STORIES

D-Lib Magazine
May 1998

ISSN 1082-9873

An Introduction to the Resource Description Framework

Eric Miller
Research Scientist
Online Computer Library Center, Inc.
Office of Research
Dublin, Ohio
emiller@oclc.org

Abstract

The Resource Description Framework (RDF) is an infrastructure that enables the encoding, exchange and reuse of structured metadata. RDF is an application of XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics. RDF additionally provides a means for publishing both human-readable and machine-processable vocabularies designed to encourage the reuse and extension of metadata semantics among disparate information communities. The structural constraints RDF imposes to support the consistent encoding and exchange of standardized metadata provides for the interchangeability of separate packages of metadata defined by different resource description communities.

Introduction

The World Wide Web affords unprecedented access to globally distributed information. Metadata, or structured data about data, improves discovery of and access to such information. The effective use of metadata among applications, however, requires common conventions about semantics, syntax, and structure. Individual resource description communities define the semantics, or meaning, of metadata that address their particular needs. Syntax, the systematic arrangement of data elements for machine-processing, facilitates the exchange and use of metadata among multiple applications. Structure can be thought of as a formal constraint on the syntax for the consistent representation of semantics.

The Resource Description Framework (RDF), developed under the auspices of the World Wide Web Consortium (W3C) [W3C], is an infrastructure that enables the encoding, exchange, and reuse of structured metadata. This infrastructure enables metadata interoperability through the design of mechanisms that support common conventions of semantics, syntax, and structure. RDF does not stipulate semantics for each resource description community, but rather provides the ability for these communities to define metadata elements as needed. RDF uses XML (eXtensible Markup Language) as a common syntax for the exchange and processing of metadata. The XML syntax is a subset of the international text processing standard SGML (Standard Generalized Markup Language [SGML]) specifically intended for use on the Web. The XML syntax provides vendor independence, user extensibility, validation, human readability, and the ability to represent complex structures. By exploiting the features of XML, RDF imposes structure that provides for the unambiguous expression of semantics and, as such, enables consistent encoding, exchange, and machine-processing of standardized metadata.

RDF supports the use of conventions that will facilitate modular interoperability among separate metadata element sets. These conventions include standard mechanisms for representing semantics that are grounded in a simple, yet powerful, data model discussed below. RDF additionally provides a means for publishing both human-readable and machine-processable vocabularies. Vocabularies are the set of properties, or metadata elements, defined by resource description communities. The ability to standardize the declaration of vocabularies is anticipated to encourage the reuse and extension of semantics among disparate information communities. For example, the Dublin Core Initiative [DC], an international resource description community focusing on simple resource description for discovery, has adopted RDF [DCRDF]. Educom's IMS Instructional Metadata System [IMS], designed to provide access to educational materials, has adopted the Dublin Core and corresponding architecture and extended it with domain-specific semantics. RDF is designed to support this type of semantic modularity by creating an infrastructure that supports the combination of distributed attribute registries. Thus, a central registry is not required. This permits communities to declare vocabularies which may be reused, extended and/or refined to address application or domain specific descriptive requirements.

The goals of RDF are broad, and the potential opportunities are enormous. This introduction to RDF begins by discussing the background context of the RDF initiative and relates it to other metadata activities. A discussion of the functionality of RDF and an overview of the model, schema and syntactic considerations of this framework follow.

Background

The history of metadata at the W3C began in 1995 with PICS, the Platform for Internet Content Selection [PICS]. PICS is a mechanism for communicating ratings of web pages from a server to clients. These ratings, or rating labels, contain information about the content of web pages: for example, whether a particular page contains a peer-reviewed research article, or was authored by an accredited researcher, or contains sex, nudity, violence, foul language, etc. Instead of being a fixed set of criteria, PICS introduced a general mechanism for creating rating systems. Different organizations could rate content based on their own objectives and values, and users -- for example, parents worried about their children's web usage -- could set their browsers to filter out any web pages not matching their own criteria. Development of PICS was motivated by the anticipation of restrictions on Internet content in the USA and elsewhere.

Through a series of meetings with the digital library community, limitations in the PICS specifications were identified, and functional requirements were outlined to address the more general problem of associating descriptive information with Internet resources based on the PICS architecture. As a result of these discussions, the W3C formed a new working group, PICS-NG Next Generation [PICSNG] to address the more general issues of resource description.

Shortly after the PICS-NG working group was chartered, it became clear that the infrastructure designed in the early document specifications [PICSMOD] was applicable in several additional applications. As a result, the W3C consolidated these applications as the W3C Resource Description Framework working group.

RDF is the result of a number of metadata communities bringing together their needs to provide a robust and flexible architecture for supporting metadata on the web. While the development of RDF as a general metadata framework, and as such, a simple knowledge representation mechanism for the web, was heavily inspired by the PICS specification [PICSSPEC], no one individual or organization invented RDF. RDF is a collaborative design effort. Several W3C Member companies are contributing intellectual resources. It is drawing upon the XML [XML] design as well as proposals submitted by Microsoft's [XMLDATA] and Netscape [MCFXML]. Other metadata efforts, such as the Dublin Core [DC] and the Warwick Framework [WF] have also influenced the design of the RDF.

The RDF Data Model

RDF provides a model for describing resources. Resources have properties (attributes or characteristics). RDF defines a *resource* as any object that is uniquely identifiable by an Uniform Resource Identifier (URI) [URI1][URI2]. The properties associated with resources are identified by property-types, and property-types have corresponding values. *Property-types* express the relationships of values associated with resources. In RDF, *values* may be atomic in nature (text strings, numbers, etc.) or other resources, which in turn may have their own properties. A collection of these properties that refers to the same resource is called a *description*. At the core of RDF is a syntax-independent model for representing resources and their corresponding descriptions [SPEC]. The following graphic (Figure 1) illustrates a generic RDF description.

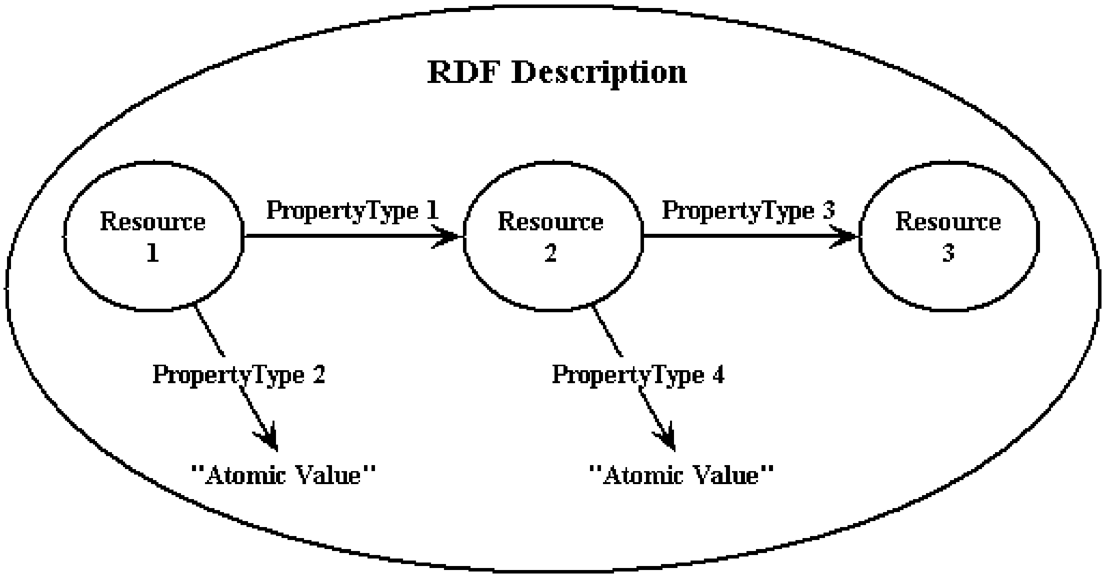

Figure 1

The application and use of the RDF data model can be illustrated by concrete examples. Consider the following statements:

1. "The author of Document 1 is John Smith"
2. "John Smith is the author of Document 1"

To humans, these statements convey the same meaning (that is, John Smith is the author of a particular document). To a machine, however, these are completely different strings. Whereas humans are extremely adept at extracting meaning from differing syntactic constructs, machines remain grossly inept. Using a triadic model of resources, property-types and corresponding values, RDF attempts to provide an unambiguous method of expressing semantics in a machine-readable encoding.

RDF provides a mechanism for associating properties with resources. So, before anything *about* Document 1 can be said, the data model requires the declaration of a resource *representing* Document 1. Thus, the data model corresponding to the statement "the author of Document 1 is John Smith" has a single resource `Document 1`, a property-type of `author` and a corresponding value of `John Smith`. To distinguish characteristics of the data model, the RDF Model and Syntax specification [SPEC] represents the relationships among resources, property-types, and values in a directed labeled graph. In this case, resources are identified as nodes, property-types are defined as directed label arcs, and string values are quoted. Given this representation, the data model corresponding to the statement is graphically expressed as (Figure 2):

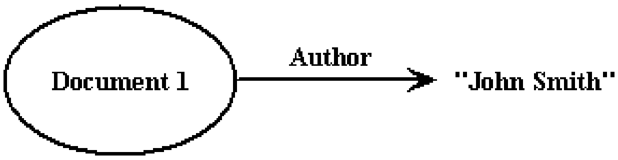

Figure 2

If additional descriptive information regarding the author were desired, e.g., the author's email address and affiliation, an elaboration on the previous example would be required. In this case, descriptive information *about* John Smith is desired. As was discussed in the first example, before descriptive properties can be expressed about the *person* John Smith, there needs to be a unique identifiable resource *representing* him. Given the directed label graph notation in the previous example, the data model corresponding to this description is graphically represented as (Figure 3):

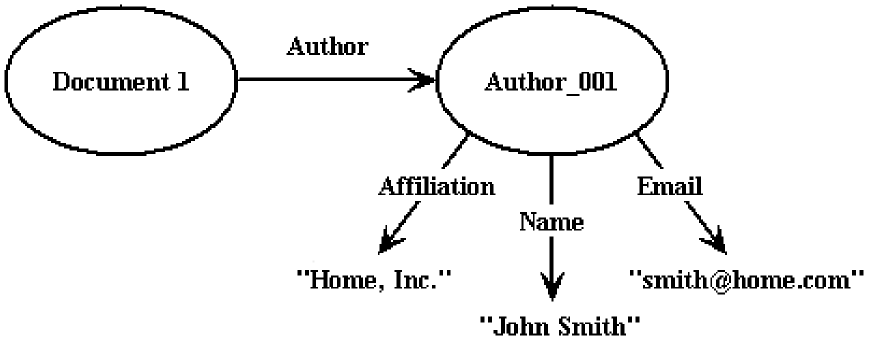

Figure 3

In this case, "John Smith" the string is replaced by a uniquely identified resource denoted by `Author_001` with the associated property-types of `name`, `email` and `affiliation`. The use of unique identifiers for resources allows for the unambiguous association of properties. This is an important point, as the person John Smith may be the value of several different property-types. John Smith may be the author of Document 1, but also may be the value of a particular company describing the set of current employees. The unambiguous identification of resources provides for the reuse of explicit, descriptive information.

In the previous example the unique identifiable resource for the author was created, but not for the author's name, email or affiliation. The RDF model allows for the creation of resources at multiple levels. Concerning the representation of personal names, for example, the creation of a resource representing the author's name could have additionally been described using "firstname", "middlename" and "surname" property-types. Clearly, this iterative descriptive process could continue down many levels. What, however, are the practical and logical limits of these iterations?

There is no one right answer to this question. The answer is dependent on the domain requirements. These issues must be addressed and decided upon in the standard practice of individual resource description communities. In short, experience and knowledge of the domain dictate which distinctions should be captured and reflected in the data model.

The RDF data model additionally provides for the description of other descriptions. For instance, often it is important to assess the credibility of a particular description (e.g., "The Library of Congress told us that John Smith is the author of Document 1"). In this case the description tells us something about the statement "John Smith is the author of Document 1", specifically, that the Library of Congress asserts this to be true. Similar constructs are additionally useful for the description of collections of resources. For instance, "John Smith is the author of Documents 1, 2, and 3". While these statements are significantly more complex, the same data model is applicable. A more detailed discussion of these issues is outside the scope of this overview, but more information is available in the RDF Model and Syntax Specification [SPEC].

The RDF Syntax

RDF defines a simple, yet powerful model for describing resources. A syntax representing this model is required to store instances of this model into machine-readable files and to communicate these instances among applications. XML is this syntax. RDF imposes formal structure on XML to support the consistent representation of semantics.

RDF provides the ability for resource description communities to define semantics. It is important, however, to disambiguate these semantics among communities. The property-type "author", for example, may have broader or narrower meaning depending on different community needs. As such, it is problematic if multiple communities use the same property-type to mean very different things. To prevent this, RDF uniquely identifies property-types by using the XML namespace mechanism [NS]. XML namespaces provide a method for unambiguously identifying the semantics and conventions governing the particular use of property-types by uniquely identifying the governing authority of the vocabulary. For example, the property-type "author" defined by the Dublin Core Initiative as the "person or organization responsible for the creation of the intellectual content of the resource" and is specified by the Dublin Core CREATOR element [DCES]. An XML namespace is used to unambiguously identify the Schema for the Dublin Core vocabulary by pointing to the definitive Dublin Core resource that defines the corresponding semantics. Additional information on RDF Schemas is discussed latter. If the Dublin Core RDF Schema, however, is abbreviated as "DC", the data model representation for this example would be (Figure 4):

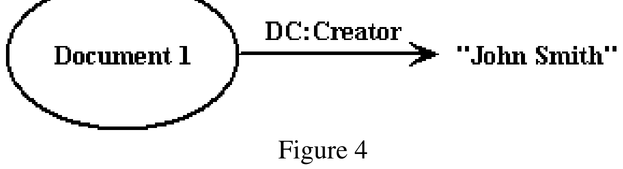

Figure 4

This more explicit declaration identifies a resource `Document 1` with the semantics of property-type `Creator` unambiguously defined in the context of `DC` (the Dublin Core vocabulary). The value of this property-type is `John Smith`.

The corresponding syntactic way of expressing this statement using XML namespaces to identify the use of the Dublin Core Schema is:

```
<?xml:namespace ns = "http://www.w3.org/RDF/RDF/" prefix ="RDF" ?>
<?xml:namespace ns = "http://purl.oclc.org/DC/" prefix = "DC" ?>

<RDF:RDF>
  <RDF:Description RDF:HREF = "http://uri-of-Document-1">
    <DC:Creator>John Smith</DC:Creator>
  </RDF:Description>
</RDF:RDF>
```

In this case, both the RDF and Dublin Core schemas are declared and abbreviated as "RDF" and "DC" respectively. The RDF Schema is declared as a boot-strapping mechanism for the declaration of the necessary vocabulary needed for expressing the data model. The Dublin Core Schema is declared in order to utilize the vocabulary defined by this community. The URI associated with the namespace declaration references the corresponding schemas. The element `<RDF:RDF>` (which can be interpreted as the element `RDF` in the context of the `RDF` namespace) is a simple wrapper that marks the boundaries in an XML document where the content is explicitly intended to be mappable into an RDF data model instance [SPEC]. The element `<RDF:Description>` (the element `Description` in the context of the `RDF` namespace) is correspondingly used to denote or instantiate a resource with the corresponding URI `http://uri-of-Document-1`. And the element `<DC:Creator>` in the context of the `<RDF:Description>` represents a property-type `DC:Creator` and a value of `"John Smith"`. The syntactic representation is designed to reflect the corresponding data model.

In the more advanced example, where additional descriptive information regarding the author is required, similar syntactic constructs are used. In this case, while it may still be desirable to use the Dublin Core CREATOR property-type to represent the person responsible for the creation of the intellectual content, additional property-types "name", "email" and "affiliation" are required. For this case, since the semantics for these elements are not defined in Dublin Core, an additional resource description standard may be utilized. It is feasible to assume the creation of an RDF schema with the semantics similar to the vCard [VC] specification designed to automate the exchange of personal information typically found on a traditional business card, could be introduced to describe the author of the document. The data model representation for this example with the corresponding business card schema defined as CARD would be (Figure 5):

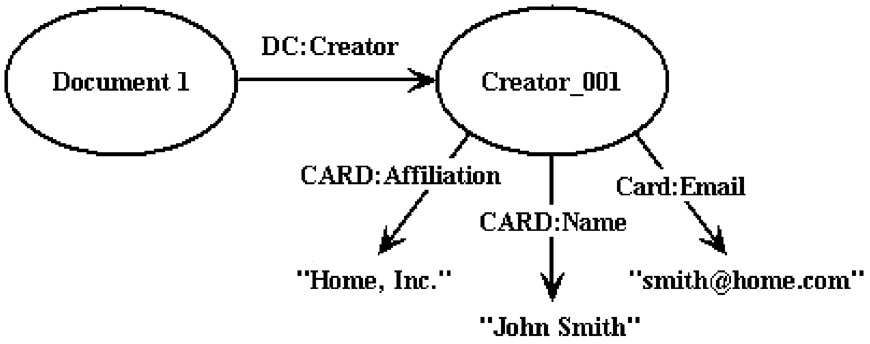

Figure 5

This, in turn, could be syntactically represented as

```
<?xml:namespace ns = "http://www.w3.org/RDF/RDF/" prefix = "RDF" ?>
<?xml:namespace ns = "http://purl.oclc.org/DC/" prefix = "DC" ?>
<?xml:namespace ns = "http://person.org/BusinessCard/" prefix = "CARD" ?>

<RDF:RDF>
  <RDF:Description RDF:HREF = "http://uri-of-Document-1">
    <DC:Creator RDF:HREF = "#Creator_001"/>
  </RDF:Description>

  <RDF:Description ID="Creator_001">
    <CARD:Name>John Smith</CARD:Name>
    <CARD:Email>smith@home.net</CARD:Email>
    <CARD:Affiliation>Home, Inc.</CARD:Affiliation>
  </RDF:Description>
</RDF:RDF>
```

in which the RDF, Dublin Core, and the "Business Card" schemas are declared and abbreviated as "RDF", "DC" and "CARD" respectively. In this case, the value associated with the property-type `DC:Creator` is now a resource. While the reference to the resource is an internal identifier, an external URI, for example, to a controlled authority of names, could have been used as well. Additionally, in this example, the semantics of the Dublin Core CREATOR element have been refined by the semantics defined by the schema referenced by CARD. This construct is similar to the Warwick Framework [WF], a recognition of separate maintainable and interchangeable packages of descriptive information used in the description of resources. The structural constraints RDF imposes to support the consistent encoding and exchange of standardized metadata provides for the interchangeability of separate packages of metadata defined by different resource description communities.

The RDF Schema

RDF Schemas are used to declare vocabularies, the sets of semantics property-types defined by a particular community. RDF schemas define the valid properties in a given RDF description, as well as any characteristics or restrictions of the property-type values themselves. The XML namespace mechanism serves to identify RDF Schemas.

A human and machine-processable description of an RDF schema may be accessed by de-referencing the schema URI. If the schema is machine-processable, it may be possible for an application to learn some of the semantics of the property-types named in the schema. To understand a particular RDF schema is to understand the semantics of each of the properties in that description. RDF schemas are structured based on the RDF data model. Therefore, an application that has no understanding of a particular schema will still be able to parse the description into the property-type and corresponding values and will be able to transport the description intact (e.g., to a cache or to another application).

The exact details of RDF schemas are currently being discussed in the W3C RDF Schema working group [SCHEMA]. It is anticipated, however, that the ability to formalize human-readable and machine-processable vocabularies will encourage the exchange, use, and extension of metadata vocabularies among disparate information communities. RDF schemas are being designed to provide this type of formalization.

Conclusions

The World Wide Web affords unprecedented access to distributed information. Metadata improves access to this information and RDF is a W3C proposed standard for defining the architecture necessary for supporting web metadata. RDF is an application of XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics for the consistent encoding, exchange, and machine processing of metadata. RDF additionally, provides means for publishing both a human-readable and a machine-processable vocabularies designed to encourage the exchange, use and extension of metadata semantics among disparate information communities.

The potential implications of widespread adoption of RDF metadata on the web are captured by Ora Lassila, editor of the RDF Model and Syntax Specification, in the following [IRDF]. [IRDF].

> Once the web has been sufficiently "populated" with rich metadata, what can we expect? First, searching on the web will become easier as search engines have more information available, and thus searching can be more focused. Doors will also be opened for automated software agents to roam the web, looking for information for us or transacting business on our behalf. The web of today, the vast unstructured mass of information, may in the future be transformed into something more manageable - and thus something far more useful.

The effective use of metadata among applications requires common conventions about semantics, syntax, and structure. The design of enabling infrastructures such as RDF to support these constructs provides the necessary foundations to support the management of information on the web and, as Ora suggests, provides the ability for transforming the web into a more useful and powerful information resource.

Acknowledgment

The author is indebted to the several persons for helping with this article. Specifically, the author would like to credit Erik Jul, Ora Lassila, Ralph LeVan, R.V. Guha, Bob Schloss, Ralph Swick, and Stuart Weibel for their comments, criticism, insight and support.

Bibliography

[DC] The Dublin Core Home Page,
URL:http://purl.oclc.org/metadata/dublin_core

[DCES] Description of the Dublic Core Elements,
URL:http://purl.oclc.org/metadata/dublin_core_elements

[DCRDF] OCLC News Release, "Dublin Core and Web MetaData Standards Converge in Helsinki", Nov. 7, 1997,
URL:http://www.oclc.org/oclc/press/971107a.htm

[DSIG] W3C Digital Signature Working Group URL:http://www.w3.org/DSig/

[IMS] Educom's IMS Instructional Metadata System Home Page,
URL:http://www.imsproject.org

[IRDF] Introduction to RDF Metadata, W3C NOTE 1997-11-13, Ora Lassila,
URL:http://www.w3.org/TR/NOTE-rdf-simple-intro

[MCFXML] R.V. Guha and Tim Bray, "Meta Content Framework using XML", June 6, 1997, http://www.w3.org/TR/NOTE-MCF-XML/

[NS] Name Spaces in XML, Tim Bray, Dave Hollander, and Andrew Layman, W3C Note, 19-January-1998,
URL:http://www.w3.org/TR/1998/NOTE-xml-names-0119

[PICS] W3C PICS The Platform for Content Selection Home Page,
URL:http://www.w3.org/PICS

[PICSMOD] PICS-NG Metadata Model and Label Syntax, Ora Lassila, Version 3.5, 5/14/97, URL:http://www.w3.org/TR/NOTE--pics-ng-metada

[PICSNG] W3C PICS-NG The Platform for Content Selection - Next Generation Home Page, URL:http://www.w3.org/PICS/NG

[PICSSPEC] PICS Label Distribution Label Syntax and Communication Protocols, Version 1.1, W3C Recommendation 31-October-96 URL:http://www.w3.org/TR/REC-PICS-labels-961031

[SCHEMA] The W3C RDF Schema Working Group, URL:http://www.w3.org/TR/WE-RDF-Schema/

[SGML] Information Processing -- Text and Office Systems -- Standard Generalized Markup Language (SGML), International Organization for Standardization, Rdf No. ISO 8879:1986, 1986.

[SPEC] Resource Description Framework (RDF) Model and Syntax, URL:http://www.w3.org/RDF/Group/WD-rdf-syntax/

[URI1] IETF RFC1738 IETF (Internet Engineering Task Force). RFC 1738: Uniform Resource Locators (URL), ed. T. Berners-Lee, L. Masinter, and M. McCahill. 1994.

[URI2] IETF RFC1808 IETF (Internet Engineering Task Force). RFC 1808: Relative Uniform Resource Locators, ed. R. Fielding, 1995.

[VC] vCard Home Page, URL:http://www.imc.org/pdi

[W3C] The World Wide Web Consortium Home Page, URL:http://www.w3.org/

[WCOL] Web Collections using XML, Alex Hopmann, Editor, March 9, 1997. URL:http://www.w3.org/Member/9703/XMLsubmit.html

[WF] "The Warwick Framework:", A Container Architecture for Aggregating Sets of Metadata, C. Lagoze, C. A. Lynch, and R. Daniel, Jr., (June 21, 1996) URL:http://cs-tr.cs.cornell.edu:80/Dienst/UI/2.0/Describe/ncstrl.cornell/TR96-1593

[XML] The W3C XML Extensible Markup Language Working Group Home Page, URL:http://www.w3.org/XML/

[XMLDATA] XML-Data, Andrew Layman, Jean Paoli, Steve De Rose, and Henry S. Thompson, June 20, 1997, URL: http://www.microsoft.com/standards/xml/xmldata.htm

---

---

hdl:cnri.dlib/may98-miller