



KUNGL
TEKNISKA
HÖGSKOLAN



CID-157 • ISSN 1403-0721 • Department of Numerical Analysis and Computer Science • KTH

The semantic web: How RDF will change learning technology standards

Mikael Nilsson



CID, CENTRE FOR USER ORIENTED IT DESIGN

Mikael Nilsson

The semantic web: How RDF will change learning technology standards

Report number: CID-157

ISSN number: ISSN 1403 - 0721 (print) 1403 - 073 X (Web/PDF)

Publication date: September 2001

E-mail of author: mini@nada.kth.se

Reports can be ordered from:

CID, Centre for User Oriented IT Design

NADA, Department of Numerical Analysis and Computer Science

KTH (Royal Institute of Technology)

SE- 100 44 Stockholm, Sweden

Telephone: + 46 (0)8 790 91 00

Fax: + 46 (0)8 790 90 99

E-mail: cid@nada.kth.se

URL: <http://cid.nada.kth.se>

The semantic web: How RDF will change learning technology standards

Mikael Nilsson, Center for User-Oriented IT-design, Royal Institute of Technology, Stockholm
September 27, 2001

The field of learning technology has entered a phase of intense work on standardization of learning technology descriptions of various kinds.

Most of the work so far has focused on XML as the encoding language for such specifications (e.g. IMS, IEEE-LOM, and SCORM). However, the World Wide Web Consortium (W3C) is putting their energy into another model for computerized descriptions, called Resource Description Framework, RDF, which is the foundation for the Semantic Web vision of Tim Berners-Lee¹.

This raises important question regarding the future of learning technologies: In what way might RDF be useful for learning technology specifications? In what sense does RDF represent the future of meta-data, and how does this affect learning technology?

It turns out that the answers to these questions requires revising some of the fundamental assumptions in the learning technology field.

RDF origins: The Semantic Web vision

The Semantic Web is the name of a long-term project recently started by W3C with the stated purpose of realizing the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.²

It was motivated by the very same problems that motivates the development of meta-data standards³ - the fact that raw media, in the form of text, HTML, images or video streams, contains meta-information that may be readily deducible from the context for the human consumer (the name of the author, the kind of material contained within, etc.), but is mostly inaccessible to computers.

Making this information available to computers in order to enhance their usefulness, was the driving vision that created the Semantic Web project.

Most traditional meta-data approaches take the view of meta-data as being mostly a digital indexing scheme to use in cataloging and digital libraries. What distinguishes the Semantic Web from these approaches to meta-data are two important things:

- The Semantic Web is **designed to allow reasoning and inference capabilities** to be added to the pure descriptions. In its simplest form, this includes stating facts such as "a hex-head bolt is a type of machine bolt"⁴, but extends to the deduction of complicated relationships. This is an important feature to allow intelligent agents and other software to not only passively swallow descriptions, but to act on them as well.
- The Semantic Web is a **web-technology that lives on top of the existing web**, by adding machine-readable information without modifying the existing Web. It is designed to be globally distributed with all this means in terms of scalability and flexibility.

The Semantic Web is a layered structure. XML forms the basis, being the transport syntax. RDF provides the information representation framework, and on top of this layer, schemas and ontologies provide the logical apparatus necessary for the expression of vocabularies, enabling intelligent processing of information.

Essential RDF

While the current Web allows you to *link* to anything from anything in a machine-understandable way, the Semantic Web will allow you to *say* anything about anything in a machine-readable way.

Seen this way, RDF is the language in which Semantic Web meta-data statements are expressed. In fact, RDF can be said to consist purely of so-called *statements*.

An RDF statement consists of three elements:

- a *subject*,

- a *predicate*, and
- an *object*.

Statements are about Web resources, so subjects and objects are URIs, machine-readable identifiers.

Objects can also be plain text strings. Saying

The document "http://www.w3c.org/2001/sw/ was created by W3C"

is represented by the triple:

("http://www.w3c.org/2001/sw/", **created by**, "W3C")

To disambiguate the different predicates that can be used, every predicate must be given a URI. In this case, there is a standard predicate available in the Dublin Core vocabulary, namely "http://purl.org/dc/elements/creator", which we can use. The triple then becomes

("http://www.w3c.org/2001/sw/", "http://purl.org/dc/elements/creator", "W3C")

This demonstrates that URIs can be used to name not only concrete digital documents on the web, but abstract entities as well. In order to talk about non-digital resources, we must give them URIs. For example, to talk about the organization W3C (i.e., use it the subject of a statement), we must give it a URI. Let's give it the URI

"http://www.w3c.org/organization".

We can now say things such as ...

"http://www.w3c.org/2001/sw/ was created by http://www.w3c.org/organization, which is an organization with the name 'W3C'"

... which is, in fact, three separate statements.

More complicated RDF expressions like this are usually represented as graphs, where the subjects and objects are nodes, and the predicates are edges⁵.

This is all there is to basic RDF - nodes-and-arcs diagrams interpreted as statements about concepts or digital resources represented by URIs⁶. However, the need for standardized vocabularies for things like "organization" and the predicate "is a" is evident. The basis for such vocabularies in RDF is **RDF Schema**⁷.

This specification provides the basic vocabulary to express relationships between terms: resources being **instances** of terms ("http://www.w3c.org/organization is an organization"), terms being **subterms** of other terms ("a hex-head bolt is a type of machine bolt") and so on.

It also provides means to restrict the usage of predicates: "is a parent of" only **applies to** persons, etc. The terms **instance**, **subterm**, **applies to** are the kind of terms defined by the RDF Schema specification.

Using the vocabulary provided by RDF Schema, it is easy to create your own semantically rich vocabularies.

Describing resources using RDF

It is not immediately obvious that the simple statement model of RDF can be used to make the Semantic Web a reality. The most fundamental benefit of RDF compared to other meta-data approaches is that using RDF, you can say **anything about anything**. Anyone can make RDF statements about any identifiable resource. Using RDF, the problems of extending meta-data and combining meta-data of different formats, from different schemas disappear, as RDF does not use closed documents.

Important uses of RDF to encode information for any resource you can name with a URI include:

describe

Since a resource can have uses outside the domain foreseen by the author, any given description (meta-data

instance) is bound to be incomplete. Because of the distributed nature of RDF, a description can be expanded, or new descriptions, following new formats (schemas), can be added. This allows for new creative uses of content in unforeseen ways. This is one of the important features of the current Web, where anyone can link to anything, that has been carried over into RDF.

certify

There is no reason why only big organizations should be able to certify content - individuals such as teachers may want to certify a certain content as a quality learning resource that is well suited for specific learning tasks. How to handle this kind of certification will be an important part of the Semantic Web.

annotate

Everything that has an identifier can be annotated. There are already attempts in this direction: Annotea⁸ is a project where annotations are created locally or on a server in RDF format. The annotations apply to HTML or XML documents and are automatically fetched and incorporated into web pages via a special feature in the experimental browser Amaya⁹.

extend

Structured content (typically in XML format) will become common. Successive editing can be done via special RDF-schemas allowing private, group consensus or author-specific versions of a common base document. The versioning history will be a tree with known and unknown branches which can be traversed with the help of the next generation versioning tools.

reuse

RDF is application independent. As the meta-data is expressed in a standard format independent of more advanced schemas that are used, even simplistic applications can understand parts of large RDF descriptions. If more advanced processing software is available (such as logic engines), more advanced treatment of the RDF descriptions is possible.

RDF and XML: Model, Syntax, Semantics

So far nothing has been said about XML. The reason is that the **RDF Model** can be defined completely without reference to XML. XML can, however, be used as a **syntax** for RDF statements. The RDF specification defines the standard syntax to encode RDF statements in XML.

But one question remains: why can't XML and XML Schema be used to represent the same kind of information that RDF expresses? XML Schemas are, after all, powerful tools to express complex requirements on XML elements. This is true, and XML and XML Schemas can be used to do some of what RDF does, but not without much trouble. The reasons are several:

- **The RDF model and the XML model are fundamentally different.** The XML data model is a text-markup oriented labeled tree. RDF, by contrast, has a very simple model consisting of labeled arcs. Of course, any specific set of RDF statements forms a graph that can be serialized in XML. But as XML and XML Schema are designed primarily for fixed, tree-like documents, they are significantly less flexible for expressing meta-data, which by its very nature is subjective, distributed and expressed in diverse forms. The RDF model, while simpler, is flexible enough to support these principles.
- **The resources used in RDF and XML Schemas are fundamentally different.** The nodes that XML Schemas talk about are nodes in an XML document, at specific places in a document structure. In RDF, the nodes are not nodes in the document itself, but rather any resources that have URIs, and more often than not live outside the RDF document itself. Thus, RDF is designed to be a meta-data language.
- **The semantics of XML Schemas and RDF are fundamentally different.** XML Schemas have a primarily syntactic interpretation, restricting the set of XML documents that can be produced. RDF, on the other hand, has a primarily semantic interpretation. While XML Schemas are used for modeling XML documents, RDF is used to model knowledge, where tree-based representations are not enough.

The difference can be formulated in this way¹⁰: XML/XML Schema is a data modeling language, and RDF is a

meta-data modeling language. When meta-data needs to be encoded as data, an XML syntax is very useful. However, *modeling* meta-data in pure XML severely restricts its flexibility.

Lessons from Using RDF in the IMS Specifications

In May 2001, IMS released version 1.2 of its meta-data specification. This major release - marking an important step forward for the specification and the underlying IEEE LOM standard - was informed by lessons learned since IMS first published its meta-data specification. But this release also had a new twist: A development effort, lead by myself, to produce an RDF binding.

The development of this specification was an important testbed for RDF-based meta-data descriptions in learning technologies, as were subsequent efforts at producing an IMS Content Packaging RDF binding.

Some of the important positive lessons we learned in this effort were:

- Interoperability with other, separate, standards is greatly increased. The reason is simple: RDF allows a **single storage model** for very different types of data and schemas. For example, storing meta-data from different specifications in the same database is straightforward. To implement searching that includes dependencies between meta-data expressed in different schemas is simplified. An example of this is the Edutella¹¹ effort to build a peer-to-peer educational meta-data exchange network, which would meet severe difficulties in using, searching in, and translating between the different formats used for VCard, Dublin Core, Dublin Core Qualifiers, IEEE LOM, SCORM etc., which are meta-data standards without the common RDF format.
- **Reuse** of existing meta-data standards is greatly simplified. For example, there has been much discussion on whether to incorporate the VCard XML syntax in the XML binding. While desirable, this creates namespacing and XML DTD problems. In the RDF binding, the VCard RDF binding can be transparently included with **no extra effort**.
- Some terms do not have exact equivalents in other meta-data standards, but relate to some existing terms by, for example, being more narrow, more broad etc. As the whole IMS RDF binding was designed as an extension to Dublin Core meta-data¹², the relationship between IMS meta-data elements and Dublin Core elements are formalized **in a machine-readable manner**. Thus, no conversion to or from Dublin Core meta-data is needed, and Dublin Core aware tools can understand the Dublin Core-parts of an IMS meta-data description.
- There has also been much discussion on the topic of **vocabularies**. While these are essential for the use and extension of IEEE LOM meta-data, there does not exist a standard way to encode and distribute them. In RDF, this problem completely disappears, as vocabularies, as we have seen, are a fundamental part of the RDF Schema specification. Not only is there a standard way to list vocabulary items, but their interdependencies can be modeled in a standard way. And as if that was not enough, efforts such as DAML¹³ and OIL¹⁴ provide means to model vocabularies as full-fledged **ontologies**¹⁵ expressed in RDF, if that is desired, while still maintaining compatibility with less-capable software.
- While extending the XML binding is certainly possible using XML Schemas, the process easily creates interoperability problems. In RDF, several independent means of meaningful extensions are available, none of which cause interoperability problems:
 - **Refinement** of the semantics of existing properties and terms by creating subterms etc. This cannot be done in a standard way using XML.
 - **Introduction of new properties** and terms describing resources. This is the kind of extension one can usually do in XML.
 - **Adding new properties** to a resource in other documents, which is possible since RDF does not work with meta-data instances as closed documents. For example, the RDF binding is designed so that translations of titles and description etc., can be managed separately. In the same way, different kinds of meta-data can be managed separately, and merged when needed¹⁶. This **modularity** is impossible to achieve in a clean and standard way using XML.

These possibilities are not only nice properties of RDF, but are completely indispensable in many cases.

- RDF already contains means for describing **meta-meta-data** (in any number of meta-steps)¹⁷, that can be as rich as ordinary meta-data.
- For meta-data that contain very complex interdependencies, such as IMS Content Packaging, the graph representation and modularity of RDF effectively **cleans up the format** and semantics of the specification.
- RDF allows for a clean integration of the different specifications in a layered way. Currently, the work on IMS Content Packaging in RDF is built on top of the IMS Meta-data RDF binding, which is built on top of the VCard RDF binding and the Dublin Core Qualifiers RDF binding, that extends the core Dublin Core RDF binding. Continuing upwards in this fashion, I see a very concrete potential for a **complete unification** of all IMS meta-data related specifications¹⁸. This is no light-minded suggestion, and the benefits, as seen in this list, are many.

These lessons are in no way coincidences. While XML was designed as a data interchange format, RDF was designed from the ground up to fulfil the role of an Internet architecture for meta-data. "**Resource Description Framework (RDF) is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web**"¹⁹. This is very clearly reflected in the findings above.

On another note, RDF presents several drawbacks, as has been made clear during the work on the RDF bindings.

- The underlying standards, notably the RDF Schema specification, but also the Dublin Core Qualifiers RDF binding, are still young, intensively discussed and possibly **subject to change**. The specifications underlying the XML binding are much more stable (even if the XML Schema specifications have changed recently). This is, of course, a temporary problem.
- **Tool support for RDF is very immature at this point**, and integration of Semantic Web technologies into the current Web is still only starting. XML support can be said to be mature in most respects. However, with the current pace of RDF adoption, tool support is rapidly increasing.
- Designing an RDF binding makes it necessary to **revisit many of the assumptions in the underlying information model**, which often is designed with an XML binding in mind. As the semantics of XML elements is not explicitly stated, much of the work in designing an RDF binding goes into defining the semantics of the elements.

This has caused minor interoperability problems between the XML binding and the RDF binding, problems that can only be remedied by designing the information model with the RDF binding in mind.

From another perspective, this is a very positive side effect, as it significantly helps sharpen the information model. This has already been observed in the design of the IMS meta-data RDF binding, but from preliminary studies seems to be even more evident in the work on an RDF binding of IMS Content Packaging.

Possibilities on the Semantic Web

From a more strategic point of view, the emerging Semantic Web presents exciting new possibilities for uses of learning technology specifications. While XML standards are very good tools for enabling interoperability by specifying import and export formats for LMSs²⁰, they tend to favor large, centrally managed, monolithic systems.

By enabling the use of learning technology specifications in Semantic Web technologies, a much wider range of applications are imaginable:

- **Intelligent software agents can be implemented**, helping the learner to find and use globally distributed learning resources.
- **Personal annotations** of any learning resource becomes a feasible technology, as demonstrated by Annotea²¹.
- **Collaborative and distributed authoring and course construction** becomes much simpler thanks to the

modularity of the information.

- **Reuse of learning material** by cross-fertilization suddenly becomes a reality, creating important synergy effects.

The Semantic Web promises to create a web-based eco-system for learning resources, freeing the material from being trapped in closed systems. One important example of this kind of technology is **Edutella**²², an RDF-based peer-to-peer system under development, being designed to allow distributed access to learning resource meta-data expressed in many different schemas. By combining meta-data from many sources in a controlled but distributed way, cross-annotation and mutual reuse of material becomes a reality.

In short, the vision of the Semantic Web is an important vision for online learning as well.

Conclusions

As has been pointed out above, the specifications produced by the different parties in the meta-data community (Dublin Core, VCard, LOM, SCORM, etc.) all have complex interdependencies, and increasingly so.

It has become evident that the reuse of vocabulary and technology between those specifications is a very difficult task, and has only been successful in some simple cases.

It should be clear that this is not always due to a fundamental interoperability problem - more often than not, it is a result of the closed nature of XML bindings. RDF bindings, on the other hand, are in general much more intensely interdependent as a consequence of the heavy reuse of vocabularies.

Using RDF in learning technology specifications would also greatly reduce the difficulties involved in adding another specification to an existing software system - it would only involve adding a new RDF Schema, which has been designed to interoperate with the existing schemas. This has important consequences for the adoption of learning technologies.

The future for RDF for learning technology specifications is bright, and the possibilities opened up by RDF and Semantic Web technologies promise to take learning technology project to a new level of applications.

But some effort from the learning technology specification community to produce the necessary specifications will be needed. In particular interoperability discussions between standards groups need to be intense and technical in order to maximize interoperation.

RDF provides an important technological platform to handle the interoperability demands of the emerging specification and vocabulary jungle.

Notes and useful links:

- Our thanks to Mikael Nilsson for submitting this feature. Mikael can be contacted by email at mini@nada.kth.se. (see our publisher's statement)
- IMS Global Learning Consortium
- IEEE Learning Object Metadata
- ADL SCORM
- W3C
- RDF page at the W3C website
- W3C Semantic Web site

1. ... Berners-Lee (up) <http://www.w3.org/2001/sw/>
2. ... (up) W3C Semantic Web Activity Statement at <http://www.w3.org/2001/sw/Activity>, part of the W3C Semantic Web site
3. ... standards (up) The Semantic Web activity is, as a matter of fact, the successor of the W3C Metadata activity. Thus, the Semantic Web is the W3C meta-data architecture.
4. ... (up) Taken from Tim Berners-Lee, James Hendler, Ora Lassila: *The Semantic Web* published in Scientific American, May 2001.
5. ... edges (up) For a more detailed explanation of these concepts, see e.g. An Introduction of RDF by Eric Miller or the RDF syntax and model specification.

6. ... URIs (up) RDF also contains an important mechanism called *reification*, that allows you to state something about another RDF statement, such as who said it, whether it is true or false, etc.
7. ... Schema (up) Defined in <http://www.w3.org/TR/rdf-schema>
8. ... Annotea (up) See the Annotea website
9. ... Amaya (up) See the Amaya website
10. ... way (up) See also <http://www.w3.org/DesignIssues/RDF-XML.html> or in more detail: www.ontoknowledge.org/oil/down/IEEE00.pdf
11. ... Edutella (up) See the Edutella website.
12. ... meta-data (up) See the Dublin Core website
13. ... DAML (up) See the DAML website
14. ... OIL (up) See the OIL pages at Ontoknowledge.org
15. ... ontologies (up) See e.g. <http://www.ontology.org/main/papers/faq.html> for a definition of the word "ontology" in this context. The paper *Combining Ontologies and Terminologies in Information Systems* by Johann Gamper, Wolfgang Nejdl and Martin Wolpers is also of interest.
16. ... needed (up) For an important example of this, see the UNIVERSAL project (<http://nm.wu-wien.ac.at/universal/>, and <http://www.ist-universal.org/>).
17. ... (up) Via *reification*, mentioned above.
18. ... specifications (up) Something similar has already been done in the UNIVERSAL project, where e.g. learner and contributor information is added on top of low-level meta-data.
19. ... Web'' (up) RDF specification: <http://www.w3.org/TR/REC-rdf-syntax>
20. ... LMSs (up) Learning Management Systems - a software system for managing an interactive learning environment.
21. ... Annotea (up) See the Annotea website
22. ... Edutella (up) See the Edutella website.