these "data elements" and those coming from other standards. During the ISBD group meeting held in November 2011 emerged the following considerations:

- DC Application Profile is still under development;
- the term "statement" is ambiguous: in OMR the meaning is "aggregated elements", while in ISBD "the information from the source" ;
- collaboration with JSC on the development of a representation of the RDA/Onix Framework in RDF;
- mapping between ISBD Area 0 and the RDA/Onix Framework;

- liaison with DCMI and with appropriate groups on translation issues;
- mappings between the ISBD and UNIMARC namespaces.

As said before, standards are important references for linked data and for semantic web. During the development of the new RDA cataloguing standard, the ISBD review group suggested that it would be useful a meeting with JSC to discuss further the differences and the similarities between them, but it was after its publication in 2010, that it was possible the meeting. ISSN Network was also interested in taking part in the meeting and renovate a tripartite agreement that was achieved in 2000 . This is coincident with another purpose of the ISBD Review Group regarding "the necessity to continue activities on harmonization of the ISBD, ISSN, RDA and other national and international cataloguing rules aligned with the FRBR model and the International Cataloguing Principles" and this tripartite meeting will be the first step toward the harmonization. The meeting addressed the topics identified as potential difficulties to the realization of interoperability among ISBD, ISSN and RDA. During the meeting to homogenize ISBD-ISSN-RDA, (Glasgow 3rd-4th November, 2011) many common issues were dealt concerning RDA rules, among them:

- sources of information;
- elements which have the same name but different definitions;
- criteria for the order of selection of the information;
- different interpretation of the data nature.

In this context, the harmonization has meant functional interoperability, so that records created according to at least one of these rules

would be reusable by an agency employing one of the other standards. Both ISBD Review Group and JSC agreed with the creation of correspondences between set elements of the two standards in RDF. As regards the diverse vocabulary which both standards recommend for the content form and media type, it was considered that a strict mapping between ISBD and RDA was not possible. Both standards are based on RDA/ONIX Framework for Resource Categorization document, which contains the general categorization used by other community's standards. Therefore during the meeting in Glasgow it was decided to declare this categorization in RDF and to make the mapping from each standard to RDA/ONIX categorization. The definitions of the ISBD and RDA elements will be reviewed to determine if they are semantically equivalent or if they are subproperty to each other. Indeed it was decided to update the Appendix A of RDA by the ISBD Review Group and to include the existing mappings and guidelines will be developed in an ISBD application profile for RDA. Meanwhile the DCMI/RDA Task Group has become Bibliographic Metadata Task Group⁴ in which ISBD Review Groups is represented. Its aim regards the definition of components of current and emerging library, publishing, and related bibliographic metadata standards as RDF vocabularies for use in developing Dublin Core application profiles and semantic mappings. Consequently, its tasks are: Explore "obvious" mappings between known element sets and between value vocabularies, and identify issues, solutions, etc.; Give feedback on the draft ISBD-RDA/ONIX alignment and methodology; Give feedback on the draft IFLA guidelines on translations of namespaces. It is necessary to wait for the report of the meeting held in London 26th April 2012, in which ISBD had much attention.

⁴http://wiki.dublincore.org/index.php/Bibliographic_Metadata_Task_Group.

Conclusions

In conclusion, the importance of the data is growing considerably and its validity and authority is becoming fundamental, this is mainly due to the relevance of its reuse considering that if it was not true it would generate and increase mistakes, creating erroneous relations. We hope that the purpose and means for the dissemination of IFLA standards, through the namespace, and the importance of considering IFLA standard as reference model, will have the expected reception. At the present moment the IFLA standards are recognized as semantic web standards for bibliographic metadata.⁵ In addition, many on-going projects are using the IFLA URIs, such as the Universitätsbibliothek of Mannheim⁶ which is adopting the ISBD URI and the British Library.⁷ Also the Deutsche Nationalbibliothek which is using URIs for the representation of FRBR entities in RDA, is to replace them with the official version of the IFLA URIs,⁸ while the Biblioteca Nacional de España that has presented in December 14th 2011 its project of catalogue published in linked data, using the IFLA ontologies: FRBR, FRAD, and ISBD. Obviously, that is not the end of the matter. In the article it has been seen the efforts for adapting the standards to the new environment, the semantic web, and also the efforts to facilitate its comprehension, so that the information which come from different libraries that apply various standards can be linked and interoperable. It is making headway toward the semantic web and the link among libraries, their languages, formats, etc., but aren't we forgetting the opposite part of the cloud not directly related to our competence, which is not structured ac-

⁵http://www.w3.org/2005/Incubator/ld/wiki/Library_Data_Resources.

⁶http://data.bib.uni-mannheim.de/dokumentation_en.html.

⁷<http://www.bl.uk/bibliographic/datasamples.html>.

⁸http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkeddata_node.html.

ording to our standards. Many people have said that the librarian language is obscure to users, at this respect linked data could make easier its understanding and guarantee the communication with others languages. IFLA has contributed for favoring this process, now it is necessary that the developers adapt systems to this new technology. our bibliographic universe.

References

- Berners Lee, Tim. "Linked Data - Design Issues". (2006). <<http://www.w3.org/DesignIssues/LinkedData.html>>. (Cit. on p. 122).
- International Federation of Library Associations and Institutions. ISBD Review Group and International Federation of Library Associations and Institutions. Cataloguing Section. Standing Committee. *ISBD International Standard Bibliographic Description*. Berlin: De Gruyter Saur, 2011. (Cit. on p. 121).

ELENA ESCOLANO RODRÌGUEZ, Biblioteca Nacional de España.
elena.escolano@bne.es

Escolano Rodríguez, E. "ISBD adaptation to SW of bibliographic data in linked data". *JLIS.it*. Vol. 4, n. 1 (Gennaio/January 2013): Art: #6305. DOI: [10.4403/jlis.it-6305](https://doi.org/10.4403/jlis.it-6305). Web.

ABSTRACT: Linked data is the current paradigm. All works, projects and applications have it as main tool, because of its potentiality. In linking data, the selected information, authorized, validated, recorded and structured in their databases for long time by cultural institutions as libraries, necessarily is going to play an important role. The work carried out by IFLA ISBD Review Group has had this goal in mind, to make possible that the information in our databases will be in the cloud, that is, "enhance the portability of bibliographic data in the semantic web environment and the interoperability of the ISBD with other content standards" as said in the Purpose of Consolidated ISBD, 2010. Many voices have spoken about the obscure language for users that represent the library language. Linked data can help in making it understandable. To reach this big objective much more, work than converting information in linked data is necessary. Declarations in RDF, definitions, and translations are essential to make really multilingual understanding, not only English semantic web.

KEYWORDS: Namespace; ISBD/XML; FRBR; Semantic web; Library Linked Data Project

ACKNOWLEDGMENT: Paper based on the Lectio Magistralis in Biblioteconomia, held at the University of Firenze, on March 6th, 2012 within the Master biennale in catalogazione, directed by prof. Mauro Guerrini. I would like to thank Claudio Venturato for the Italian translation of the paper.

Submitted: 2012-04-25
Accepted: 2012-08-31
Published: 2013-01-15

