

Synthesis on RDF/RDFS 2

EIT Digital Master in Data Science

Marcos Bernal España – marcosbernal3@gmail.com

Yolanda de la Hoz Simón - yolanda93h@gmail.com

Index

What are the main principles of this model?..... 3

What are good modelisation practices? 4

References..... 5

What are the main principles of this model?

RDF is a standard model which main objective is the data interchange over web for multiple machines. This model has different features that facilitate data merging and interchanging even if the underlying schemas differ, it specifically supports the evolution of schemas over time without requiring specific changes in consumer data.

RDF is the response to the existing problem with data embedded in html pages. This data are design to be understood by humans, this means that most of the data available in a web site is present in a non-structure way (natural language, images, videos, etc) and also this data is embbded in a html page whose design and structure it's changing dinamicly over time. This makes more difficult to detect changes and share or process data by multiple machines.

The main objctives of this model are the following:

- It allows structured and semi-structured data to be mixed, exposed, and shared across different applications.
- Facilitate data merging even if the underlying schemas differ.
- RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple")

Tim Berners-Lee outlined four principles of linked data in his "Linked Data" note of 2006, [1] paraphrased along the following lines:

1. Use URIs to name (identify) things.
2. Use HTTP URIs so that these things can be looked up (interpreted, "dereferenced").
3. Provide useful information about what a name identifies when it's looked up, using open standards such as RDF, SPARQL, etc.
4. Refer to other things using their HTTP URI-based names when publishing data on the Web.

Tim Berners-Lee gave a presentation on linked data at the TED 2009 conference.[2] In it, he restated the linked data principles as three "extremely simple" rules:

1. All kinds of conceptual things, they have names now that start with HTTP.
2. If I take one of these HTTP names and I look it up...I will get back some data in a standard format which is kind of useful data that somebody might like to know about that thing, about that event.
3. When I get back that information it's not just got somebody's height and weight and when they were born, its got relationships. And when it has relationships, whenever it expresses a relationship then the other thing that it's related to is given one of those names that starts with HTTP.

What are good modelisation practices?

In the modelisation steps we should have into account the following considerations, before starting, in order to reduce time and also design a reusable RDF according with the existing data.

1. Start with a robust domain model developed following a structured process and methodology.
2. Research existing terms and their usage and maximise reuse of those terms.
3. Where new terms can be seen as specialisations of existing terms, create subclass and subproperties.
4. Where new terms are required, create them following commonly agreed best practice.
5. Publish within a highly stable environment designed to be persistent
6. Publicise the RDF schema by registering it with relevant services

In the modelisation process, it is recommended to follow the best practices commented in *Best Practices for Publishing Linked Data*. The best practices discussed in this document are summarized as the following steps:

1. PREPARE STAKEHOLDERS:
Prepare stakeholders by explaining the process of creating and maintaining Linked Open Data.
2. SELECT A DATASET:
Select a dataset that provides benefit to others for reuse.
3. MODEL THE DATA:
Modeling Linked Data involves representing data objects and how they are related in an application-independent way.
4. SPECIFY AN APPROPRIATE LICENSE:
Specify an appropriate open data license. Data reuse is more likely to occur when there is a clear statement about the origin, ownership and terms related to the use of the published data.
5. GOOD URIs FOR LINKED DATA:
The core of Linked Data is a well-considered URI naming strategy and implementation plan, based on HTTP URIs. Consideration for naming objects, multilingual support, data change over time and persistence strategy are the building blocks for useful Linked Data.
6. USE STANDARD VOCABULARIES:
Describe objects with previously defined vocabularies whenever possible. Extend standard vocabularies where necessary, and create vocabularies (only when required) that follow best practices whenever possible.
7. CONVERT DATA:
Convert data to a Linked Data representation. This is typically done by script or other automated processes.
8. PROVIDE MACHINE ACCESS TO DATA:
Provide various ways for search engines and other automated processes to access data using standard Web mechanisms.
9. ANNOUNCE NEW DATA SETS:
Remember to announce new data sets on an authoritative domain. Importantly, remember that as a Linked Open Data publisher, an implicit social contract is in effect.
10. RECOGNIZE THE SOCIAL CONTRACT:
Recognize your responsibility in maintaining data once it is published. Ensure that the

dataset(s) remain available where your organization says it will be and is maintained over time.

Finally, for the publication of the vocabulary or ontology on the Web it's recommended to choose a recipe from *Best Practice Recipes for Publishing RDF vocabularies* [4]. This recipes introduces general principles and an example configuration for use with an Apache HTTP server.

References

- [1] Tim Berners-Lee (2006-07-27). "Linked Data". Design Issues. W3C. Retrieved 2010-12-18.
- [2] Jump up^ "Tim Berners-Lee on the next Web".
- [3] Best Practices for Publishing Linked Data, W3C. Bernadette Hyland. Ghislain Ateazing, Boris Villazón-Terrazas
- [4] Best Practice Recipes for Publishing RDF Vocabularies, W3C. Diego Berrueta, Jon Phipps.