The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-24129-6_13

Assumpção, F.S., Santarem Segundo, J.E., Santos, P.L.V.A.C.: RDA Element Sets and RDA Value Vocabularies: Vocabularies for Resource Description in the Semantic Web. In: Garoufallou, E., Hartley, R.J., Gaitanou, P. (eds.) MTSR 2015. CCIS, vol. 544, pp 147-158. Springer (2015)

RDA Element Sets and RDA Value Vocabularies: Vocabularies for Resource Description in the Semantic Web

Fabrício Silva Assumpção[1], José Eduardo Santarem Segundo[1,2], and Plácida Leopoldina Ventura Amorim da Costa Santos[1]

[1] UNESP – Univ Estadual Paulista, Graduate Program in Information Science, Marília, Brazil
{fsassumpcao,placida}@marilia.unesp.br
[2] Universidade de São Paulo (USP), Department of Education, Information and Communication, Ribeirão Preto, Brazil
santarem@usp.br

Abstract. Considering the need for metadata standards suitable for the Semantic Web, this paper describes the RDA Element Sets and the RDA Value Vocabularies that were created from attributes and relationships defined in Resource Description and Access (RDA). First, we present the vocabularies included in RDA Element Sets: the vocabularies of classes, of properties and of properties unconstrained by FRBR entities; and then we present the RDA Value Vocabularies, which are under development. As a conclusion, we highlight that these vocabularies can be used to meet the needs of different contexts due to the unconstrained properties and to the independence of the vocabularies of properties from the vocabularies of values and vice versa.

Keywords: Resource Description and Access (RDA). Resource Description Framework (RDF). Functional Requirements for Bibliographic Records (FRBR). Vocabularies.

1 Introduction

In Information Science, the representation of resources has been based on several instruments, including metadata standards, which are created for specific contexts and focused on specific technological environments. With the Semantic Web initiative, there is an attempt to develop and implement technologies that allow the creation of descriptions of resources accessible and processable not only by its syntax, but also by its semantics.

In this sense, Information Science needs metadata standards suitable for the Semantic Web, that is, metadata standards appropriated for the creation of representations accessible by applications that use Semantic Web technologies. Based on this need, some initiatives arise in descriptive cataloging in order to create suitable metadata standards and/or to adapt those already existing.

One of these initiatives occurs in parallel with the development and implementation of Resource Description and Access (RDA) [17] and it has as its main goal the creation of vocabularies of properties and vocabularies of values based on RDA. The results of this initiative have been released under the names RDA Element Sets and RDA Value Vocabularies.

Considering the contributions of this initiative, in this paper we aim to present the RDA Element Sets and the RDA Value Vocabularies, describing their development, classes, properties and values.

2 RDF Data Model in Resource Description

Descriptive cataloging deals with the description of formal aspects of information resources and establishes names and titles to provide access to these resources. In order to do it, descriptive cataloging comprises instruments for description that were created over the course of time; some of these instruments are the metadata standards and the content standards.

Over the past decades, we have faced changes in descriptive cataloging as a result of the development of information and communication technologies. Such changes require different views of the treatment of information resources and the use of practices for information organizations on the Web [2].

One of these changes involves the traditional approach that has defined catalogs' structures since the 19th century, when cataloging practices and instruments began to be formalized. The traditional catalog record, which "is composed of the values of multiple properties associated with a bibliographic entity, for example its title and physical description" [7], results from this approach.

Revision of this approach is necessary because in the Resource Description Framework (RDF) data model [16] – one of the base technologies for the Semantic Web – "the focus is on individual metadata statements represented by three-part data triples in the form subject-predicate-object" [6]. RDF is a data model that allows us to describe any kind of resources using triples composed by subject-predicate-object or, as we prefer to use in this paper, resource-property-value [2] [8].

With the focus changing, we will have individual statements, each one describing a property of the resource; for example, "The book has the title The Lord of the Rings", "The book was published in 2005" and "The book was written by J. R. R. Tolkien", rather than a single record with all the properties together, as we can see in MARC 21 Format for Bibliographic Data. "The RDF approach is very different from the traditional library catalog record exemplified by MARC21, where descriptions of multiple aspects of a resource are bound together by a specific syntax of tags, indicators, and subfields as a single identifiable stream of data that is manipulated as a whole. In RDF, the data must be separated out into single statements that can then be processed independently from one another; processing includes the aggregation of statements into a record-based view, but is not confined to any specific record schema or source for the data. Statements or triples can be mixed and matched from many different sources to form many different kinds of user-friendly displays." [6]

We can see the record-based approach not only in catalogs' structures but also in metadata and content standards used in descriptive cataloging, such as the MARC 21 Format for Bibliographic Data and Anglo-American Cataloguing Rules (AACR2), respectively. Thus, as the approach changes, we will need metadata and content standards suitable for RDF data model.

Considering this necessity, some initiatives have been undertaken for the following purposes: (1) creating metadata standards suitable for RDF, for instance, the Bibliographic Framework Initiative (BIBFRAME) and the vocabularies created from FRBR, FRAD, FRSAD, ISBD, and RDA, and (2) adapting the standards already existent such that they can be used in RDF, for instance, MODS RDF Ontology, MADS/RDF, and the vocabularies created from MARC 21 and UNIMARC.

These initiatives apply the concepts of "vocabulary", "vocabularies of properties", and "vocabularies of values". To provide a better understanding of RDA Element Sets and RDA Value Vocabularies, in the next section we briefly discuss such concepts and their relationships to descriptive cataloging.

3 Vocabularies

In RDF descriptions, each statement is composed of a resource, a property and a value; this later may be literally described (a literal value) or ascribed to another resource [8]. Following the previous examples, the resource is "the book", the properties are "has the title", "was published in" and "was written by", and the values are "The Lord of the Rings" and "2005" (literal values) and "J. R. R. Tolkien" (a resource). In this case, "J. R. R. Tolkien" is considered a resource because we may continue to describe it; for example, "J. R. R. Tolkien was born in 1892".

The properties should be from vocabularies and the values may be taken from the vocabularies. By vocabulary, we mean a set of terms. So, a set of terms used as properties is a vocabulary of properties and a set of terms used as values is a vocabulary of values.

These kinds of vocabularies are familiar to those within Information Science. Vocabularies of properties may be understood as metadata standards [8]: predetermined sets of methodologically constructed and standardized metadata (descriptive elements or attributes that represent characteristics of a resource or that are assigned to it) [1]. Dublin Core, for example, is a vocabulary of properties because it provides a set of properties (terms) to describe resources.

Vocabularies of values are similar to subject headings lists and authority files; these well-known instruments of Information Science provide sets of terms (topical and chronological terms, personal, corporate and geographic names, etc.) to be used as values. In addition, there are vocabularies used to represent languages, countries, document and content types, etc. Some examples of these vocabularies are the lists of languages and country codes used in MAR 21 Formats, the list of terms used as general material designation in AACR2r, and the lists of codes for illustrations, target audience, form of item, nature of content and literary form in 008 field of MARC 21 Format for Bibliographic Data.

In RDF statements, resources and properties should be identified by Uniform Resource Identifiers (URIs), while values should be identified by a URI only if they are resources, that is, when they are not literal values [8]. So, vocabularies, in the semantic Web context, define URIs for their properties and values.

4 Vocabularies Created from RDA

RDA was published in 2010 as result of the AACR foundation's revision that began in 1997. "RDA essentially standardizes how metadata content is identified, transcribed and generally structured, although it is independent of any specific metadata encoding. RDA also identifies a general set of metadata elements, and in many cases provides a controlled vocabulary for use as the content of an element." [9]

One of the bases of the RDA is the conceptual model Functional Requirements for Bibliographic Records (FRBR) that, expanded by Functional Requirements for Authority Data (FRAD), provides a set of entities, attributes and relationships for RDA.

In a meeting held in 2007, representatives of RDA developers and the Dublin Core Metadata Initiative (DCMI) recommended some activities in order to create a "metadata standard that is compatible with the Web Architecture and that is fully interoperable with other Semantic Web initiatives" [4]. The activities recommended were: "development of an RDA Element Vocabulary; development of an RDA DC Application Profile based on FRBR and FRAD; and disclosure of RDA Value Vocabularies using RDF/RDFS/SKOS." [4]

Starting from these recommendations, the DCMI RDA Task Group was established in 2007 with the objective "To define components of the draft standard RDA - Resource Description and Access as an RDF vocabulary for use in developing a Dublin Core application profile" [5].

In 2011, based on a review of its goals and activities, the group changed its name to the Bibliographic Metadata Task Group. In January 2014, the vocabularies of properties created from attributes and relationships defined in RDA were published on the Open Metadata Registry under the name RDA Element Sets and, in June of the same year, they were also released on the RDA Registry platform [13,14,15]. Some vocabularies of values are already published in RDA Registry under the name RDA Value Vocabularies, while some remain under development in the Open Metadata Registry [12].

The platforms used for publishing these vocabularies, Open Metadata Registry [15] and RDA Registry [10], provide "information about the metadata standard in a machine-actionable format, capable of integration into applications" [9]. In these platforms, the statements about classes, properties and values are available for humans (in a HTML interface) and for machines (in Turtle, Notation 3, N-Triples, RDF/XML, RDFa, Microdata, JSON-LD and RDF/JSON) [15].

5 RDA Element Sets

RDA Element Sets comprise seven vocabularies: one with classes, five with properties used for each class, and one with properties unconstrained by FRBR classes.

5.1 Vocabularies of Classes

In the Semantic Web, a class is defined as a set of individuals or even "an abstraction mechanism for grouping resources with similar characteristics" [11]. "A class is much like a class in a scientific taxonomic sense: it is a grouping of like resources that all belong together based on some common characteristics that make them members of the same set." [3]

We recognize the concept of class used in the Semantic Web as being similar to the concept of entity used in FRBR and RDA. For instance, intellectual creations may be gathered into a class "Work" and its creators may be gathered into classes like "Person", "Family" and "Corporate body". Based on this similarity, a vocabulary of classes from RDA entities was defined. This vocabulary comprises the following terms: Work, Expression, Manifestation, Item, Agent, Person, Family, and Corporate Body.

The classes corresponding to the entities person, family and corporate body are subclasses of the Agent class, since these entities share attributes and relationships between them. Agent is not a class present in FRBR, but it was defined in FRBR-object oriented (FRBRoo) [9]. By defining Person, Family and Corporate Body as subclasses of Agent, it became unnecessary to duplicate statements about the properties, as we see in section 5.2.

The terms of the vocabulary of classes are used to describe to which class the resource belongs. To describe it we might use the property *type*, provided by RDF language [16], and the URI representing the class:

- Work: http://www.rdaregistry.info/Elements/c#C10001
- Expression: http://www.rdaregistry.info/Elements/c#C10006
- Manifestation: http://www.rdaregistry.info/Elements/c#C10007
- Item: http://www.rdaregistry.info/Elements/c#C10003
- Agent: http://www.rdaregistry.info/Elements/c#C10002
- Person: http://www.rdaregistry.info/Elements/c#C10004
- Family: http://www.rdaregistry.info/Elements/c#C10008
- Corporate Body: http://www.rdaregistry.info/Elements/c#C10005.

The vocabulary of classes is identified by http://www.rdaregistry.info/Elements/c/. In RDA Registry, this namespace is represented by the prefix *rdac*.

5.2 Vocabularies of Properties Restricted to FRBR

There are different vocabularies of properties and different namespaces for each class (Work, Expression, Manifestation, Item and Agent). These five vocabularies contain properties created from attributes and relationships defined in RDA.

The property *has title proper* (Fig. 1), for example, was created from the attribute *Title proper* defined in RDA, rule 2.3.2. This property should be applied to manifestations, so it belongs to Manifestation vocabulary, which is identified by the namespace http://rdaregistry.info/Elements/m/ (prefix: *rdam*). We can use the property *has title proper* in a RDF statement by applying one of its URIs:

- http://www.rdaregistry.info/Elements/m/#P30156 (*rdam:P30156*): "Abstract or numeric URIs that don't contain a language-specific label can help maintain the language-neutral sense of the element or concept." [9];
- http://www.rdaregistry.info/Elements/m/#titleProper.en (*rdam:titleProper.en*): lexical URI with an English label; this URI redirect to numeric URI and might be modified over the time; URIs with labels in other languages might be created, as well.

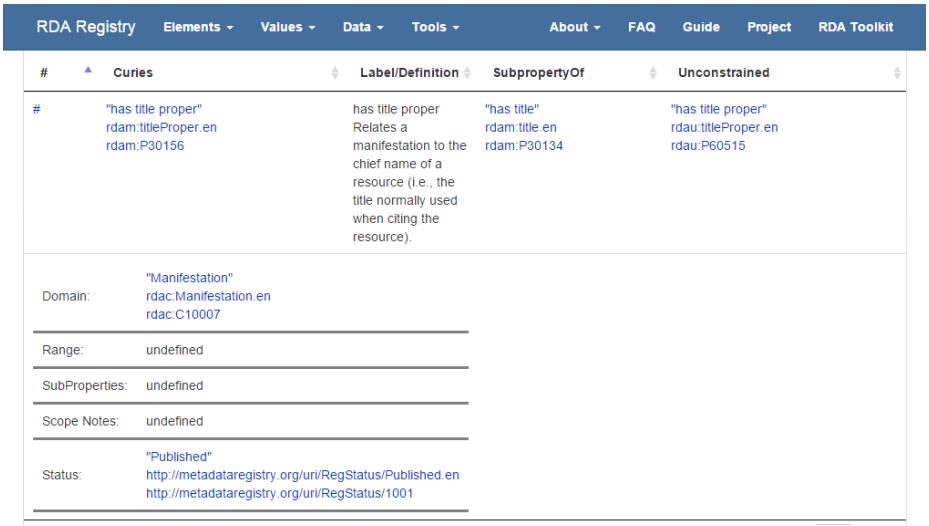


Fig. 1. The property *has title proper* in RDA Registry

Some properties are defined as sub-properties. This relation of sub-property occurs when we have a property with a broader meaning and another with a narrower meaning. The property *has title proper* (Fig. 1) is a sub-property of the property *has title* (http://www.rdaregistry.info/Elements/m/#P30134 or *rdam:P30134*).

The domain of each property is defined in these five vocabularies. The domain specifies in which classes a property can be used. The property *has title proper*, for example, is intended to be used for resources belonging to the Manifestation class, so the domain of this property is this class. In Fig. 1, the domain of the property *has title*

proper is identified by *rdac:C10007* (abbreviated form of http://www.rdaregistry.info/Elements/c/#C10007).

The range specifies which values can be used for the property or the class to which the values should belong. The range is defined for some of the properties of the five vocabularies. For instance, the range of the property *has title proper* (Fig. 1) is undefined, so this property can accept any kind of value, whether a literal or a resource, while the values of the property *has author* (Fig. 2) should be an individual of the class Agent.

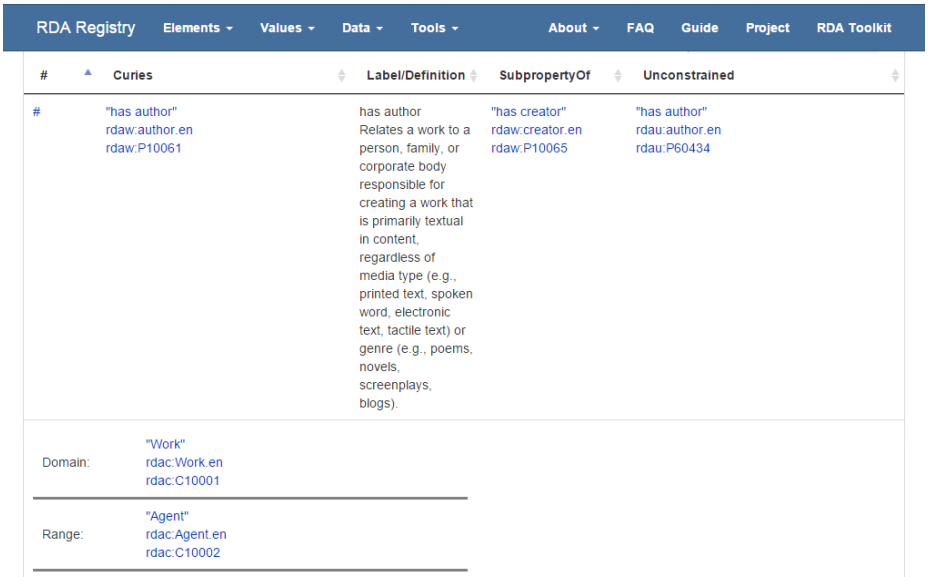


Fig. 2. The property *has author* in RDA Registry

Here we present an example of some properties used for works, expressions, manifestations, and item descriptions.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdaw="http://rdaregistry.info/Elements/w/"
xmlns:rdae="http://rdaregistry.info/Elements/e/"
xmlns:rdam="http://rdaregistry.info/Elements/m/"
xmlns:rdai="http://rdaregistry.info/Elements/i/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <rdf:Description rdf:about="http://example.com/work-1">
    <rdf:type
rdf:resource="http://rdaregistry.info/Elements/c/C10001"/
>
    <rdaw:P10088>The lord of the rings</rdaw:P10088>
    <rdaw:P10061 rdf:resource="http://example.com/person-
1"/>
```

```
  </rdf:Description>
  <rdf:Description
rdf:about="http://example.com/expression-1">
  <rdf:type
rdf:resource="http://rdaregistry.info/Elements/c/C10006"/
>
  <rdae:P20001>text</rdae:P20001>
  <rdae:P20006>English</rdae:P20006>
  <rdae:P20037 rdf:resource="http://example.com/person-
2"/>
  <rdae:P20231 rdf:resource="http://example.com/work-1"/>
</rdf:Description>
<rdf:Description
rdf:about="http://example.com/manifestation-1">
  <rdf:type
rdf:resource="http://rdaregistry.info/Elements/c/C10007"/
>
  <rdam:P30156>The lord of the rings</rdam:P30156>
  <rdam:P30088>Boston</rdam:P30088>
  <rdam:P30176>Houghton Mifflin Company</rdam:P30176>
  <rdam:P30011>2005</rdam:P30011>
  <rdam:P30181>1178 pages</rdam:P30181>
  <rdam:P30002>unmediated</rdam:P30002>
  <rdam:P30001>volume</rdam:P30001>
  <rdam:P30004>ISBN 978-0-618-64015-7</rdam:P30004>
  <rdam:P30139
rdf:resource="http://example.com/expression-1"/>
  <rdam:P30135 rdf:resource="http://example.com/work-1"/>
</rdf:Description>
<rdf:Description rdf:about="http://example.com/item-1">
  <rdf:type
rdf:resource="http://rdaregistry.info/Elements/c/C10003"/
>
  <rdai:P40047>Available only for university
students</rdai:P40047>
  <rdai:P40049
rdf:resource="http://example.com/manifestation-1"/>
</rdf:Description>
</rdf:RDF>
```

There are no specific vocabularies for the classes Person, Family and Corporate Body. These classes use the properties defined for Agent since they are subclasses of Agent and then they inherit the relevant properties. However, in the vocabulary of properties for Agent, some properties can be used for all the subclasses, while some can be used only for one subclass. The property *is singer of* can be used for all Agent

subclasses, but *has date of birth* can be used only for Person. In this sense, the class Agent makes the use and maintenance of vocabularies easier since it is not necessary to define several domains for a property that can be used for persons, families, and corporate bodies.

The domains and ranges defined in these five vocabularies make the properties useful only in the context in which FRBR conceptual model is implemented to some degree. Thus, the properties of these vocabularies are considered to be restricted to FRBR. However, as we describe in the next section, the Bibliographic Metadata Task Group also considered the need for vocabularies from RDA that can be used independent of the FRBR.

5.3 Unconstrained Properties

The FRBR conceptual model was developed in the library community and has not been broadly used for system design in this community. Outside the library community, there is little discussion about this model. Considering that, the development of vocabularies for exclusive use with FRBR would be a disadvantage, since they would not promote RDA use in the library community or outside that community.

To overcome the restriction imposed by FRBR and to make RDA vocabularies useful for any Semantic Web application, a vocabulary of FRBR unconstrained properties was created, published in namespace http://www.rdaregistry.info/Elements/u/ [14].

Similar to properties presented in section 5.2, the unconstrained properties were also created from RDA attributes and relationships. The difference, however, is that the domains and ranges of the unconstrained properties are undefined, thus the use of these properties does not require the use the classes Work, Expression, Manifestation, Item, Agent, Person, Family and Corporate Body. The property *has publisher's name* (Fig. 3) is an example of an unconstrained property created from the attribute *Publisher's name* defined in RDA, rule 2.8.4 [17].

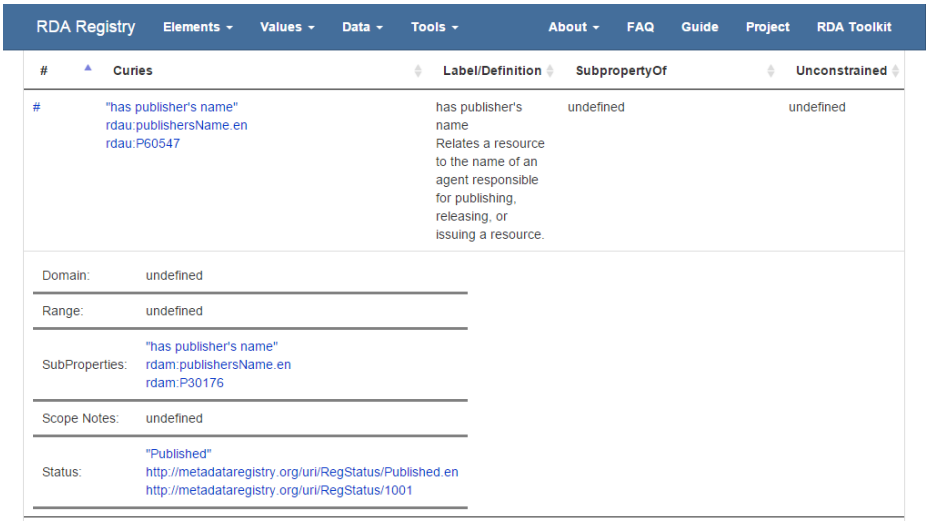
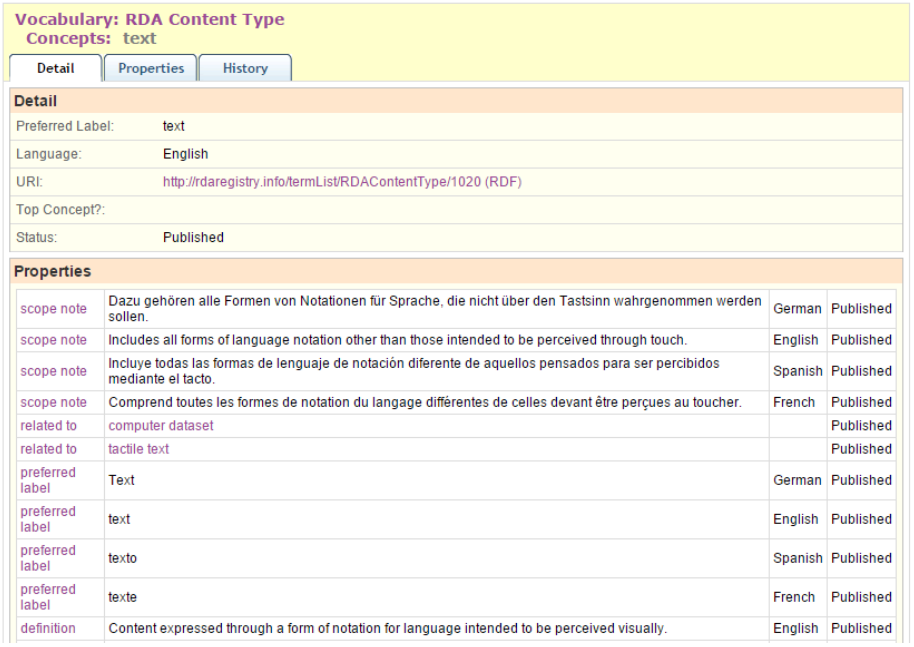


Fig. 3. The unconstrained property *has publisher's name* in RDA Registry

Despite their independence from FRBR, the unconstrained properties are associated with the restricted ones. In this relationship, unconstrained properties are considered broader than the restricted ones, and then the restricted properties are declared to be sub-properties of the unconstrained properties.

6 RDA Value Vocabularies

In addition to the vocabularies of classes and properties, there are vocabularies of values created from RDA. Some RDA attributes have their values defined by a list of terms, for example, *Carrier type* (RDA 3.3), *Content type* (RDA 6.9) and *Illustrative content* (RDA 7.15) [17]. These lists of terms were used to create the vocabularies for values. Certain vocabularies are already published in RDA Registry and in Open Metadata Registry, while others remain under development. Fig. 4 shows the term *text* that is part of the Content type vocabulary.



Vocabulary: RDA Content Type
Concepts: text

Detail | Properties | History

Detail

Preferred Label:	text
Language:	English
URI:	http://rdaregistry.info/termList/RDAContentType/1020 (RDF)
Top Concept?:	
Status:	Published

Properties

scope note	Dazu gehören alle Formen von Notationen für Sprache, die nicht über den Tastsinn wahrgenommen werden sollen.	German	Published
scope note	Includes all forms of language notation other than those intended to be perceived through touch.	English	Published
scope note	Incluye todas las formas de lenguaje de notación diferente de aquellos pensados para ser percibidos mediante el tacto.	Spanish	Published
scope note	Comprend toutes les formes de notation du langage différentes de celles devant être perçues au toucher.	French	Published
related to	computer dataset		Published
related to	tactile text		Published
preferred label	Text	German	Published
preferred label	text	English	Published
preferred label	texto	Spanish	Published
preferred label	texte	French	Published
definition	Content expressed through a form of notation for language intended to be perceived visually.	English	Published

Fig. 4. The term *text* from the Content Type vocabulary

Although the vocabularies of values have been created from RDA attributes and RDA attributes have been used also for creating the vocabularies of properties, these two kinds of vocabularies may be used in an independent form. For example, it is not mandatory that the value of the property *has content type* to be a term from Content type vocabulary.

7 Conclusions

After the brief explanation provided in this paper, we may remark on some conclusions regarding the RDA Element Sets and the RDA Value Vocabularies.

The decision to make the vocabularies of values independent of the vocabularies of properties allows the vocabularies to be opened in a way such that they can better meet the needs arising in different contexts. We can say the same for properties that usually have date or language codes as values but are presented in RDA Element Sets without ranges.

Since the unconstrained properties can be used in systems that are not FRBR-based, the unconstrained vocabulary may be used in the library community and beyond, for example, in museums, archives, publishers, and e-commerce.

We also think that an unconstrained vocabulary is very similar to metadata standards already used for data interchange in cataloging. One of these standards is MARC 21 Format for Bibliographic Data that, like an unconstrained vocabulary, does not separate bibliographic data into work, expression, manifestation, and item entities.

This similarity may promote the use of RDA unconstrained vocabularies for the publication of current bibliographic data as RDF triples.

Although the RDA standard is independent of any metadata standard and RDA vocabulary developers aim to make the use of these vocabularies independent of RDA [9], we highlight that the close relationship between RDA and RDA vocabularies has benefits and costs. On the positive side, we think that the high level of compatibility between them may encourage institutions that use RDA to use RDA vocabularies for publishing their data in RDF. On the other hand, the fact that RDA is a closed standard – with access only by subscription or purchase – may discourage the use of the RDA vocabularies by institutions that do not use RDA, since using open standards is a key feature for Semantic Web projects.

References

1. Alves, R.C.V.: Metadados como elementos do processo de catalogação. Ph.D. thesis, Ciência da Informação, Univ Estadual Paulista, Marília, Brazil (2010). http://hdl.handle.net/11449/103361
2. Catarino, M.E., Souza, T.B.: A representação descritiva no contexto da web semântica. Transinformação 24(2), 78-90 (2012)
3. Coyle, K.: Understanding the Semantic Web: Bibliographic Data and Metadata. ALA Library Technology Reports 46(1), 5-32 (2010)
4. Data Model Meeting, http://www.bl.uk/bibliographic/meeting.html
5. DCMI/RDA Task Group Repurposing Proposal, http://wiki.dublincore.org/index.php?title=DCMI/RDA_Task_Group_Repurposing_Proposal&oldid=1724
6. Dunsire, G., Hillmann, D., Phipps, J.: Reconsidering Universal Bibliographic Control in Light of the Semantic Web. Journal of Library Metadata 12(2-3), 164-176 (2012)
7. Dunsire, G., Willer, M.: Standard library metadata models and structures for the Semantic Web. Library Hi Tech News 28(3), 1-12 (2011)
8. Ferreira, J.A., Santos, P.L.V.A.C.: O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. Informação & Sociedade: estudos 23(2), 13-23 (2013)
9. Hillmann, D., Coyle, K., Phipps, J., Dunsire, G.: RDA Vocabularies: Process, Outcome, Use. D-Lib Magazine 16(1-2), (2010)
10. Open Metadata Registry, http://metadataregistry.org
11. OWL Web Ontology Language: Reference, http://www.w3.org/TR/2004/REC-owl-ref-20040210
12. Publication of RDA Content, Carrier, and Media Type Vocabularies, http://www.rda-jsc.org/rdatypespublish.html
13. Publication of the RDA Element Vocabularies, http://www.rda-jsc.org/RDAelementvocabs.html
14. RDA Registry Website, http://www.rda-jsc.org/RDAregistrywebsite.html
15. RDA Registry, http://www.rdaregistry.info
16. RDF 1.1 Primer, http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624
17. Resource Description and Access, http://access.rdatoolkit.org