



Report WP 2.3 KnowledgeWeb General Assembly Crete 2-3 June 2005

Carlos F. Enguix
DERI Galway, National University of Ireland

With contributions from members who assisted the meeting
Anna Zhdanova (University of Innsbruck, DERI Innsbruck)
Diana Maynard, Sheffield University, England
Enrico Franconi (Free University of Bozen-Bolzano, Italy)
Ian Blacoe (University of Liverpool, England)
Max Völkel (University of Karlsruhe, Germany)
Robert Stevens (Manchester University, England)
Stamatia Dasiopoulou (ITI/CERTH, Greece)
Stefan Decker (DERI Galway, Ireland)

Abstract.

EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB

Document Identifier	KWEB/2005/Report WP 2.3 June 2005/v1.0
Project	KWEB EU-IST-2004-507482
Version	v1.0
Date	June 17th, 2005
State	final
Distribution	public

Knowledge Web Consortium

This document is part of a research project funded by the IST Programme of the Commission of the European Communities as project number IST-2004-507482.

University of Innsbruck (UIBK) - Coordinator

Institute of Computer Science
Technikerstrasse 13
A-6020 Innsbruck
Austria
Fax: +43(0)5125079872, Phone: +43(0)5125076485/88
Contact person: Dieter Fensel
E-mail address: dieter.fensel@uibk.ac.at

France Telecom (FT)

4 Rue du Clos Courtel
35512 Cesson Sévigné
France. PO Box 91226
Fax: +33 2 99124098, Phone: +33 2 99124223
Contact person : Alain Leger
E-mail address: alain.leger@rd.francetelecom.com

Free University of Bozen-Bolzano (FUB)

Piazza Domenicani 3
39100 Bolzano
Italy
Fax: +39 0471 315649, Phone: +39 0471 315642
Contact person: Enrico Franconi
E-mail address: franconi@inf.unibz.it

Centre for Research and Technology Hellas / Informatics and Telematics Institute (ITI-CERTH)

1st km Thessaloniki - Panorama road
57001 Thessaloniki-Thessaloniki
Greece. Po Box 361
Fax: +30-2310-464164, Phone: +30-2310-464160
Contact person: Michael G. Strintzis
E-mail address: strintzi@iti.gr

National University of Ireland Galway (NUIG)

National University of Ireland
Science and Technology Building
University Road
Galway
Ireland
Fax: +353 91 526388, Phone: +353 87 6826940
Contact person: Christoph Bussler
E-mail address: chris.bussler@deri.ie

École Polytechnique Fédérale de Lausanne (EPFL)

Computer Science Department
Swiss Federal Institute of Technology
IN (Ecublens), CH-1015 Lausanne
Switzerland
Fax: +41 21 6935225, Phone: +41 21 6932738
Contact person: Boi Faltings
E-mail address: boi.faltings@epfl.ch

Freie Universität Berlin (FU Berlin)

Takustrasse 9
14195 Berlin
Germany
Fax: +49 30 83875220, Phone: +49 30 83875223
Contact person: Robert Tolksdorf
E-mail address: tolk@inf.fu-berlin.de

Institut National de Recherche en Informatique et en Automatique (INRIA)

ZIRST - 655 avenue de l'Europe -
Montbonnot Saint Martin
38334 Saint-Ismier
France
Fax: +33 4 7661 5207, Phone: +33 4 7661 5366
Contact person: Jérôme Euzenat
E-mail address: Jerome.Euzenat@inrialpes.fr

Learning Lab Lower Saxony (L3S)

Expo Plaza 1
30539 Hannover
Germany
Fax: +49-511-7629779, Phone: +49-511-76219711
Contact person: Wolfgang Nejdl
E-mail address: nejdl@learninglab.de

The Open University (OU)

Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA
United Kingdom
Fax: +44 1908 653169, Phone: +44 1908 653506
Contact person: Enrico Motta
E-mail address: e.motta@open.ac.uk

Universidad Politécnica de Madrid (UPM)

Campus de Montegancedo sn
28660 Boadilla del Monte
Spain
Fax: +34-913524819, Phone: +34-913367439
Contact person: Asunción Gómez Pérez
E-mail address: asun@fi.upm.es

University of Liverpool (UniLiv)

Chadwick Building, Peach Street
L697ZF Liverpool
United Kingdom
Fax: +44(151)7943715, Phone: +44(151)7943667
Contact person: Michael Wooldridge
E-mail address: M.J.Wooldridge@csc.liv.ac.uk

University of Sheffield (USFD)

Regent Court, 211 Portobello street
S14DP Sheffield
United Kingdom
Fax: +44 114 2221810, Phone: +44 114 2221891
Contact person: Hamish Cunningham
E-mail address: hamish@dcs.shef.ac.uk

Vrije Universiteit Amsterdam (VUA)

De Boelelaan 1081a
1081HV. Amsterdam
The Netherlands
Fax: +31842214294, Phone: +31204447731
Contact person: Frank van Harmelen
E-mail address: Frank.van.Harmelen@cs.vu.nl

University of Karlsruhe (UKARL)

Institut für Angewandte Informatik und Formale
Beschreibungsverfahren - AIFB
Universität Karlsruhe
D-76128 Karlsruhe
Germany
Fax: +49 721 6086580, Phone: +49 721 6083923
Contact person: Rudi Studer
E-mail address: studer@aifb.uni-karlsruhe.de

University of Manchester (UoM)

Room 2.32. Kilburn Building, Department of Computer
Science, University of Manchester, Oxford Road
Manchester, M13 9PL
United Kingdom
Fax: +44 161 2756204, Phone: +44 161 2756248
Contact person: Carole Goble
E-mail address: carole@cs.man.ac.uk

University of Trento (UniTn)

Via Sommarive 14
38050 Trento
Italy
Fax: +39 0461 882093, Phone: +39 0461 881533
Contact person: Fausto Giunchiglia
E-mail address: fausto@dit.unitn.it

Vrije Universiteit Brussel (VUB)

Pleinlaan 2, Building G10
1050 Brussels
Belgium
Fax: +32 2 6293308, Phone: +32 2 6293308
Contact person: Robert Meersman
E-mail address: robert.meersman@vub.ac.be

Executive Summary

The General KnowledgeWeb Assembly at Crete on the 2-3 of June was held in the Alde-mar Knossos Royal Hotel in Hersonissos, Crete. The agenda of WP 2.3 started at 11 o'clock on the 2nd of June, and lasted till 4 o'clock in the afternoon due to the fact there was a joint session with all research workpackages including a panel session with all work package leaders. The list of people who attended WP 2.3 meetings include (ordered by alphabetical order):

- Anna V. Zhdanova, DERI Innsbruck, Austria
- Carlos F. Enguix, DERI Galway, Ireland
- Diana Maynard, Sheffield University, England
- Enrico Franconi, Free University of Bozen-Bolzano, Italy
- Ian Blacoe, University of Liverpool, England
- Max Völkel, University of Karlsruhe, Germany
- Robert Stevens, Manchester University, England
- Stamatia Dasiopoulou ITI/CERTH, Greece
- Stefan Decker, DERI Galway, Ireland

Other people who did not attend the meetings but are part of the workpackage include Jerome Euzenat, Valentina Tamma, and Rokia Bendaoud. Sessions were held in the Hermes room, including powerpoint presentations showing the agenda to be covered and a presentation by Max Völkel with regard to ontology versioning and ontology versioning use cases.

Robert Stevens from Manchester University, has become an active member of the work package. Robert is a biologist who is also a doctor in Computer Science. Robert is a Bioinformatics Lecturer in the BioHealth Informatics Group at the University of Manchester. He has around 80 publications in international conferences, workshops, journals and so on. He was involved in the TAMBIS project for transparent access and integration of biological databases. Now one of his main interests is in the definition of formal biological ontologies. He is involved in the transformation of the Gene Ontology controlled vocabulary into a description-logics OWL based ontology. He is interested in contributing to the development of an ontology-based versioning system to the Gene Ontology which is part of the Open Biological Ontologies. Current Gene Ontology versions are maintained by CVS repositories which handle only syntactic differences among ontologies. In other words CVS is not able to differentiate class versions for instance, being able only to differentiate text/file differences.

The Workpackage will include three different use cases with quite different needs:

- Anna V. Zhdanova's People Portal for Ontology Consensus Framework
- Sebastian Kruk MarcOnt Digital Library system and
- Robert Stevens, Gene Ontology versioning system

The agenda mainly centered in the gathering of end-user requirements and use cases for ontology versioning. Also some issues about intra and inter collaboration among workpackage members were discussed. From 14:00 to 14:30 there was a joint session with WP2.5: Semantic Web Language Extensions, session coordinated and led by Jeff Pan and Ian Horrocks. WP 2.3 proposed an extension to RDQL and SPARQL to enable the querying of versions through bi-temporal database features such as valid-time and transaction-time and context information. The proposal was merely practical and did not include the intended semantics associated to the query language extensions, which in fact triggered some debate in the room. Our viewpoint was from a database perspective considering that both RDQL and SPARQL are SQL-like query languages for semi-structured/graph-based data. The point to be discussed is that SPARQL could simulate or include features available in SQL3 such as nested/correlated queries with the required closure of query results, include bi-temporal data such as found in temporal SQL, and possibly include procedural capabilities such as calls to external functions, use of surrogate methods, use of path expressions as in object oriented or graph databases and so on.

Contents

Executive Summary	i
Introduction	iv
1 Summary conclusions KW General Assembly, 24-25/01/2005, Hannover, Germany	2
2 Summarized Review of the current deliverables: D2.3.1 and D2.3.2	2
3 Current view status and future development of T2.3.3.2, T2.3.3.3, T2.3.4, T2.3.5, T2.3.6, T2.3.7	3
4 Deadline status and summary of the current deliverables: D2.3.3.a and D2.3.5.a	4
5 Discussion collaboration problems inside WP 2.3 and Joint prototyping and publication efforts of WP 2.3 with rest of packages	5
Collaboration inside WP 2.3	5
Collaboration among WP 2.3 and rest of WP	5
6 Requirements gathering for RDF and ontology versioning	6
The Gene Ontology and the Open Biological Ontologies: a summarised survey	7
GO format (deprecated)	8
OBO Format	9
XML Format	12
OWL Format	12
MySQL Version	13
7 Conclusions: Workplan/infrastructure and Task assignments	15
Workplan and Infrastructure	15
Gene Ontology Workplan	15
Task Assignment	16
8 Next Agenda: M18-M24	17

Introduction

The agenda proposed for the General Assembly in Crete 2-3 June 2005 for WP 2.3 was the following:

1. Summary conclusions KW General Assembly, 24-25/01/2005, Hannover, Germany
2. Summary First Knowledge Web Review Karlsruhe 14/03/2005
3. Summarized Review of the current deliverables:
 - (a) D2.3.1: Specification of a methodology for ontology syntactic and semantic versioning
 - (b) D2.3.2: Specification of knowledge acquisition and modeling of the process of the consensus
4. Current view status and future development of:
 - (a) T2.3.3.1 Report that describes how versioning systems for different ontology languages can be build on top of the RDF versioning system.
 - (b) T2.3.3.2 Evolving the RDF prototype delivered in M12 into a full RDF versioning system.
 - (c) T2.3.3.3 Implementing a full RDF-Schema versioning system on top of the RDF versioning system Including semantic diffs (transitive closure etc.).
 - (d) T2.3.4 Implementing a first prototype of a WSMO-Core versioning system on top of the RDF versioning system built in T2.3.3.3.
 - (e) T2.3.5 Implementation of a Semantic Web consensus making environment.
 - (f) T2.3.6 Defining criteria for ontology evaluation and building language dependant prototypes for evaluating ontologies especially RDF-Schema, OWL-light-minus and WSMO-Core ontologies.
 - (g) T2.3.7 Defining negotiation/argumentation techniques for agents complying to different ontologies in order to allow communication between them.
5. Deadline status and summary of the current deliverables:
 - (a) D2.3.3.a Full RDF versioning system including a specification for the mechanisms for versioning different ontology languages on top of the RDF versioning system. Leading Partner: UKARL Type: Report Deadline: June 05 (M18)
 - (b) 2.3.5.a Integration of Consensus Making Environment with RDF versioning system Leading Partner: UIBK Type: Prototype with report Deadline: Aug 05 (M20)

6. Discussion collaboration problems inside the workpackage 2.3 and Joint prototyping and publication efforts of Workpackage 2.3 with rest of packages:
 - (a) WP2.1: Scalability
 - (b) WP2.2: Heterogeneity
 - (c) WP2.4: Semantic Web Services
 - (d) WP2.5: Semantic Web Language Extensions
7. Requirements gathering for RDF and ontology versioning Open Issues: improvements, discussions, etc.
8. Conclusions
9. Discussion issues and agenda for next meeting

1 Summary conclusions KW General Assembly, 24-25/01/2005, Hannover, Germany

The General Knowledge Web Assembly in Hannover on the 24-25 of 2005, with regard to WP 2.3 concluded with the following summary:

- Versioning: New Requirements for Versioning Generic Versioning: Blank Nodes, Partial Implementation, Analysis of snapshot versus operation in ontology versioning and evolution, Alignment of versions, Cross Version Querying and publication of versioning API
- Deliverable T2.3.7 Defining negotiation/argumentation techniques for agents complying to different ontologies in order to allow communication between them: Presentation By Jerome and Valentina
- Deliverable D2.3.2: Specification of knowledge acquisition and modeling of the process of the consensus by Anna, Prototype for consensus building How to Enable and Enhance Collaboration, Protocol/bargaining/argumentation
- Deliverable D2.4.1: Semantic requirements for web services description
- General Interests of WP Members: a majority of members were interested in ontology evolution and ontology versioning. Other interests include consensus framework, formal semantics of ontology versioning, ontology modularization and query languages for ontology versioning.

2 Summarized Review of the current deliverables: D2.3.1 and D2.3.2

- D2.3.1: Specification of a methodology for ontology syntactic and semantic versioning
 1. Methodology for syntactic versioning of plain RDF, based upon set theoretic difference
 2. Brief indication of semantic versioning for ontologies on top of plain RDF
 3. B-Node enrichment using inverse functional properties to uniquely identify anonymous nodes
 4. Definition of an HTTP/based architecture for committing and updating versions on the Web
 5. Report about economic aspects of versioning and evolution

6. Report about inferencing based upon neural networks and fuzzy logics for versioning and evolution of ontologies
- D2.3.2: Specification of knowledge acquisition and modeling of the process of the consensus
 1. Knowledge Acquisition and Consensus Modeling Methodologies specification
 2. Models for Knowledge Acquisition and Consensus Making
 3. Report chapter about tools for Knowledge Acquisition and Consensus Making
 4. Specification for Next Generation Knowledge Acquisition and Modeling of Process of Consensus

3 Current view status and future development of T2.3.3.2, T2.3.3.3, T2.3.4, T2.3.5, T2.3.6, T2.3.7

- T2.3.3.2 Evolving the RDF prototype delivered in M12 into a full RDF versioning system: associated deliverables (D2.3.3.a Full RDF versioning system including a specification for the mechanisms for versioning different ontology languages on top of the RDF versioning system) to be delivered at the end of June 2005 by Max Völkel
- T2.3.3.3 Implementing a full RDF-Schema versioning system on top of the RDF versioning system Including semantic diffs (transitive closure etc.): associated deliverables (D2.3.3.b Full RDF versioning system. Including a specification for versioning different ontology languages on top of the RDF versioning system. RDF Schema (RDFS) versioning as a proof-of-concept for the layered approach) to be delivered at the end of December 2005 by Carlos F. Enguix with potential contributions from Max Völkel and Robert Stevens
- T2.3.4 Implementing a first prototype of a WSMO-Core versioning system on top of the RDF versioning system built in T2.3.3.3.: associated deliverables (D2.3.4 Implementing an OWL-light-minus and a first prototype of a WSMO-Core versioning system on top of the RDF versioning system) to be delivered at the end of June 2006 by Carlos F. Enguix with potential contributions from Max Völkel and Robert Stevens
- T2.3.5 Implementation of a Semantic Web consensus making environment: Associated deliverables:
 1. Internal D2.3.5.a Integration of Consensus Making Environment with RDF versioning system to be delivered at the end of August 2005 by Anna V. Zhdanova

2. D2.3.5.b Consensus Making Environment: a prototype for consensus framework with ontology versioning support should be delivered at the end of December 2005 by Anna V. Zhdanova with potential contributions from Robert Stevens, Max Völkel and Carlos F. Enguix
- T2.3.6 Defining criteria for ontology evaluation and building language dependant prototypes for evaluating ontologies especially RDF-Schema, OWL-light-minus and WSMO-Core ontologies: associated deliverables (D2.3.6 Prototypes of language dependent tools for ontology evaluation) delivered at the end of December 2005 by Diana Maynard with potential contributions and collaboration from Robert Stevens, Max Völkel, Carlos F. Enguix and Anna V. Zhdanova
 - T2.3.7 Defining negotiation/argumentation techniques for agents complying to different ontologies in order to allow communication between them: associated deliverables (D2.3.7 Report on negotiation/argumentation techniques among agents complying to different ontologies) delivered at the end of December 2005 by Valentina Tamma and Jerome Euzenat. According to conversations maintained at the KWeb General Assembly in Crete June 2-3 2005 Jerome indicated that the deliverable is almost finished and it may be submitted earlier.

4 Deadline status and summary of the current deliverables: D2.3.3.a and D2.3.5.a

D2.3.5.a Integration of Consensus Making Environment with RDF versioning system (Anna) Leading Partner: UIBK

- D2.3.3.a Full RDF versioning system including a specification for the mechanisms for versioning different ontology languages on top of the RDF versioning system

Author:	Max Völkel
Leading Partner:	UKARL
Type:	Report
Deadline:	June 2005 (M18)

According to Max this deliverable was almost completed and will fulfil its deadline at the end of June 2005.

The main issues contained in this deliverable are the following according to a summary received from Max:

- Software Development Java-based RDF versioning system with basic features like: Add initial model, Add model as successor to another model, Retrieve model based on timestamp or parent, Add and retrieve metadata per model, Bnode enrichment support, Set-based rdf diff

- Report: Description of setup and usage of the versioning system, Explanation of concepts such as: Bnode enrichment Model-based versioning, Ontology versioning on top of SemVersion.
- D2.3.5.a Integration of Consensus Making Environment with RDF versioning system

Author:	Anna V. Zhdanova
Leading Partner:	UIBK
Type:	Prototype with report
Deadline:	August 2005 (M20)

5 Discussion collaboration problems inside WP 2.3 and Joint prototyping and publication efforts of WP 2.3 with rest of packages

Collaboration inside WP 2.3

In this session it was discussed the production of a joint state-of-the-art journal paper on the basis of the deliverable D2.3.2: Specification of knowledge acquisition and modeling of the process of the consensus between Stamatia Dasiopoulou from ITI/CERTH, Greece, Anna V. Zhdanova, from DERI Innsbruck, Austria and Diana Maynard, from Sheffield University, England

The incorporation of Robert Stevens as an active member of WP 2.3 may enhance the collaboration of WP 2.3 members in the versioning of ontologies in the Gene Ontology (GO) ¹ as a non-trivial important project that requires a consensus and cohesive participation of integrating members.

Up-till now there has been limited collaboration among WP 2.3 members. The Knowledge Web Portal is currently being updated in order to enhance collaboration among WP 2.3 members indicating potential contributions in the delivery of research prototypes, reports and deliverables among leadership members and collaborators.

Collaboration among WP 2.3 and rest of WP

- WP2.1: Scalability: there is no clear collaboration among WP 2.3 and WP 2.1. It may be an important issue in the evaluation of scalability in ontology versioning of distributed ontologies and consensus framework.

¹<http://www.geneontology.org/>

- WP2.2: Heterogeneity: same scenario, there is no clear collaboration among WP 2.3 and WP 2.2. There might be opportunities for also evaluating the suitability of the prototypes developed for ontology versioning and consensus framework with respect to ontology heterogeneity: for instance evaluate the adaptability and suitability of Semversion (ontology versioning prototype) for biological ontologies (Gene Ontology), Web Services (WSMO), medical ontologies (Unified Medical Language System UMLS) and so on.
- WP2.4: Semantic Web Services: deliverable D2.3.4 Implementing an OWL-lightminus and a first prototype of a WSMO-Core versioning system on top of the RDF versioning system requires collaboration among WP 2.3 and WP 2.4. This deliverable is due on June 2006.
- WP2.5: Semantic Web Language Extensions: this WP has attempted to gather requirements from all WP's in order to define use cases for extending current Semantic Web languages. The General Assembly in Crete June 2-3 2005 included a joint session of WP 2.5 and rest of workpackages. The contribution of WP 2.3 was the inclusion of bi-temporal database attributes having temporal SQL as a guideline for such inclusion (valid-time and transaction time) and the inclusion of context information in the form of quads. No formal semantics were defined for the language extensions proposed by WP 2.3. In fact language extensions were proposed from a pragmatic and practical perspective indicating the needs for querying ontology versions. Nevertheless it was proposed by Jeff Pan a follow up of the discussion in the mailing lists.

6 Requirements gathering for RDF and ontology versioning

In WP 2.3 we will have use cases for Anna V. Zhdanova's Consensus Framework and People's Portal Prototype (DERI Innsbruck), Sebastian Kruk's MarcOnt Digital Library (DERI Galway) and Robert Stevens' Gene Ontology (Manchester University).

Max Völkel has been gathering some use cases from Anna V. Zhdanova's Consensus Framework and People's Portal Prototype.

With regard to Sebastian Kruk's MarcOnt Digital Library it is expected to start soon the process of gathering requirements and use cases. From initial conversations it appears that MarcOnt needs are pretty different from the rest of use cases.

The major effort might be centered in the gathering of use cases and case studies from the Gene Ontology. What follows next is a preliminary list of user requirements and use cases associated to the Gene Ontology. Robert Stevens provided us with seed information that enabled us to compile a summarised survey of the GO.

The Gene Ontology and the Open Biological Ontologies: a summarised survey

*"The goal of the Gene Ontology (GO) consortium is to produce a controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing. GO provides three structured networks of defined terms to describe gene product attributes."*²

The Gene Ontology "per se" is not an Ontology in the formal sense, it is rather a cross-species controlled biological vocabulary as previously indicated above. The Gene Ontology is divided in three disjoint sub-ontologies, currently stored in big flat files or also stored in persistent repositories such as a relational database (MySQL database). The three sub-ontologies are divided into vocabularies that describe gene products in terms of:

- Molecular functions
- Associated biological processes and
- Cellular components

The GO ontology permits to associate biological relationships among molecular functions, the involvement of molecular functions in biological processes and the occurrence of biological processes at a given time and space in cells [Con01]. Whereas the molecular function defines what a gene product does at the biochemical level, the biological process normally indicates a transformation process triggered or contributed by a gene product involving multiple molecular functions. Finally the cellular component indicates the cell structure a gene product is part of.

The Gene Ontology contains around 20.000 concepts which are convertible to OWL. The latest statistics about the GO could be found at the GO site³:

Current term counts: as of June 20, 2005 at 6:00 Pacific time
17946 terms, 94.2% with definitions.
6984 (38.9%) Molecular functions
9410 (52.4%) Biological processes
1552 (8.6%) Cellular components
There are 998 obsolete terms not included in the above statistics (Total Terms=18944)

According to [SWLG03] the GO is a handcrafted ontology accepting only "is-a" and "part-of" relationships. The hierarchical organization is represented via a directed-acyclic-graph (DAG) structure similar to the representation of Web pages or hypertext systems. Members of the Consortium group contribute to updates and revisions of the GO. The Go is maintained by editors and scientific curators who notify GO users of

²Extracted from the OBO site <http://obo.sourceforge.net/>

³Ontologies and Definitions <http://www.geneontology.org/GO.downloads.shtml#ont>

ontology changes via email, or at the GO site by monthly reports ⁴. Please note that ontology creation and annotation of GO terms in databases (association of GO terms with gene products) are two different operations. Each annotation should include its data provenance or source (a cross database reference, a literature reference, etc).

GO is available in several different formats:

- OBO Flat File Format (obo extension)
- GO Flat File Format (ontology extension)
- XML (RDF/XML) File Format (rdf-xml extension)
- OWL (RDF/XML) File Format (owl extension)
- MySQL Version

The structure of a GO Term is as follows:

- term name (e.g. cell)
- a GO identifier/accession number: an arbitrary (non-semantic, meaningless) unique, zero-padded seven-digit identifier prefixed by GO (e.g. GO:0005623)
- optional synonyms (e.g. synonym of apoptosis= type I programmed cell death)
- dbxrefs or optional database cross references: identifiers used to maintain cross references among databases (e.g. term retinal isomerase activity has the database cross reference EC:5.2.1.3 which is the accession number of this enzyme activity in the Enzyme Commission database)
- definition (e.g. "The action characteristic of a gene product." [GO:curators])
- comment (e.g. Note that this term refers to both the old and new (post-1998))

GO format (deprecated)

The GO ontology in the GO format is distributed into three flat files:

- Biological Process (process.ontology)
- Molecular Function (function.ontology)
- Cellular Component (component.ontology)

⁴see <http://www.geneontology.org/MonthlyReports/>

Each GO flat file contains the following headers containing metadata associated to the generation of the terms:

```
!autogenerated-by: DAG-Edit version 1.315
!saved-by: midori
!date: Fri Jan 03 17:14:37 GMT 2003
!version: $Revision: 1.10 $
!type: % ISA Is a
!type: < PARTOF Part of \Gene\_Ontology ; GO:0003673
```

Each definition in the GO ontology, which is stored in the GO.defs file, contains the following:

- term: the name of the term to which the definition refers (e.g. (+)-camphor biosynthesis)
- goid: the term's unique identifier (e.g. GO:0046211)
- definition: the definition of the term (e.g. The formation from simpler components of (+)-camphor, a bicyclic monoterpene ketone.)
- definition_reference: one or more references for the definition (e.g. GO:ai)
- comment: a definition may also include an optional comment

Example of a snapshot of an ontology in the GO format:

```
!autogenerated-by: DAG-Edit version 1.419 rev 3
!saved-by: gocvs
!date: Sat Jun 18 21:00:14 PDT 2005
!version: $Revision: 3.269 $
!type: % is_a is a
!type: < part_of part of
!type: ^ inverse_of inverse of
!type: | disjoint_from disjoint from
$Gene_Ontology ; GO:0003673
<biological_process ; GO:0008150
%behavior ; GO:0007610 ; synonym:behaviour
%adult behavior ; GO:0030534 ; synonym:adult behaviour
%adult feeding behavior ; GO:0008343 ; synonym:adult feeding behaviour % feeding behavior ; GO:0007610
%adult locomotory behavior ; GO:0008344 ; synonym:adult locomotory behaviour % locomotory behavior ; GO:0007610
%adult walking behavior ; GO:0007628 ; synonym:adult walking behaviour
%flight behavior ; GO:0007629 ; synonym:flight behaviour
%jump response ; GO:0007630
%chemosensory jump behavior ; GO:0007636 ; synonym:chemosensory jump behaviour % chemosensory behavior
%behavioral response to cocaine ; GO:0048148 ; synonym:behavioural response to cocaine % response to cocaine
%behavioral response to ethanol ; GO:0048149 ; synonym:behavioural response to ethanol % response to ethanol
%behavioral response to ether ; GO:0048150 ; synonym:behavioural response to ether % response to ether
%behavioral response to nicotine ; GO:0035095 ; synonym:behavioural response to nicotine % response to nicotine
```

OBO Format

This format is an extension of the tag-value format of the GO definitions file permitting the use of unknown or unrecognized tags. It was designed to improve the shortcomings of

the old GO flat format having in mind human readability, extensibility, non-redundancy and machine processable (easy to parse by programs). Unlike the GO flat file format that deals with rooted DAG graphs, the OBO format deals with unrooted or multiple rooted, cyclic, directed graphs.

An OBO Flat file document is structured as follows:

```
<header>
[<Object type>]
<tag>: <value>
<tag>: <value>
...
```

All tag-value pairs occur on a single line or broken by

```
\<newline>
```

combinations. This is an example of tag-value pairs:

```
[Term]
id: GO:0019383
name: (+)-camphor catabolism
def: "The catabolism of (+)-camphor." [GO:ma "Michael
Ashburner \"was responsible for creating this term\""]
comment: This is a gratuitous example\nof an escaped newline
```

The format header includes several metadata items related to versioning information: version, date, saved-by and auto-generated-by. The tags include the following information which allows the definition of formal vocabularies or ontologies and the further translation into description-logics based OWL ontologies:

- id: unique identifier of the current term
- name: unique term name
- alt_id: alternate identifier. A term may have n alternative identifiers
- namespace: the namespace associated. If not indicated the default namespace is associated
- definition: the definition of the term. At the most one definition per term
- comment: a comment for this term
- subset: the term subset which belongs the term
- synonym: a term may have n synonyms
- related_synonym, exact_synonym, broad_synonym, narrow_synonym: these will define siblings, equivalent terms, hypernym and hyponym terms respectively. A term may have n related synonyms.

- `is_a`: indicates "is a" relationships. A term may have several parent is-a relationships (multiple inheritance). Only root terms have no is-a relationship.
- `relationship`: indicates a relationship among terms. A term may have n relationships.
- `is_obsolete`: indicates the term is obsolete or to be deprecated
- `use_term`: indicated the term to be used if this term is obsolete
- `domain`: indicates the type of class of children terms to be related to this term
- `range`: indicates the type of class of parent terms to be related to this term
- `is_cyclic`: indicates that cycles are allowed (in a way it is contradicting the DAG structure of GO)
- `is_transitive`: indicates the relationship is transitive. Useful for reasoners to infer new relationships.
- `is_symmetric`: indicates the relationship is symmetric from parent-to-child and child-to-parent.

This is an example of a snapshot of an ontology in OBO format:

```
format-version: GO_1.0
!any comment here
typeref: relationship.types
subsetdef: goslim Generic GO Slim"
version: $Revision: 1.10 $
date:April 18th, 2003
saved-by: jrichter
remark: Example file

[Term]
id: GO:0003674
name: molecular_function
def: "The action characteristic of a gene product." [GO:curators]
subset: goslim

[Term]
id: GO:0016209
name: antioxidant activity
is_a: GO:0003674
def: "Inhibition of the reactions brought about by dioxygen or
peroxides. \Usually the antioxidant is effective because it can
itself be more easily \oxidized than the substance protected. The
term is often applied to \components that can trap free radicals,
thereby breaking the chain \reaction that normally leads to
extensive biological damage." \
[ISBN:0198506732]
```

XML Format

Actually it is serialized as simple RDF/XML including two namespaces, one for RDF itself and one for the GO vocabulary. ⁵ Example of a snapshot of the GO in XML format:

```
<?xml version="1.0" encoding="UTF-8"?> <!DOCTYPE go:go>
<go:go xmlns:go="xml-dtd/go.dtd#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <go:version timestamp="Wed May 9 23:55:02 2001" />
  <rdf:RDF>
    <go:term rdf:about="go#GO:0003673">
      <go:accession>GO:0003673</go:accession>
      <go:name>Gene_Ontology</go:name>
      <go:definition></go:definition>
    </go:term>
    <go:term rdf:about="go#GO:0003674">
      <go:accession>GO:0003674</go:accession>
      <go:name>molecular_function</go:name>
      <go:definition>The action characteristic of a gene product.</go:definition>
      <go:part-of rdf:resource="go#GO:0003673" />
      <go:dbxref>
        <go:database_symbol>go</go:database_symbol>
        <go:reference>curators</go:reference>
      </go:dbxref>
    </go:term>
```

OWL Format

Serialized also as RDF/XML format defining subClassOf (is-a) and part-of relationships. It seems that the GO consortium is not making use of versioning-related OWL-lite statements such as:

- versionInfo: information of the version
- priorVersion: the previous ontology version from which it has as ancestor
- backwardCompatibleWith: compatible with previous ontology versions
- incompatibleWith: incompatible with previous ontology versions
- DeprecatedClass: class preserved for backward-compatibility purposes but may disappear in future versions
- DeprecatedProperty: property preserved for backward-compatibility purposes but may disappear in future versions

This is an example of a snapshot of the GO in OWL format: ⁶

⁵see also <http://www.geneontology.org/GO.format.shtml#XML>

⁶see also <http://archive.godatabase.org/latest-termdb/go.daily-termdb.owl.gz>

```
<?xml version="1.0"?> <rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns="http://www.geneontology.org/owl#"
xml:base="http://www.geneontology.org/owl/">
  <owl:Ontology rdf:about=""/>
  <Class rdf:ID="GO_0000001">
    <rdfs:label>mitochondrion inheritance</rdfs:label>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">The distribution of mitochondri
<!-- organelle inheritance -->
    <rdfs:subClassOf rdf:resource="#GO_0048308"/>
<!-- mitochondrion distribution -->
    <rdfs:subClassOf rdf:resource="#GO_0048311"/>
  </Class>
  <Class rdf:ID="GO_0000002">
    <rdfs:label>mitochondrial genome maintenance</rdfs:label>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">The maintenance of the structur
<!-- mitochondrion organization and biogenesis -->
    <rdfs:subClassOf rdf:resource="#GO_0007005"/>
  </Class>
```

MySQL Version

The MySQL version of the GO ontology consists of four databases:

- termdb: ontologies, definitions and mappings to other databases
- assocdb: contains the termdb database plus associations to gene products
- seqdb: contains the assocdb database plus protein sequences for gene products
- seqdblite: contains the assocdb database with IEA associations excluded (Inferred from Electronic Associations, when normally no curator has checked the annotation to verify its accuracy)

The GO database releases are available, either as RDF XML or as a MySQL database dump. The GO Database is built from GO format flat files for solely querying purposes. These are some of the typical queries performed against a MySQL or Postgres Database:

- Find all the children of a term (is-a relationship)
- Fetch every descendent of a given term (all is-a relationships, hyponyms)
- Find the ancestors of a given term (all is-a parents, all hypernyms)
- Find ancestor terms and their relationships
- Find common parents for a set of sibling terms
- Find the distance between two nodes in the graph
- Find every term that has been annotated in a given way

- Find correlations between terms
- Transitive Correlations

The following figure fig. 1 depicts the relational model for the database schema associated to the GO ontology:

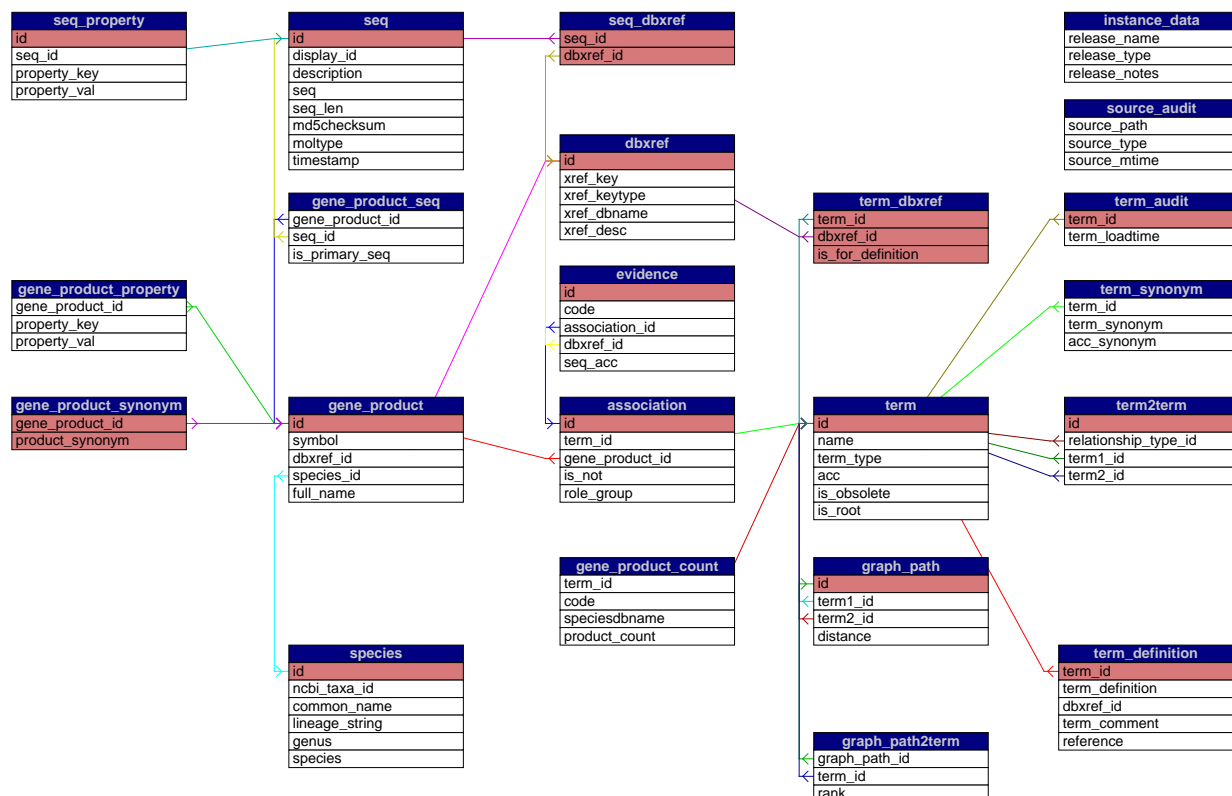


Figure 1: Relational Database Schema for the GO Ontology

There are some efforts to formalise the language used to term names in the Gene Ontology (GO) and the Open Biomedical Ontologies (OBO) such as Obol⁷. The GO is contained within OBO in three disjoint sub-ontologies: Molecular function, Biological process and Cellular component. There are also efforts to map OBO to OWL Dialects as previously indicated in the GO OWL flat file format.

GO currently uses CVS for versioning and storing the syntax of the ontology. It is expected WP 2.3 may have to define an API to CVS for capturing GO snapshot versions, and comments indicating the reason for changes.

⁷Obol <http://www.fruitfly.org/~cjm/obol/>

It is expected we will set up two layers on top of Semversion RDF versioning: RDF Schema (RDF-S) and a dialect of OWL to the Gene Ontology.

Manchester University uses the Distributed, Loosely-controlled and Evolving Engineering processes of Ontologies (DILIGENT) Methodology⁸ as an ontological engineering framework. The DILIGENT methodology comprises five main activities: initial semi-automatic ontology construction (build process), adaptation and curation of ontologies (local adaptation process), ontology change analysis (analysis process), revision and comparison of local and shared ontologies (revision process) and update of local ontologies and reuse and adaptation of concepts (local update process)

It is quite common that new approaches, methodologies, and software artifacts suddenly create new needs for a given community. We identify a clear opportunity in the development of a versioning system adapted to the GO ontology that enables the editing and update of ontologies in a distributed environment from a much finer granularity, at the concept level instead of the file-resource level as currently supported by CVS. Also time travel queries that enable to retrieve older database annotations, annotated according to previous GO ontology versions may be important for hypothesis testing (selection among different hypothesis) and re-creation of "*in-silico*" biological experiments.

It seems that the GO Consortium and the OBO community are still using quite simplistic and non-optimized approaches for the editing, updating and versioning of ontologies and notification of ontology changes to users. We also identify a clear opportunity for developing a consensus modelling framework in order to facilitate and improve the ontology engineering life-cycle with regard to the development of the GO and OBO ontologies. Such consensus environment could for instance automate the notification of users almost in real-time using event-oriented rule-based systems.

7 Conclusions: Workplan/infrastructure and Task assignments

Workplan and Infrastructure

In WP 2.3 we will have use cases for Anna V. Zhdanova's Consensus Framework and People's Portal Prototype (DERI Innsbruck), Sebastian Kruk's MarcOnt Digital Library (DERI Galway) and Robert Stevens' Gene Ontology (Manchester University).

Gene Ontology Workplan

1. Select several seed GO snapshots from CVS

⁸DILIGENT Bibliography http://www.aifb.uni-karlsruhe.de/Publikationen/showPublikationen?id_db=2055

2. Define OBOL-RDFS vocabulary
3. Analyze CVS comments/mine patterns and analyze documentation
4. Define consensus modelling for the GO.

Task Assignment

We concluded the meeting by indicating the tasks to be associated to each WP 2.3 members. Other members not included in the list are encouraged to participate at least as potential contributors.

- Robert Stevens: providing snapshots of the Gene Ontology. Robert and Jeff Pan from Manchester University suggested the celebration of a joint workshop in Manchester in order to gather use cases and user requirements from the GO. Most probably this workshop will be held either in August or in September 2005 (to be confirmed). Alternatively it is expected to organise monthly phone conferences over skype to enhance collaboration among members.
- Max Völkel and Carlos F. Enguix: RDF versioning system. It is expected that Carlos will visit Max in Karlsruhe on the beginning of July.
- Carlos F. Enguix and Max Völkel: learning and conversion of the GO/OBOL ontologies and conversion to an RDF-S vocabulary
- Enrico Franconi: semantic Diffs operators or formalisation in the definition of Semantic Diffs in ontologies. Enrico Franconi has previous experience in the field of Description Logics and temporal logics and temporal databases. It would be interesting to include his expertise especially in the formal definition of bi-temporal database querying features and context information in Semversion.
- Anna V. Zhdanova: Consensus modelling framework with dynamic of changes in snapshots in the GO. There should be a tight collaboration between Anna and Robert Stevens in the versioning of the GO, in order to define a prototype and framework for a consensus modelling framework in the GO.
- Max Völkel, Carlos F. Enguix and Andreas Harth (DERI Galway): scalable versioning system coupled with the YARS prototype. YARS is a scalable and distributed RDF repository based upon the use of local b-trees based upon the use of distributed hash tables.⁹

⁹Project Web site <http://sw.deri.org/2004/06/yars/yars.html>

8 Next Agenda: M18-M24

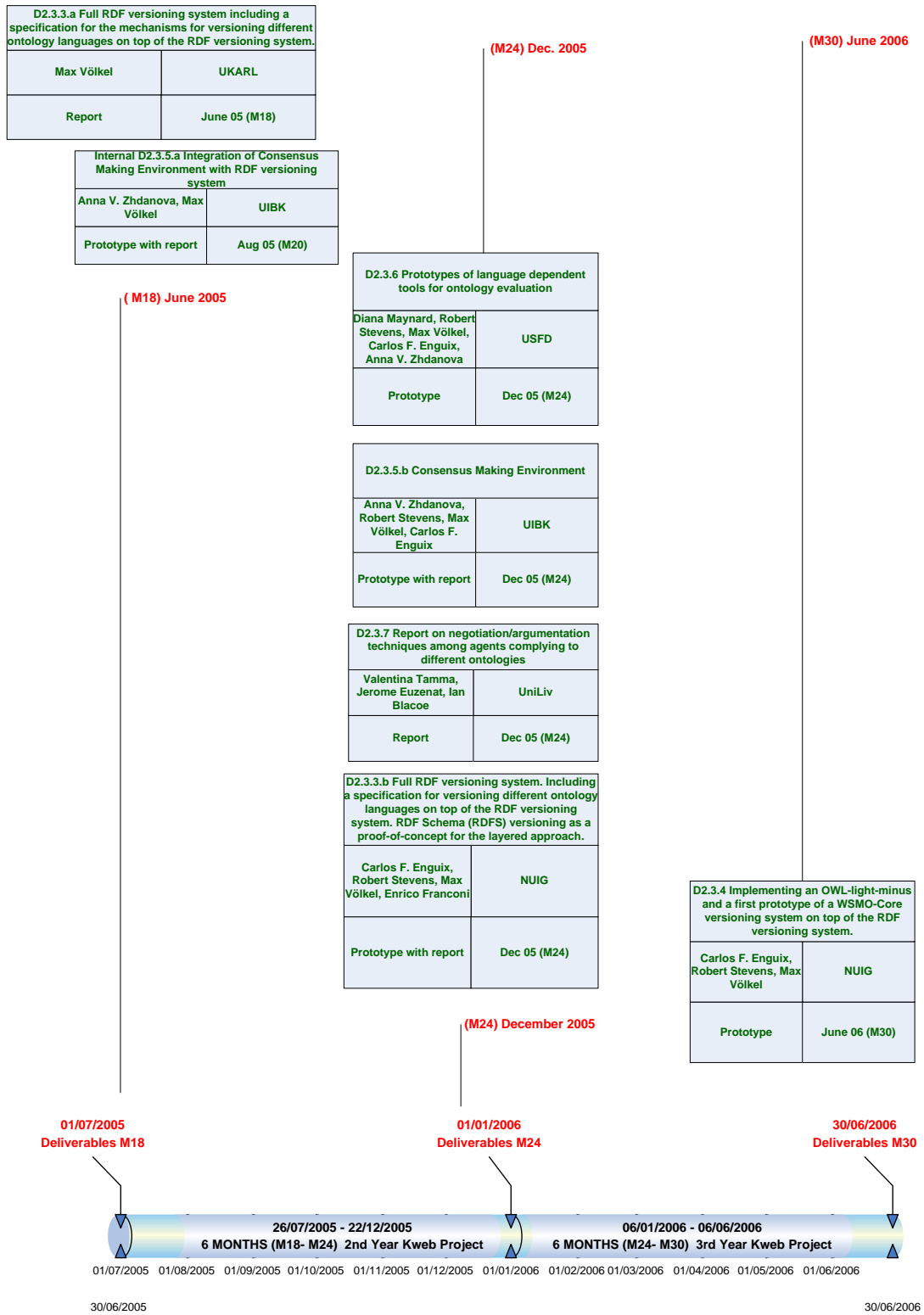
The coming semester (M18-M24) represents a critical stage with major milestones with regard to WP 2.3 as it is expected to be released a first prototype of an RDF-S versioning system on top of Semversion RDF system and a first prototype for a consensus making environment adapted for the GO.

In order to be successful in the release of deliverables (reports and research prototypes) it would be necessary to enhance the collaboration among WP 2.3 members, which requires a consensus among members in the set up of goals and priorities, availability in their respective personal agendas and a willingness to tighten relationships among participating entities and people.

We conclude the report indicating the possible options to enhance collaboration among WP 2.3 members:

- Frequently submit messages to WP 2.3 mailing list: kweb-wp23@lists.deri.org
- Organize monthly internet phone conferences over skype
- Publish and advertise on the mailing list the skype identifiers and internet chatting and conferencing identifiers of members
- Set up a collaboration site (BSWC) for submitting related materials
- Set up of a wiki for inclusion of ideas as a collaborative dashboard: this would be very convenient for the GO versioning effort
- Exchange visits among participating entities and people

The following diagram, a modified PERT diagram which is available on the KnowledgeWeb portal, summarises the involvement of members on each task for the remaining life-span of the KnowledgeWeb project



References

- [Con01] The Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Research*, 11(8):1425–33, 2001.
- [SWLG03] Robert Stevens, Chris Wroe, Phillip Lord, and Carole Goble. Ontologies in bioinformatics. In Stefan Staab and Rudi Studer, editors, *Handbook on Ontologies in Information Systems*, pages 635–657. Springer, 2003.