Society of Archivists Data Standards Group

A Guide to Archival and Related Standards
Standards applicable to archives; for the digital delivery of repository guides, finding-aids, and images of material from collections.

Title: Resource Description Framework (RDF)

Name of Standards Developing Organisation: W3C


Replaces: n/a

Abstract
The Resource Description Framework (RDF) is a model for representing information about resources on the Web in a way that is processable by machines and that facilitates data sharing. It is the standard for encoding data for the Semantic Web.

Description
The Web is a network of interconnected information that is organised primarily for human consumption. RDF is a way of making meaningful statements about resources and it is intended to convey information in a way that enables computers to process that information in a useful way. The idea is that any real world or abstract concepts can be represented on the Web using RDF. It is a relatively simple method to express information which then allows computer applications to do useful things with the information. Research can benefit from RDF because it is possible to make new discoveries that no single document has asserted.

The RDF statements are made in the form of ‘triples’, with an entity being represented by a subject, a predicate and an object. The subject is the resource, the predicate defines traits or aspects of the resource and expresses the relationship between the subject and the object. A statement such as John Ruskin was born in London consists of the subject, ‘John Ruskin’, the predicate ‘was born in’ and the object ‘London’.

Subjects, predicates and objects are names for concrete or abstract entities in the real world. There are standards built on RDF that describe how logical inferences may be made between facts. RDF is currently seen by many as the best way to represent knowledge in a distributed environment without substantial loss of meaning. The idea is that applications can bring together the data in RDF files by means of the links between the files. By naming not only the subject, but also the object and the relationship between the subject and object, it is possible to bring a higher level of meaning to data, to mix and share data across different applications. The information may be used by applications
RDF does not specify a vocabulary of descriptive properties, such as ‘author’, ‘title’, ‘location’. Instead, it specifies ways to name and describe properties that could include author, title etc. RDF is based on the idea of identifying things using identifiers called *Uniform Resource Identifiers (URIs)* which are then made available on the Web as *http URIs*. URLs (Uniform Resource Locators, which can be thought of as Web addresses) are a type of URI, and, in RDF, URIs commonly appear in the form of URLs. Typically, http URIs can be ‘dereferenced’, which means that they are ‘actionable’ via Web applications such as Web browsers which can access actual data on the Web, but this is not strictly necessary. Whatever their form, the important thing to understand is that URIs are just names for entities, even if they may appear a little different to what we normally think of as names. It is important to understand that RDF can be used to represent information about things that can be identified on the Web, even when they cannot be directly retrieved on the Web. For example, I can be represented with the URI http://www.archiveshub.ac.uk/janefoaf.rdf. This is my foaf file, which is an XML representation of RDF used for describing individuals. This URI can be used when referencing me in an RDF statement. RDF is not tied to a particular format. It is commonly expressed as XML, but it is important not to see RDF as being an XML specification.

RDF is an important component of the Semantic Web (http://semanticweb.org). In the Semantic Web the meaning of information and services is defined, allowing computers to ‘understand’ the information and make use of it. The idea is to move away from a Web of documents and towards a Web of knowledge. Linked Data (http://linkeddata.org/) is increasingly being seen as an integral part of the Semantic Web because it is essentially about linking up the information so that a person, or machine, can explore the web of data. Linked Data requires the use of URIs to identify things, and many people argue that it also requires the use of RDF to provide meaningful machine-processable information about those things.

In many ways, the importance of RDF lies in its ability to help us publish data on the Web in a way that is effective. Up till now, we have tended to think in terms of publishing documents on the Web, which comprise headings, paragraphs, lists, etc. Data is about things (entities) that have properties that have values.

RDFa is an additional W3C recommendation for expressing RDF data within XHTML (HTML that is XML compliant). It introduces the use of XHTML attributes (hence the ‘a’ in RDFa) to provide machine-readable indicators for browsers and for other programs to interpret. It is really about using attributes to provide metadata for machines that define the properties of links, or, put another way, that define relationships between web pages that are linked together. For example, an attribute can be used to define a link as being about the use of a licence. Whilst a human can click on a link and view the licence, a machine requires some additional markup within the XHTML to enable it to interpret the link. Similarly, to define a person’s name, an attribute can be used to declare that this is a person. The vocabulary used needs to be specified, and then a machine can process information as being a ‘title’ or an ‘author’ or a link to a ‘licence’.

RDF is increasingly being used to create richer data for the Web, but it is a gradual process and there are huge challenges involved. At present it is relatively difficult to publish RDF data and we are still working out how to make the best use of it.
Next month

Author name (replace text)
Author affiliation (replace text)