Available online at www.sciencedirect.com
SCIENCE DIRECT®
Web Semantics: Science, Services and Agents on the World Wide Web 3 (2005) 247–267
JOURNAL OF Web Semantics
www.elsevier.com/locate/websem

# Named graphs

Jeremy J. Carroll[a], Christian Bizer[b,*], Pat Hayes[c], Patrick Stickler[d]

[a] *Hewlett-Packard Labs, Bristol, UK*
[b] *Freie Universität Berlin, Germany*
[c] *IHMC, FL, USA*
[d] *Nokia, Finland*

Received 15 August 2005; accepted 8 September 2005

**Abstract**

The Semantic Web consists of many RDF graphs nameable by URIs. This paper extends the syntax and semantics of RDF to cover such named graphs. This enables RDF statements that describe graphs, which is beneficial in many Semantic Web application areas. Named graphs are given an abstract syntax, a formal semantics, an XML syntax, and a syntax based on N3. SPARQL is a query language applicable to named graphs. A specific application area discussed in detail is that of describing provenance information. This paper provides a formally defined framework suited to being a foundation for the Semantic Web trust layer.


*Keywords:* Semantic Web; RDF; SPARQL; Provenance; Digital signatures

## 1. Introduction

A simplified view of the Semantic Web is a collection of web retrievable RDF documents, each containing an RDF graph. The RDF Recommendation [1–4] explains the meaning of any one graph, and how to merge a set of graphs into one, but does not provide suitable mechanisms for talking about graphs or relations between graphs. The ability to express metainformation about graphs is required for:

**Data syndication**: Systems need to keep track of provenance information and provenance chains.

**Restricting information usage**: information providers might want to attach information about intellectual property rights or their privacy preferences to graphs in order to restrict the usage of published information [5,6].

**Access control**: a triple store may wish to allow fine-grain access control, which appears as metadata concerning the graphs in the store [7].

**Signing RDF graphs**: as discussed in [8], it is often necessary to keep the graph that has been signed distinct from the signature, and other metadata concerning the signing, which may be kept in a second graph.

* Corresponding author. Tel.: +49 30 838 54057; fax: +49 30 838 52027.
*E-mail address:* chris@bizer.de (C. Bizer).

1570-8268/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.websem.2005.09.001

**Expressing propositional attitudes:** such as modalities and beliefs [9].

**Scoping assertions and logic:** where logical relationships between graphs have to be captured [10–12].

**Ontology versioning and evolution:** OWL [13] provides various properties to express metadata about ontologies. In OWL Full, these ontologies are RDF graphs. Ontology versioning and evolution is discussed in [14,15].

RDF reification has well-known problems in addressing these use cases as previously discussed in [16]. To avoid these problems the use of quads has been proposed by several authors [7,17–19]. These consist of an RDF triple and a further URIref or blank node or ID. The proposals vary widely in the semantics of the fourth element, using it to refer to information sources, to model IDs or statement IDs or more generally to 'contexts'.

We propose a general and simple variation on RDF, called *named graphs*. A named graph is an RDF graph which is assigned a name in the form of a URIref. The name of a graph may occur either in the graph itself, in other graphs, or not at all. Graphs may share URIrefs but not blank nodes.

Named graphs can be seen as a reformulation of quads in which the fourth element's distinct syntactic and semantic properties are clearly distinguished, and the relationship to RDF's triples, abstract syntax and semantics is clearer.

Named graphs are a deliberately small step on top of the RDF and OWL Recommendations. This allows their use with tools built as implementing those recommendations, in a backward compatible way, with little or no code modifications.

The first half of the paper covers: the abstract syntax and semantics of named graphs; their relationship with RDF, OWL, TRIPLE, Guha's contexts and SPARQL RDF dataset. We then discuss the TriX, TriG and RDF/XML syntaxes for named graphs and the query language SPARQL.

The second half describes how named graphs can be used for Semantic Web publishing, looking in particular on provenance tracking and how it interacts with the choices consumers of Semantic Web information make about which information to trust. We provide a vocabulary for Semantic Web publishing with its formal semantics. The vocabulary includes support for digital signatures and addresses performative acts, such as asserting RDF.

This paper is an extended version of the paper presented at the World Wide Web Conference (WWW 2005) [20].

## 2. Abstract syntax

RDF syntax is based on a mathematical abstraction: an RDF graph is defined as a set of triples. These graphs are stored in documents which can be retrieved from URIs on the Web. Often these URIs are also used as a name for the graph, for example with an `owl:imports`. To avoid confusion between these two usages we distinguish between named graphs and the RDF graph that the named graph encodes or represents. A named graph is an entity with two functions *name* and *rdfgraph* defined on it which determine respectively its name, which is a URI, and the RDF graph that it encodes or represents. These functions assign a unique name and RDF graph to each named graph. In this way, a named graph is a resource, identified by its name, and so it can be described in the usual open way using RDF.

More formally, let $U$ be the set of all URI references, $B$ an infinite set of RDF blank nodes, and $L$ the set of all legal RDF literals (all three sets as defined in [4]); $U$, $B$ and $L$ are pairwise disjoint; let $V = U \cup B \cup L$ be the set of *nodes*; then the set $T = V \times U \times V$ is the set of all RDF triples.[1] The set of RDF graphs $G$ is the power set of $T$. A named graph is a pair $\mathrm{ng} = (n, g)$ with $n$ in $U$ and $g$ in $G$. We write $name(ng) = n$ and $rdfgraph(ng) = g$. To enforce the blank node scoping rules ([3]) we make the global assumption that blank nodes cannot be shared between different named graphs, i.e. if $ng$ and $ng'$ are different named graphs then the sets of blank nodes which occur in triples in $rdfgraph(ng)$ and in $rdfgraph(ng')$ are disjoint.

All of the above definitions may be relativized to a particular set of URIrefs, or to a particular set of named graphs. Any set of named graphs can be thought of as a partial function from $U$ into the power set of $T$.

[1] We have removed the legacy constraint that a literal cannot be the subject of a triple.