

---

# Automating a Highly Expressive Knowledge Space: Dual Representation

2007-12-28

## Table of Contents

Introduction .....	1
Adopted Rule Language: RIF BLD .....	2
Macros for Dual Representation Archetypes .....	2
A Dual Representation Convention for XML->RDF .....	4
Value Proposition of Dual Representation in CMS and RESTful Applications .....	5
Bibliography .....	5

**Abstract.** SemanticDB represents a methodology for building and maintaining a highly flexible content management system built around a centralized vocabulary. This vocabulary incorporates a formal semantics (via a combination of an OWL ontology and SWRL,RIF,N3 rule sets) for hierarchical document composition and an abstract framework for modelling terms in a domain in a way that facilitates semi-automated use of native XML and RDF representations. It relies on a very recent set of technologies (Semantic Web technologies) to automate certain aspects of data entry, structure, storage, display, and retrieval with minimal intervention by traditional database administrators and computer programmers. A SemanticDB instance is built around a domain model expressed using a Data Node Model. From the domain model various XML/RDF management components are generated. The Data Mason takes a domain model and generates XML schemas, formal ontologies, XSLT transforms, stored queries, XML templates, document definitions, etc. The ScreenCompiler takes an abstract representation of a data entry form (with references to terms in a domain model) and generates a user interface in a particular host language (currently XForms)

## Introduction

The model described here is the same described in the Semantic Technology Conference 2006 proceedings for *Automating the Management of a Semi-structured Database* [DNM]. It is a semi-structured modelling language which uses RDF to describe info set items that participate in functional mappings to RDF. The functional mappings can be formally defined using GRDDL [GRDDL]. The model maps information items parsed from an XML document with namespace nodes which share a common base URI, to *Associative Box (Abox)* membership assertions (i.e., statements such as "item set #23211 is a member of a collection / class with the same name)" as well as other assertions about these information items.

The mapping consists of several transformation rules which convert RDF statements made between these *Data Nodes* (OWL Classes and OWL Properties), including: *contains*, *"is a specialization of"*, *"inherits structural constraints"*. These statements are converted into Description Logic (OWL-DL) expressions. Besides class membership assertions, the RDF graphs which are the range of these functional mappings (GRDDL result graphs) will also include *contains* relations made between instances of classes. This relation is interpreted as the inverse property ([OWL]) of `<http://www.w3.org/2001/04/infoset#parent>` as defined in the *RDF Schema for the XML Information Set* [[RDFS-INFO]]. These RDF graphs will conform to the OWL-DL expressions generated from an instance of the model which represents ontological commitment within a particular domain (vehicle almanacs, digital library resource management, etc.).

# Adopted Rule Language: RIF BLD

---

A variation of the human-readable concrete syntax for Rule Interchange Format BLD (modified to use compact URIs) is used to describe the semantics which require rule-based dialects for the additional expressiveness over OWL. Where OWL is sufficient, the Manchester OWL syntax [MANOWL] is used. The *dnode* namespace prefix is used when forming compact URIs [CURIE] in the namespace of the Data Node model.

## Macros for Dual Representation Archetypes

The `dnode:type` property is used to associate an XSD datatype to a leaf node in a *directed multigraph* of *Data Nodes* (an RDF graph of a data node model). These leaf data nodes are interpreted as RDF properties with the same URI. Depending on the XSD datatype assigned to these leaf nodes, some of them are interpreted as `owl:DatatypeProperties` with a pairwise disjoint set of values for their range, each with human-readable string labels. Others will be interpreted as simple `owl:DatatypeProperties` with an `rdfs:range` corresponding to the assigned datatype. The former subcategory of leaf nodes can be thought of as *Controlled Vocabularies*.

```
class: dnode:DataNode
```

```
class: dnode:RootNode
```

```
EquivalentTo: dnode:DataNode and not ( dnode:contained-by some dnode:DataNode )
```

```
class: dnode:LeafNode
```

```
SubClassOf: dnode:DataNode
```

```
(dnode:type min 1)
```

```
not ( dnode:contains some dnode:DataNode )
```

```
class: dnode:ControlledVocabulary
```

```
EquivalentTo: dnode:LeafNode and ( dnode:type value owl:DataRange )
```

In addition, a single `dnode:RootNode` is designated as the data node whose members (in an XML instance) will be the root node (or document element) in the corresponding XML instance document. The `dnode:contained-by` property is introduced for convenience to emphasize that these data nodes (and their instances) are true root nodes and cannot be contained by another data node.

There are several rules involving antecedents which include atoms that make use of the *contains* property between two (non-leaf) data nodes. These data nodes primarily comprise the backbone of a data node model. The first rule gives the general interpretation for the use of the *contains* property:

```
forall ?A ?C
```

```
  And(rdfs:subClassOf(?A _:univRest)  
    owl:Restriction(_:univRest)  
    owl:onProperty(_:univRest dnode:contains)  
    owl:allValuesFrom(_:univRest _:disjClass)  
    owl:unionOf(_:disjClass _:aList)  
    list:member(_:aList ?B1)  
    list:member(_:aList ?B2)  
    ...  
    list:member(_:aList ?BN)
```

```
  :- And(dnode:contains(?A ?B1)  
    dnode:contains(?A ?B2)
```

Automating a Highly Ex-  
pressive Knowledge Space:

---

```
... Dual Representation
dnode:contains(?A ?BN)
dnode:DataNode(?A)
dnode:DataNode(?B1)
...
dnode:DataNode(BN))
```

Which can be paraphrased as:

“When used between two data nodes (A and B), the *contains* relationship indicates that members of the first data node (an OWL class) which appear in a source XML instance document *can* be related to members of the second data node (also an OWL class) by the same property in the corresponding faithful RDF rendition of the source”

The list:member logic function (or built-in) is burrowed here as shorthand notation for constructing an anonymous class consisting of the boolean disjunction operation applied over each *contained* data node

The remaining rules deal with the owl:DatatypeProperties. These are either those with XSD datatypes as their rdfs:range:

```
Forall ?L ?P ?DT
  And(rdfs:subClassOf(?P _:Rest)
    owl:Restriction(_:Rest)
    owl:onProperty(_:Rest ?L)
    owl:allValuesFrom(_:Rest ?DT)
    rdfs:domain(?L ?P) :- And(dnode:contains(?P ?L)
      dnode:LeafNode(?L)
      dnode:type(?L ?DT)
      owl:disjointWith(?DT owl:DataRange))
```

or pair-wise disjoint sets of *values* (owl:DataRange).

```
Forall ?L ?P ?DT
  And(rdfs:subClassOf(?P _:Rest)
    owl:Restriction(_:Rest)
    owl:onProperty(_:Rest ?L)
    owl:allValuesFrom(_:Rest ?ENUM)
    rdfs:domain(?L ?P) :- And(dnode:contains(?P ?L)
      dnode:ControlledVocabulary(?L)
      dnode:enumeratedClassFn(?L ?ENUM))
```

The dnode:enumeratedClassFn logic function (arity 1) takes the dnode:ControlledVocabulary instance and returns an anonymous class which corresponds to the value space for the controlled vocabulary.

Finally, two sets of rules (one which is burrowed from the set of semantics associated with Description Logic subsumption and attributed to the rdfs:subClassOf property) cover other dual representation modeling patterns:

```
Forall ?A ?B ?C
  dnode:contains(?A ?C) :- And(Or(
    dnode:inheritsConstraints(?A ?B)
    rdfs:subClassOf(?A ?B)
  )
  dnode:DataNode(?P)
  dnode:DataNode(?L)
```

*Need some remark about how inheritsConstraints has weaker semantics than rdfs:subClassOf*

There is an analogy between how these rules are applied and the way in which Description Logic dialects (such as OWL) act as templates for well-defined portions of a much more expressive language. The RDF-based OWL dialect maps RDF statements to general forms of Description Logic expressions. Data node model RDF statements about data nodes are mapped to more specific Description Logic expressions. In particular, these expressions can be interpreted as describing infoset constraints which can be captured in XPath-based document processing dialects such as XSLT, XPath, XQuery, XProc, etc. By describing infoset constraints in tandem with model-theoretic constraints, a fully functional mapping from XML documents to RDF graphs can be provided. In this way an XML / RDF repository can guarantee a closed, fully-automated, faithful rendition of its information space. This setup is referred to as *dual representation*.

## A Dual Representation Convention for XML->RDF

The general pattern or convention setup by these rules and their interpretation is as follows. As an example, below is a data node model describing 5 data nodes in a data node multigraph.

```
ex:Root a dnode:DataNode;
        dnode:contains ex:NodeA, ex:NodeB, ex:hasFooB.
ex:NodeA a dnode:DataNode;
        dnode:contains ex:hasFooA.
ex:hasFooA a dnode:LeafNode;
        dnode:type xsd:string.
ex:NodeB a dnode:DataNode.
ex:hasFooB a dnode:ControlledVocabulary
        dnode:type ("value1" "value2" "value3" .. ).
```

An XML instance documents which conform to a data node model will be of the form:

```
<ex:Root>
  <ex:NodeA>
    <ex:hasFooA>.. string ..</ex:hasFooA>
  </ex:NodeA>
  <ex:NodeB />
  <ex:hasFooB>.. string ..</ex:hasFooB>
</ex:Root>
```

The corresponding faithful RDF rendition (through the use of a generated XSLT transformation) will be in the form:

```
[ ] a ex:root
    dnode:contains
      [ a ex:NodeA ; ex:hasFooA ".. string .." ],
      [ a ex:NodeB ].
    ex:hasFooB " .. string .. "
```

The full data node model framework includes an additional mechanism (not covered in this overview) which allows certain combinations of dnode:contains asserted between members of specific data nodes to instead be transformed (in the corresponding RDF) into assertions using a specific property instead. This provides more expressive relations between members of the data nodes besides the contains relationship whose semantics is limited to infoset constraint and thus very ambiguous.

# Value Proposition of Dual Representation in CMS and RESTful Applications

This guided modeling of a multigraph of OWL classes and RDF properties and the XML-based dialects that are generated from such a modeling is the way by which Content Management Systems with native support for the XML and RDF syntax as well as native support for GRDDL-like mechanisms are bootstrapped with predefined GRDDL transforms, XML schemas, OWL ontologies, and even fully generated XForms user interfaces.

This last capability (auto-generated screens) is yet another example of how the formally expressed infoset constraints can be used to facilitate operations on the XML source documents. In particular, a templated user interface description (which references the data nodes in the corresponding data node model) is processed by a *script* which takes as (additional input) the data node model and generates a fully featured XForms 1.1 document. The XForms 1.1 document is pre-built to perform user-guided XML operations on XML instances which make use of the namespace associated with the data node model.

These XML documents are sent back and forth over the wire (in very RESTful fashion) to the underlying CMS which updates the corresponding faithful rendition of each document with respect to the XSLT transform. In this way, the first-class representation (the original XML document) becomes the primary exchange format for data entry, replication, RESTful operations, etc.

The uniformity of the initial XML vocabulary solves the problem of impedance with associating CRUD user interfaces with RDF graphs directly. The dual representation capability also emphasizes representation state transfer operations where the entire state of a resource (an XML document in a CMS with a corresponding faithful RDF rendition) is sent back and forth from the CMS's HTTP server to a remote, browser-based application. This provides a uniform granularity for other features common in a CMS including ACL-based security policy, where access control is delegated at the document / graph level. In addition, this (to some extent) obviates the need to support a fine-grained update language for manipulating RDF graphs. Updates to the RDF graph can be made at the coarse-grained level associated with the HTTP PUT method.

## Bibliography

- [CPR] Institute of Medicine (Corporate Author), Elaine B. Steen (Editor), Don E. Detmer (Editor): The Computer-Based Patient Record: An Essential Technology for Health Care, 15 January 1997. National Academy Press; Rev Sub edition. <http://www.iom.edu/CMS/3809/22303.aspx>
- [REFTERM] Smith, B - Kusnierczyk, W - Schober, D - Ceusters, W: Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain Care, 8 November 2006. KR-MED 2006. [http://mor.nlm.nih.gov:8000/pubs/pdf/krmed2006\\_proceedings.pdf](http://mor.nlm.nih.gov:8000/pubs/pdf/krmed2006_proceedings.pdf)
- [WEBARCH] Jacobs, Ian - Walsh, Norman: Architecture of the World Wide Web, Volume One. W3C Recommendation, 15 December 2004. World Wide Web Consortium. <http://www.w3.org/TR/webarch/>
- [DLP] Grosf, B - Volz, R - Horrocks, I - Decker, S: Description Logic Programs: Combining Logic Programs with Description Logic. 2003. <http://www.cs.man.ac.uk/~horrocks/Publications/download/2003/p117-grosf.pdf>
- [SWRL] Grosf, B - Horrocks, I - Patel-Schneider, P - Boley, H - Tabet, S - Dean, M: SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission, 21, May, 2004. <http://www.w3.org/Submission/SWRL/>
- [SWTRANS] Ruttenberg, A - Clark, T - Bug, W - Samwald, M - Bodenreider, O - Chen, H - Doherty, D - Forsberg, K - Gao, Y - Kashyap, V - Kinoshita, J - Luciano, J - Marshall, M - Ogbuji, C - Rees, J - Stephens, S - Wong, G

Automating a Highly Ex-  
pressive Knowledge Space:

- Wu, E - Zaccagnini, D - Hongsermeier, T - Neumann, E - Herman, I - Cheung, K: Advancing translational research with the Semantic Web. 9 May 2007. BMC Bioinformatics. <http://www.biomedcentral.com/1471-2105/8/S3/S2>

- [DL2OWL] Horrocks, I - Patel-Schneider, P - Harmelen, F: From SHIQ and RDF to OWL: The Making of a Web Ontology Language., 2003. Journal of Web Semantics. <http://www.cs.man.ac.uk/~horrocks/Publications/download/2003/HoPH03a.pdf>
- [OBR] Ross, C - Kumar, A - Mejino, J - Cook, D - Detwiler, B: A Strategy for Improving and Integrating Biomedical Ontologies. 2005. AMIA 2005. <http://ontology.buffalo.edu/bio/OBR.pdf>
- [MVC] Burbeck, S: Applications Programming in Smalltalk-80 (TM): How to use Model-View-Controller (MVC). 1992. University of Illinois at Champagne-Urbana. <http://st-www.cs.uiuc.edu/users/smarch/st-docs/mvc.html>
- [SW] Berners-Lee T- Hendler J - Lassila O: The Semantic Web, May, 2001. Scientific American.
- [WONWEB] Masolo, C - Borgo, S - Gangemi, A - Guarino, N - Oltramari, A - Schneider, L: The WonderWeb Library of Foundational Ontologies: Preliminary Report, 29, May, 2003.
- [OWLTIME] Hobs, J - Pan, F: Time Ontology in OWL . 27 September, 2006. <http://www.w3.org/TR/owl-time/>.
- [SKOS] Miles, A - Brickley, D: SKOS Core Vocabulary Specification. 2 November 2005. <http://www.w3.org/TR/swbp-skos-core-spec/>.
- [PO\_CPR] PKC Corporation: A Problem Oriented Approach to the Computerized Patient Record. 1998. <http://www.pkc.com/papers/pomr.pdf>.
- [POMR] Weed L: Medical Records that Guide and Teach 1968. New England Journal of Medicine.
- [POMR\_KR] Bayegan, E: Knowledge Representation for Relevance Ranking of Patient-Record Contents in Primary-Care Situations. 2002. Department of Computer and Information Science, Norwegian University of Science and Technology. <http://www.idi.ntnu.no/~bayegan/pub/ebthesis.pdf>
- [KR] Davis, R - Shrobe, H - Szolovits, P: What is a Knowledge Representation?, 1993. MIT Computer Science and Artificial Intelligence Laboratory. <http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html>
- [HL7] Smith, B - Vizenor, L - Ceusters, W: Foundation for the Electronic Health Record: An Ontological Analysis of the HL7's Reference Information Model. Date?. [http://ontology.buffalo.edu/medo/HL7\\_2004.pdf](http://ontology.buffalo.edu/medo/HL7_2004.pdf)
- [GALEN\_EXP] Rector, A - Rogers, J. Ontological & Practical Issues in using a Description Logic to Represent Medical Concepts: Experience from GALEN. December, 2005. School of Computer Science, The University of Manchester.
- [GALEN] Rector, A - Nowlan, W. The GALEN project. Comput Methods Programs Biomed. October, 1994.
- [GRDDL-PRIMER] Halpin, H - Davis I. GRDDL Primer. W3C Recommendation, 28 June 2007. World Wide Web Consortium. <http://www.w3.org/TR/grddl-primer/> [<http://www.w3c.org/TR/grddl-primer>].
- [RDF-MT] Hayes, P. RDF Semantics. W3C Recommendation. 10, February, 2004. <http://www.w3.org/TR/rdf-mt/>
- [TURTLE] Beckett, D: Turtle - Terse RDF Triple Language, 4, December, 2007. <http://www.dajobe.org/2004/01/turtle/>
- [FOAF] Miller, L - Brickley, D: Turtle - Terse RDF Triple Language, 4, December, 2007. <http://xmlns.com/foaf/spec/>
- [RETE-UL] Doorenbos, R: Production Matching for Large Learning Systems, 31, January, 1995. Computer Science Department, Carnegie Mellon University.. <http://reports-archive.adm.cs.cmu.edu/anon/1995/CMU-CS-95-113.pdf>

Automating a Highly Ex-  
pressive Knowledge Space:

[CABG] American College of Cardiology/American Heart Association: ACC/AHA guidelines and Indications for  
Coronary Artery Bypass Graft Surgery, 1991. American Heart Association. <http://circ.ahajournals.org/cgi/reprint/83/3/1125.pdf>