Resource Description Framework: Lingua Franca for the Semantic Web

Justine Withers

San Jose State University

Abstract

Resource Description Framework (RDF) is a lowest-common-denominator model of encoding data and metadata. Although its simple graph format creates single statements linking a subject (property) with an object (value), every statement potentially can be linked to any other statement, creating a huge web of data. This is the foundation for the Semantic Web, connecting many databases and data syntaxes with a hope for more intelligent computing and data that contains its own meaning.

Resource Description Framework: Lingua Franca for the Semantic Web

Open any first-level foreign language textbook. What do you learn?

Carlos conoce Linda. (Carlos knows Linda.)

Miette est mon ami. (Miette is my friend.)

Hundo estas besto. (A dog is an animal.) (Cresswell & Hartley, 1959)

Dette er en bok. (This is a book.) (Haugen, n.d.)

Simple sentences: a subject, verb, and object. If you can use only these types of sentences, it will take you a while to communicate, but you will eventually get your point across. That is what the Resource Description Framework (RDF) model does: make simple connections that can be arranged to tell a larger story and be readily understood by a machine.

Caplan (2003) defines RDF as a "*data model* [emphasis mine] for representing resources, their properties, and the values of those properties, and in theory the data model could be expressed in any number of syntaxes" (p. 20).

RDF was born from the idea of the Semantic Web and Tim Berners-Lee's desire to create a "global database." In 1998, Berners-Lee described it thus:

> When looking at a possible formulation of a universal Web of semantic assertions, the principle of minimalist design requires that it be based on a common model of great generality. Only when the common model is general can any prospective application be mapped onto the model. The general model is the Resource Description Framework.

The RDF specifications are maintained by W3C, currently in Recommendation Status, and have not been significantly updated since 2004. *RDF 1.1 Concepts and Abstract Syntax* is in progress along with revisions and additions to the larger sphere of RDF specifications. (W3C, 2012a). Current standards and proposed changes are available at http://www.w3.org/standards/techs/rdf#w3c_all.

RDF Basics

Pollock (2009) highlights an important distinction of RDF. It is designed for machines, not people. It encodes information from any number of schema into simple statements that can be easily shared among systems. “Translation into the language of RDF properties and classes puts heretofore disparate languages of description into a unified grammatical foundation” (Baker, 2012, p 117). Baker compares RDF’s use as a common denominator among “local data formats” to “English as a second language,” which can serve as “a basis for communication among non-native English speakers” (p. 128). All of these “foreign languages” are defined by a relevant namespace, “web-accessible versions of the metadata schema” (Caplan, 2003). [Author note: I wrote down this quote and neglected to record the page number before returning the book to the library.]

RDF is usually expressed in XML, although it does not have to be (Moore, 2012, p. 40).

Dublin Core was one of the first “vocabularies to be declared as a set of RDF properties,” way back in 1999, although its implementation has not yet been entirely smoothed out (Baker, 2012, p. 119–121). The Dublin Core Metadata Initiative Bibliographic (DCMI) Metadata Task Group reorganized in 2012 with the goal of defining “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" (DCMI, 2012).

SKOS (Simple Knowledge Organization System) uses RDF to represent "knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web" (W3C, 2012b).

In October, 2011, The Library of Congress announced it is working on a system for converting MARC records to RDF (Marcum, 2011).

Sanchez, et al., (2012) provide an example of RDF's limits. In creating a system for aggregating and organizing open-access research, the RDF system of binary statements ("to link two individuals or an individual and a value" [W3C, 2006]) inadequately describes the more complex relationships that usually exist between pieces of information. They instead "propose the use of the Web Ontology Language Description Logic (OWL-DL) to take advantage of tools that enable efficient reasoning about the knowledge represented in the ontologies of records and to maintain compatibility with RDF ontologies" (p. 54). Note that OWL is based on RDF.

RDF in Detail

Although some consider RDF a theory more than a practice, resources abound, both for its technical specifics and for its role in the wider world of linked data and the Semantic Web.

All technical specifications are available on the W3C RDF Web site (W3C, 2012, February).

Moore (2012) concisely describes RDF, places it in the wider world of Semantic Web technology, and provides a skeptic's view of its benefits and future viability.

Karen Coyle is widely known and influential on the topic of linked data. Her Web site contains presentations on RDF basics and its potential use in libraries and knowledge management systems (Coyle, n.d.).

Linked Data Tools offer detailed and clear tutorials on RDF and its use in linked data, including queries (Linked Data Tools, 2009).

The Linked Data organization hosts a website that serves as a central resource for the linked data community. Although many contributors assume working knowledge of RDF structure, their applications deepen understanding of the model (Linked Data, n.d.).

Despite the condescending title, I found Pollock's *Semantic Web for Dummies* (2009) has the most succinct explanation of RDF, its uses, and its potential.

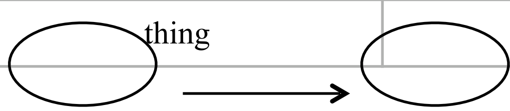
RDF in Practice

RDF is built on the "triple" or "statement." Any triple must have an subject, predicate, and object. Let's return to our simple sentences in various languages.

Carlos	conoce	Linda.
subject	predicate	object

A similar statement could be made about an information resource:

Resource	has title	Very Simple and Cool Explanation of RDF
subject	predicate	object
thing	property	value



Each of these statements can be "graphed."

Resource	has title	"Title"

node arc node (Pollock, 2009)

Nodes can have multiple arcs and link to other node-arc-node connections. The graph structure differentiates it from relational and hierarchical data formats (Pollock, 2009).

When learning a language, the teacher warns of false cognates: words that seem like they would have the same meaning in both languages but do not. For example, if you tell someone in Madrid that you are embarazada, they will ask when the baby is coming, not why your cheeks are red.

Various schema have false cognates as well. For example, the resource in the example above has a title. In Dublin Core, the title element holds the name of the resource. In Friend of a Friend, it is an appellation: Mr., Lady, Sra. How does the computer know which is which? The namespace declaration and prefixes tell the system how to interpret elements.

Another thing to keep in mind, is that our RDF triples can be grammatically correct and make no sense, as pointed out by P. Schreur (Linked Data as Revolution presentation to Northern California Technical Services Group, May 18, 2012). Just like we can say *the dog has rain* or *my shoes are friendly,* we can join a book to a video format or an author name to a place. The correctness of a statement relies on the cataloger's input.

More relevant to RDF statements is the ambiguous nature of words. A human can read the sentence *the sheep are on the green* and understand that the sheep are eating grass, not inhabiting a colorspace. A computer is stupid this way and needs to know exactly what we mean by *sheep* and *green* and *are.* Dunsire (2012) effectively explains the importance of "using a special form of identifier, the uniform resource identifier or URI" in compensating for a computer's inability to understand ambiguous terms.

> In principle, a URI is any unique combination of letters, numbers, and punctuation, although some punctuation marks are not allowed....A particular form of URI, the http URI, exploits the existing http uniform resource locator or URL....An http URI looks just like a URL but it does not necessarily act like a URL. For example, the URI for the Library of Congress Subject Heading (LCSH) "mathematical physics" is http://id.loc.gov/authorities/subjects/sh85082129. (p. 725)

A normal Web browser will use that URI to find the LC web page on the mathematical physics subject heading. A Semantic Web browser will "return machine-readable RDF statements rather than a human-readable Web page" (p. 726).

An RDF statement could just as easily link a resource to an ISBN (International Standard Book Number), a URI maintained by ISBN Agencies in various countries or indeed, use the ISBN to identify the resource itself. In addition, although the subject must be a URI of some sort, the object can be a literal string, such as a date or local term (Dunsire, 2012, p. 726).

So let's look at a sample of RDF statements, expressed in XML. (Sample from Pollock, 2009, with annotations by the author).

First, tell the machine to expect some XML.

```
<?xml version="1.0"?>
```

Then, tell the machine that this is a Resource Description Framework (RDF) format.

```
<rdf:RDF
```

The next two lines declare the namespaces. The first says, Whenever you see “rdf,” go to the W3 site to decipher the property. The second says, When you see “dc,” go to the Dublin Core site (the domain name hosted by the Permanent URL organization [PURL]) to decipher the property.

```
xmlns:rdf=”http://www.w3.org/1999/02/22-rdf-syntax-ns#”
xmlns:dc=”http://purl.org/dc/elements/1.1”>
```

Here is the start of our first RDF triple. The Description element acts as the subject, saying that anything wrapped in this element is about this resource: the web page http://me.jtpollock.us.

```
<rdf:Description rdf:about=”http://me.jtpollock.us/”>
```

Now here is the predicate (has title) and object (Jeff’s homepage!) of the first triple. Because of the “dc” prefix, the machine knows to use Dublin Core syntax to interpret this statement.

```
<dc:title>Jeff’s homepage!</dc:title>
```

We can say something else about this resource and create a new triple with the same original node (our subject). This predicate (creator) will arc to a different node (object). The web page was created by JT Pollock, who is described by an FOAF URI. Notice it is pointing to a separate RDF file.

```
<dc:creator rdf:resource=”http:/me.jt.pollock.us/foaf.rdf#me”/>
```

Here ends our RDF description statements.

```
</rdf:Description>
```

Don’t forget to close the entire RDF structure.

```
</rdf:RDF>
```

RDF in the Future

Recent studies are looking at ways to reduce the unwieldiness of large-scale RDF databases (Ohsawa, Amagasa & Kitagawa, 2012; MahmoudiNasab & Sakr, 2012).

Baker (2012, p. 129) emphasizes the importance of preserving RDF vocabularies. Because the schemes are maintained by many institutions and individuals, their lifetime is only as long as that of their creators. Baker points to the arrangement between DCMI and the Friend of a Friend project: DCMI mirrors FOAF data and will take it over if FOAF ends operations. Mutual support between organizations will help keep RDF data available.

Alemu, Stevens, and Ross (2012) assert that RDF and its descendants are presently an academic abstraction and "there are no viable semantic web related metadata solutions in widespread use [at the present time]" (p. 44). Howarth (2012), Baker (2012), and Dunsire (2012) offer a more optimistic view, at least in the bibliographic world, describing several initiatives that rely on RDF triples to link bibliographic data, including the translation of Library of Congress Subject Headings in 2009. Dunsire is especially positive: "RDF representations of the [Functional Requirements] models offer a powerful way of relating them to other bibliographic schemas and supporting the transformation of metadata statements from one format to another." However, he goes on to say that it will be the *"cataloger of tomorrow"* [emphasis mine] using RDF statements and the Semantic Web (p. 740).

Heath and Bizer (2011) describe the spread of linked data, using RDF triples, as "significant." Their examples cover cross-domain research, geographic data, media, government information, libraries,and commerce. In my estimation, it remains to be seen if and when legacy systems will be converted to RDF (such as library catalogs) and if a lay Web content creator will use URIs or have the ability to convert content to meaningful triples.

References

Alemu, G., Stevens, B., & Ross, P. (2012). Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: A social constructivist approach. *New Library World, 113*(1/2), 38–54.

Baker, T. (2012). Libraries, languages of description, and linked data: a Dublin Core perspective. *Library Hi Tech, 30*(1), 116–133.

Berners-Lee, T. (1998). Semantic Web roadmap. Retrieved from http://www.w3.org/DesignIssues/Semantic.html

Caplan, P. (2003). *Metadata Fundamentals for All Librarians.* Chicago: American Library Association.

Coyle, K. (n.d.). Karen Coyle's home page. Retrieved from http://www.kcoyle.net

Cresswell, J. and Hartley, J. (1959). *Teach yourself Esperanto.* London: English Universities Press.

Dublin Core Metadata Initiative (DCMI). (2012, May 8). Bibliographic Metadata Task Group. Retrieved from http://wiki.dublincore.org/index.php/Bibliographic_Metadata_Task_Group

Dunsire, G. (2012). Representing the FR family in the Semantic Web. *Cataloging & Classification Quarterly, 50,* 724–741.

Haugen, E. (n.d.). *Beginning Norwegian. 3rd ed.* New York: Appleton-Century-Crofts, Inc.

Heath, T., & Bizer, C. (2011). Web of data. In *Linked Data: Evolving the Web into a Global Data Space* (chapter 3). Retrieved from http://linkeddatabook.com/editions/1.0/#htoc23

Howarth, L.C. (2012). FRBR and linked data: Connecting FRBR and linked data. *Cataloging & Classification Quarterly, 50,* 763–776.

Linked Data Tools. (2009). Tutorial 2: Introducing RDF/XML. Retrieved from http://www.linkeddatatools.com/introducing-rdf-part-2

MahmoudiNasab, H., & Sakr, S. (2012). AdaptRDF: adaptive storage management for RDF databases. *International Journal of Web Information Systems, 8*(2), 234–250.

Marcum, D. (2011, October 31). A bibliographic framework for the digital age. Retrieved from http://www.loc.gov/marc/transition/news/framework-103111.html

Moore, M. (2012, March). The semantic web: An introduction for information professionals. *Indexer, 30* (1), 38–43.

Ohsawa, S., Amagasa, T., & Kitagawa, H. (2012). RDF packages: a scheme for efficient reasoning and querying over large-scale RDF data. *International Journal of Web Information Systems, 8*(2), 212–233.

Sánchez, A., Auxilio Medina, M., Starostenko, O., Benitez, A., & López Domínguez, E. (2012). Organizing open archives via lightweight ontologies to facilitate the use of heterogeneous collections. *Aslib Proceedings, 64*(1), 46–66.

W3C. (2006, April 12). Defining n-ary relations on the Semantic Web. Retrieved from http://www.w3.org/TR/swbp-n-aryRelations/

W3C. (2012, February 3). Resource Description Framework (RDF). Retrieved from http://www.w3.org/RDF/

W3C. (2012a). RDF current status. Retrieved from http://www.w3.org/standards/techs/rdf#w3c_all

W3C. (2012b). Introduction to SKOS. Retrieved from http://www.w3.org/2004/02/skos/intro