RDF Next Version

Ivan Herman and Sandro Hawke
W3C

History

- Current RDF has been published in 2004
- Significant deployment since then
  - implementation experiences
  - users' experiences
- Some cracks, missing functionalities, etc, came to the fore
- There are significant communities that have not picked up RDF
  - e.g., Web Developers

The question

- Shall we
  - live with those issues and go on with our lives?
  - dump it and start all over again from scratch?
  - do some minimal changes?

The W3C "RDF Next Step" Workshop

- W3C organized a Workshop in June 2010
  - 32 submissions, 28 accepted for publication, 18 were presented at the workshop
  - 2 busy days at Stanford (courtesy of NCBO)

©2003 Robert Canfield Photography
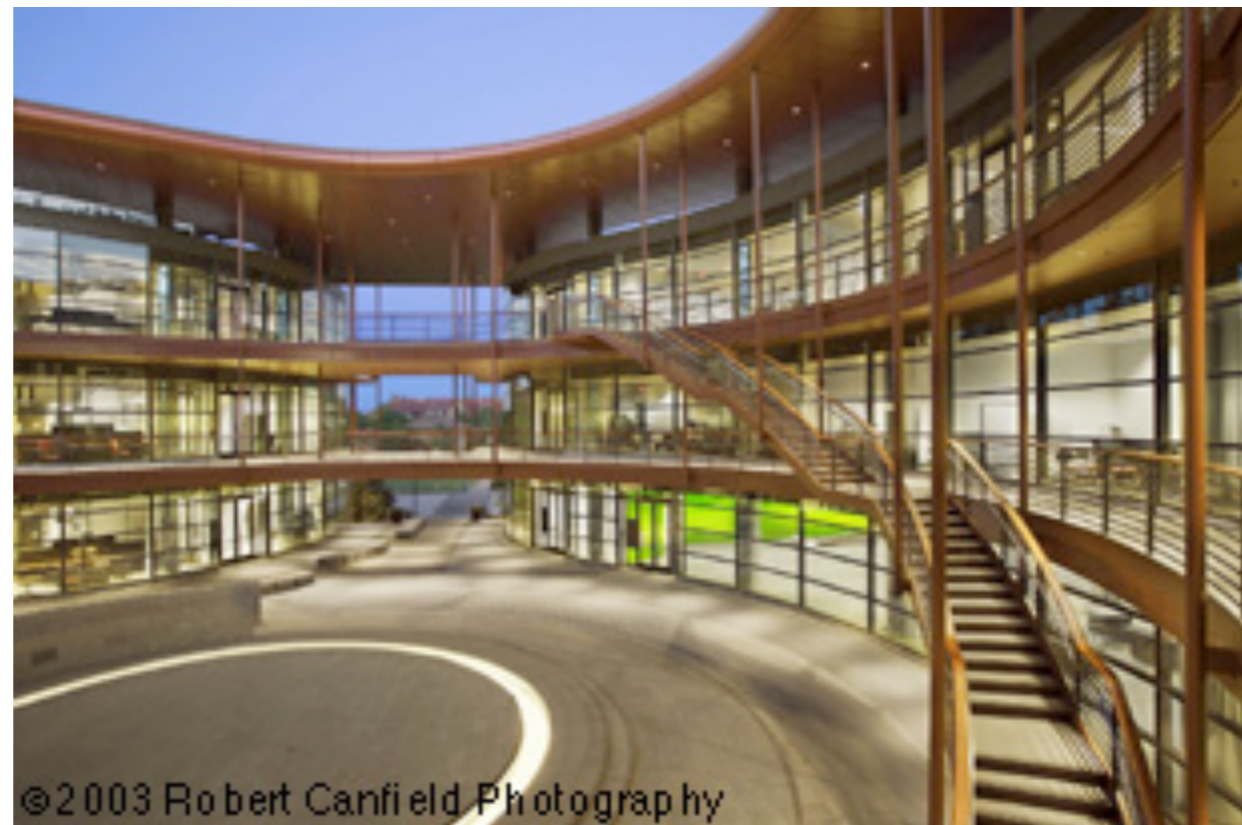
Workshop participants

What we did…

- Try to answer the question: live with it, redo it, mend it…
  - if something has to be changed, what is it and with what priority?
- Give a list of possible work items, with priorities

The general feeling…

- Yes, it is probably o.k. to touch some issues
- But we have to be *very* careful not to send the wrong signal to adopters, tool providers, etc.
- *I.e.: keep the changes to the minimum*

The Workshop straw poll result

Proposed Work Item	WG must do it	WG should/may do it	WG must not do it
Standardize Model for Graph Identification	✔✔✔✔✔✔✔✔✔✔✔✔✔✔	✔✔✔	
Modify Semantics to Support Graph Identification	✔✔✔✔✔✔✔✔✔✔✔✔✔✔	✔✔	
Switch to Improved Inference Rules	✔✔✔✔✔✔✔✔✔	✔✔✔✔✔✔✔	
Apply Fixes to Known Spec Errors	✔✔✔✔✔✔✔✔✔✔✔✔✔		
Standardize a JSON RDF Syntax	✔✔✔✔✔✔✔	✔✔✔✔✔✔✔✔✔✔	
Make Turtle a W3C Standard	✔✔✔✔✔✔✔✔✔✔✔	✔✔✔	
Add Graphs to Turtle	✔✔✔✔✔✔✔✔✔	✔✔✔✔	
Add Graphs to RDF/XML	✔✔✔✔✔	✔✔✔✔✔✔✔✔✔✔✔	✖✖
Revise Blank Node Semantics	✔	✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔	
Specifiy Linked Data Style of RDF	✔✔	✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔	✖
Weakly Deprecate some RDF/XML Features	✔✔✔	✔✔✔✔✔✔✔✔✔✔✔✔✔✔	
Define alternatives to owl:sameAs	✔✔✔✔	✔✔✔✔✔✔✔✔✔✔✔✔✔✔	✖✖✖
Weakly Deprecate some Data Model Feature	✔✔	✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔	✖
Namespace Profiles		✔✔✔✔✔✔✔✔✔✔✔✔✔✔	
Weakly Deprecate some RDF Semantics Features		✔✔✔✔✔✔✔✔✔✔✔✔	
Have Explicit Support for Annotations		✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔✔	✖✖✖
Align RDF Semantics with SPARQL	✔	✔✔✔✔✔✔✔✔✔✔✔✔	✖✖✖✖
Improve rdf:List Support in RDF/XML		✔✔✔✔✔✔✔✔✔✔✔	✖✖✖✖
Explain How to Determine What a URI Means		✔✔✔✔✔✔✔✔✔	✖✖✖✖✖✖
Allow Literals as Subjects		✔✔✔✔✔✔✔	✖✖✖✖✖✖✖✖✖✖✖✖

8

Follow up

- Workshop report published:
  - http://www.w3.org/2009/12/rdf-ws/Report.html
- W3C Team began working on chartering
- …but felt the larger community should be asked
- A questionnaire was published in August 2010
  - http://www.w3.org/2002/09/wbs/1/rdf-2010/results
- And, of course, lots of discussion on various fora

In what follows…

- We will list the major items that came up during the discussions
  - mostly the Workshop+Questionnaire, plus some others
- We will list them in the order of
  - may end up in the RDF Working Group Charter, if accepted by the W3C members
  - should be done but not clear yet where and how
  - postpone it for now

Before going into details…

- What about backward compatibility?
- The proposal is to stay backward compatible, i.e.:
  - any valid RDF graphs of today should remain valid in terms of a new version of RDF
  - any RDF or RDFS entailment drawn today should be valid entailment in terms of a new version of RDF and RDFS

"Charter candidates"
Part I: required features

The obvious issues

- There are some errata that have to be taken care of
  - exact relationship to IRI-s
  - more flexible references to XML versions
  - there is an error in the Entailment lemma (see H. ter Horst's mail)
  - etc
- Not worth discussing them here

"Graph identification"

- A.k.a. "named graphs", "quoted graphs", "knowledge bases"
- Is on the top of all priority lists…
  - there is a notion in SPARQL already
  - widely used in practice
  - other applications, like provenance, needs it

"Graph identification"

- But the semantics is not absolutely clear
  - e.g., are we talking about a mutable or immutable collection of triples?
  - maybe we have two different concepts here…

Turtle serialization syntax

- Another top priority request by the community
- We have
  - a stable "team submission" as a specification
  - SPARQL Query Language
  - N3
- But there is no *standard* reference
- Additional syntax should also be added for graph identification

JSON serialization syntax

- Is essential for Web Application; Javascript programmers
  - may do not want to have a different parser
  - may not *really* know RDF, nor do they care about all details
- The syntax may not be a complete syntax; to be decided as we go. E.g.,
  - e.g., no blank nodes, only syntax for Skolemized nodes
  - hide the difference between a URI-as-a-string and URI-as-identification
  - etc.

JSON serialization syntax

- We do not start from scratch
  - Ian Davis' RDF/JSON
  - Manu Sporny' JSON-LD
  - Sandro Hawke's musings
- The syntax may also include tools for lists, graph identification, etc.

A common theme for Turtle and JSON: profiles

- RDFa 1.1 has the notion of "profile files":
  - a separate file containing
    - prefix definitions
    - "term" definition, ie, mapping from a word to a URI
  - an RDFa file can refer to a profile file on the Web to "include" all those definitions
  - great help to hide many of the namespace and URI complexities
- Do we want to have such profiles in Turtle and/or JSON?
  - came up in the discussions
  - not explicitly part of the charter, the WG may decide if it wants to go there

Deprecation

- Some features may be deprecated: reification, containers, …
- This is planned to be a "weak" deprecation:
  - feature is not removed from the specification
  - no commitment that a future release of RDF would remove it
  - an explicit advice to the community *not* to use the given feature
  - maybe: a "sub" semantics for RDFS that makes use of the absence of containers (see later)

Reconcile semantics documents

- A number of semantics extensions and features have appeared in other W3C Recommendations
- These have their logical place in the RDF document
- Goal is *not* to do the work again; rather collect, maybe by reference, all in one place

Reconcile semantics documents: examples

- **rdf:plainLiteral**: plain literal with language tag, but as an explicit datatype
  - was needed by OWL 2 and RIF to make handling of literals clear
  - it should be, logically, part of the RDFS document
- **POWDER's bridge between URI-s and strings**:
  - wdrs:matchesregex and wdrs:notmatchesregex relate resources to regular expressions which that resource matches
  - is an extension of the core RDFS semantics

Reconcile semantics documents: examples

- SPARQL 1.1 defines a <u>"finite" version of RDF(S) semantics</u>
  - part of the SPARQL 1.1 entailment regimes' document
  - is a "sub-semantics" of RDF(S)
    - can be implemented by rule engines alone for example
  - a variant of the "ter Horst Semantics"
  - should be easily referencable for implementations in general

"Charter candidates"
Part II: time-permitting features

Harmonize plain literal management

- We currently have plain literal, xsd:string, rdf:plainLiteral…
  - it leads to, e.g., convoluted SPARQL queries
- These should be harmonized somehow on the semantic level

Update the RDF Primer

- Refresh the vocabulary examples being used
- Possibly with multi-syntax examples (like, e.g., the OWL 2 documents)
- Linked Data guidelines could be included
  - "follow your nose"
  - issues around the usage of owl:sameAs or others
  - etc.

"Work candidates"
(not yet clear where and how)

What does this mean?

- A number of issues have been identified that
  - clearly reflect necessary or highly useful work
  - it is not necessarily part of an RDF "Core" work…
  - … but should be done somewhere at some point
- Ideally, W3C should start a separate group to handle those
  - at this moment we do not know when and how that will be possible (for very practical reasons)

RDF API-s

- Mainly the Web Developers' community needs API-s
  - there is already an established set of API-s in Java, Python, etc
  - but not that much in Javascript
  - Web Developers are a different breed than, e.g., Jena users
- The RDFa Working Group develops an RDFa API
  - this API will have a "subset" that is, essentially, an RDF API
  - but some of the details will not be handled by them

Provenance

- Clearly a major problem to be solved
  - there are already a number of Provenance vocabularies, but no standard
  - touches upon general metadata management
- W3C may start a separate group on Provenance vocabularies
  - there is a W3C Incubator group that should publish a report soon and may have laid the groundwork for that

RDFS++/RDFS 3.0/OWLPrime

- Goal: define a subset of (essentially) OWL 2 RL
  - more palatable to developers
    - eg, no property restrictions, that Web developers do not really grasp anyway
    - include (in)equalities, characterization of properties, maybe class intersections, keys, and chains
  - can be "referenced" as an *entity*, not only a set of rules
- Should be done, not clear where and how
  - would a W3C Note or submission be enough, or does it need a full Recommendation status?

"Create standard for deployment of linked data"

- Issues that arose:
  - httpRange-14 as a standard
  - "Cool URI-s for the Semantic Web"
  - "follow your nose principles"
  - social contracts around URI-s
  - etc.
- Not clear that these should be Recommendations
- Some of the issues are listed as, possibly, part of a renewed RDF Primer

Similarity/equivalence properties

- Mainly on the Linked Data Cloud owl:sameAs is widely used
  - the usage is not necessarily semantically correct
  - a vocabulary should be defined to reflect the various usages
    - could be very close to the relevant SKOS terms, actually…
- Some elements may be part of an updated RDF Primer

No work planned

RDF/XML improvements

- RDF/XML is generally disliked
  - no one wants to spend time on improving it…
- New features (e.g., graph identification) may not find its way to RDF/XML
- The RDF Group may handle the reported bugs in the specification, but that is about it

Redo the RDF Semantics

- There is a disconnect between the formal, model-theoretic semantics of RDF(S) and applications
  - there are also theoretical inadequacies, too
- But the overall feedback was: don't touch it, deployment has learned to live with it, there is no consensus on what to replace it with, etc.

Remove RDF restrictions

- Blank nodes as predicates, literals as subject
- Literals as subject has deeply divided the community
  - some are violently against it, others ask for it
  - clearly no consensus at the moment!

And now?

- W3C team should finalize the charter soon
- Then the W3C process kicks in
  - AC members vote with a yea or nay
  - if the vote is positive: work can begin in early 2010

The planned staff

Co-Chair:
David Wood,
Talis, Washington DC

Co-Chair:
Guus Schreiber,
VU, Amsterdam

W3C Staff Contact
Sandro Hawke
W3C, Boston

W3C Staff Contact
Ivan Herman
W3C, Amsterdam

The planned timing

- The group should start in February 2011, end in January 2013
- First versions ("First Public Working Draft"): May 2011
- The technical work should be completed ("Last Call"): August 2012
- This is followed by implementation testing

Thank you for your attention

This slides are available on-line:
  http://www.w3.org/2010/Talks/1111-Shanghai-IH/

Shoot with the questions!