

Ontolog UBL Ontology Project Status Report

Purpose

This document summarizes the formation of the Ontolog Forum and establishment of the Ontolog UBL Ontology Project. It was prepared for the OASIS UBL Committee by Kurt Conrad on 2003-04-28.

Contents

Executive Summary	2
1.0 Background	3
1.1 What is an Ontology?	3
1.2 Origins of the Ontolog Forum	3
1.3 Ontolog Forum Logistics	4
2.0 Ontolog UBL Ontology Project	5
2.1 Overview	5
2.2 Expected Relationship to and Impact on UBL Schemas	5
2.3 Project Goals	6
2.4 Technical Goals	6
2.5 Project Management Strategy	6
2.6 Methodology	7
3.0 Project Status by Activity Area	8
3.1 Technology Selection (ongoing)	8
3.2 Determining Domain and Scope (ongoing)	9
3.3 Considering the Reuse of Existing Ontologies (complete)	9
3.4 Enumeration of important terms in the ontology (commencing)	9

Executive Summary

The Ontolog Forum was established as an independent effort in September, 2002, in large part to focus on ontological issues relating to the development of UBL. In March, 2003, the Ontolog Forum launched a project to develop formalized ontologies based on UBL.

It is expected that this project will:

- Provide a useful demonstration of ontological engineering.
- Facilitate the training of interested parties.
- Provide the basis for meaningful input to the OASIS UBL effort.
- Enable more advanced software processing of UBL documents.

To date, the UBL Ontology Project Team has:

- Selected an adapted a methodology.
- Started the articulation of use cases dealing with
 - The automated reconciliation of Purchase Orders and Invoices
 - The identification of the proper structure of an address from context
 - The mapping between different standards and representation languages
- Commenced mapping of UBL constructs to the Suggested Upper Merged Ontology (SUMO)
 - Purchase Orders
 - Invoices
 - Shipping Documents

1.0 Background

1.1 What is an Ontology?

[Adapted from Section 1.1 of “Web Ontology Language (OWL) Use Cases and Requirements” see <http://www.w3.org/TR/2003/WD-webont-req-20030331/#onto-def>]

An ontology defines the terms used to describe and represent an area of knowledge. Ontologies comprise computer-usable definitions of basic concepts in a domain and the relationships among those concepts.

Although XML DTDs and XML Schemas are sufficient for exchanging data between parties who have agreed to the definitions beforehand, their lack of semantics prevent machines from reliably understanding the meaning of new XML vocabularies and automatically identifying appropriate processing behaviors.

Ontologies, in contrast, are usually expressed using a logic-based language, so that detailed, accurate, consistent, sound, and meaningful distinctions can be made among the classes, properties, and relations. Applications that using ontologies can be considered "intelligent," in the sense that they can more accurately work at the human conceptual level.

The word ontology has been used to describe artifacts with different degrees of structure. These range from simple taxonomies (such as the Yahoo hierarchy), to metadata schemes (such as the Dublin Core), to logical theories. Highly-formalized ontologies specify descriptions for the following kinds of concepts:

- Classes (general things) in the many domains of interest
- The relationships that can exist among things
- The properties (or attributes) those things may have

1.2 Origins of the Ontolog Forum

In March and April of 2002, Peter Yim and other members of the OASIS Universal Business Language (UBL) Committee with backgrounds and interests in ontologies attempted to establish an ongoing discussion dealing with ontological issues under auspices of UBL Library Content Subcommittee (oasis-open.org/committees/ubl/lcsc/).

Two factors made this approach problematic: 1) Early feedback indicated more interest in learning about ontologies than applying experience with ontological development 2) Tight timetables, deadlines, and priorities within the UBL effort made it an inappropriate forum for the conversation.

After discussions with Jon Bosak, the chair of the UBL committee, the conveners of the Ontolog Forum (Kurt Conrad, Leo Orbst, and Peter Yim) decided to reconstitute the discussion outside of the UBL effort by establishing as an open forum to:

- 1) Discuss practical issues and strategies associated with the development of both formal and informal ontologies used in business.
- 2) Identify ontological engineering approaches that might be applied to the UBL effort.

Where the discussion raises specific issues or leads to specific proposals or recommendations for the UBL effort, the Ontolog Forum was asked to separately document and forward their recommendations

Ontolog UBL Ontology Project Status Report – 2003-04-28

to the UBL Chair for consideration. Further, it was requested that such input be limited to actionable recommendations, not just general issues that would require further study.

In addition to the primary statements of purpose, the Ontolog Forum was established to:

- Distribute news and information about the Ontology development community.
- Be pragmatic and not take a purely philosophical or academic approach to the topic.
- Contribute to the development of future standards. The ongoing relationship with the UBL effort is one example of this.
- Encourage forum members to propose and decide upon their own objectives.
- Eventually host a repository of business ontologies.

A general invitation was sent out in September 2002. The Ontolog Forum currently has 30 active members and 36 observers. The following UBL members are also subscribed to Ontolog:

Michael Adcock
Bill Burcham
Sally Chan
Kurt Conrad
Eduardo Gutentag
Monica Martin

Tim McGrath
Bill Meadows
Marion Royal
Lisa Seaburg
Peter Yim

Dong Soo Kim
Sue Probe
Patrick Yee

1.3 Ontolog Forum Logistics

A formal governance policy has not been adopted, but activities follow the OASIS process as closely as possible. Thus far, we have relied almost exclusively on a consensus model.

The Ontolog Forum uses infrastructure provided by CIM Engineering, Inc. (www.cim3.com).

Most of the work is done via the mailing list. Messages can be posted by members to ontolog-forum@ontolog.cim3.net. A number of delivery options are available for those who don't want the daily traffic. Archives can be accessed by anybody at ontolog.cim3.net/forum/ontolog-forum/.

With the launch of the Ontolog UBL Ontology Project, the infrastructure was augmented in a number of important ways:

- A shared, web-accessible work space was made available at ontolog.cim3.net/file/.
- A community Wiki has been established at ontolog.cim3.net/wiki/. An increasing amount of work has started to be done via Wiki pages (taking of meeting minutes, summarizing topic threads, etc.). A summary of changes to the Wiki space is also being distributed on a weekly basis.
- A schedule of weekly conference calls have been established, using TightVNC for screen-sharing support.

2.0 Ontolog UBL Ontology Project

2.1 Overview

As of March, 2003, there had been considerable discussion around the idea of creating an ontology based on the UBL schemas. After discussing this idea with some of the principles of the UBL effort, the general consensus was that the current specification (oasis-open.org/committees/ubl/lcsc/0p70/) was stable enough to begin engineering an ontology from it.

It also appeared that the project would mesh nicely with the goals that motivated many to join the Ontolog Forum in the first place, most notably to:

- Learn about ontologies (concepts, language, best practices)
- Identify a lifecycle process for developing ontology-based systems
- Increase awareness and understanding of ontology tools
- Work with a group of people on a common ontology
- Apply ontologies to real-world applications, especially eBusiness

To date, 23 members have contributed to the UBL Ontology Project:

Patrick Cassidy	Paul Murray	Adam Pease
Michael Daconta	Farrukh Najmi	Lisa Seaburg
Dean Black	Bo Newman	Norma Slattery
Peter Elkin	Duane Nickull	Bob Smith
Sam Hunting	Leo Obrst	Michael Uschold
Shiang-Yu Lee	Jack Park	Peter Yim
Bill McCarthy	Sue Probert	John Yunker
Tim McGrath	Marion Royal	

2.2 Expected Relationship to and Impact on UBL Schemas

The UBL schemas represent a starting point for the formalization process. The resulting ontology is expected to extend and formalize UBL English definitions and formalize relationship semantics (both hierarchical and non-hierarchical relationships).

It is hoped that the resulting UBL-based ontology will assist the UBL project in expressing the semantics inherent in complex business processes and contexts. Specifically, the Ontolog team hopes to provide “early warnings” and “leading indicators”, as appropriate, to support the formalization of the UBL Context Methodology.

In the event that the Ontolog team “gets stuck”, it plans to look beyond the UBL schemas, but not extend the conceptual scope of the project. The reuse of existing upper ontologies, as described later in this report, is one method for controlling this type of scope creep.

It is possible that a need to look beyond UBL schemas will point to specific UBL modeling issues (e.g., gaps or sub-optimizations). In this case, the Ontolog team will work to formulate actionable feedback and recommendations, as per our charter.

The Ontolog team hopes to develop an ontology that is an accurate modeling of the UBL domain and results in some level of validation, acceptance, approval, or adoption by the UBL committee, as appropriate.

Ontolog UBL Ontology Project Status Report – 2003-04-28

2.3 Project Goals

- Leverage the expertise of the UBL community to complete and validate our ontology work.
- Facilitate cross-pollination between the ontology and XML standards communities. Specifically, identify methods for bringing increased semantic formalization to markup systems.
- Bring together the academic and business (implementation) ontology development communities.
- Provide Ontolog members an opportunity to demonstrate an ontology development process and the business value of formalized ontologies.
- Provide Ontolog members an opportunity to learn about building ontologies and other semantic web concepts.
- Develop useful models for the management of virtual projects and ontology development efforts.
- Develop and demonstrate metrics for the creation and use of business ontologies.

2.4 Technical Goals

- Leverage as much of the UBL committee's work, as possible (don't reinvent the wheel).
- Leverage open source processes, technologies, and philosophy.
- Map to multiple upper ontologies.
- Demonstrate multiple tools and methodologies, where there is sufficient interest and resources.
- Implement a real-life, public-domain application in parallel with the development of the ontology. This should be one which uses the ontology in a non-trivial way to illustrate the reasoning capability that the ontology is intended to enable.

2.5 Project Management Strategy

The team has gravitated towards an iterative project management model, as opposed to more formal engineering models. Thus, a number of project activities that might otherwise be handled in series are being worked in parallel: scoping, process definition, technology selection, and modeling. While this approach brings with it the risk of having to repeat activities based on newly-articulated requirements and drivers, the team perceives the risk as slight, especially in light of the considerable expertise that a number of members bring to the team.

The iterative model has allowed consensus on various methodology and technical considerations to emerge more quickly than might otherwise have been possible with a more formal engineering approach. Generally, the team is finding that many such issues can be decided by the consensus of a very small subset of the active participants.

On the other hand, iterative projects typically require a higher level of ongoing effort to monitor and maintain alignment among project activities. To date, no identifiable alignment issues have surfaced with regard to the technical aspects of the project.

Consensus around the project goals and drivers has been somewhat slower to emerge, however. Early discussion around project goals tended to focus, instead, on technology options. The current task to articulate use cases has served as a lightning rod for issues of project scope and the relationship of the Ontolog project to the UBL effort. Weekly conference calls have proved invaluable for driving shared understanding and agreement in this area.

Ontolog UBL Ontology Project Status Report – 2003-04-28

On balance, the iterative model is proving to be well-suited to the type of demonstration project that we are currently engaged in. That being said, it is foreseeable that a more formal development model may be called for at some point in the future.

As this is considered to be a demonstration project, the project team has put considerable emphasis on knowledge transfer and sharing among the participants. In addition to knowledge sharing associated with the technical aspects of the project the team has also provided internal training on such topics as communications infrastructure (Wikis, purple numbers, VNC) and how to document use cases.

The focus on training is expected to increase considerable in the very near future. The team has just started to look at ways of distributing the work to formalize the UBL schemas. This is almost certain to result in the development of a process for getting non-experts comfortable enough with upper ontologies, ontology representation languages, and ontology tools to be able to contribute to the formalization process. As exposure to the practical aspects of ontological engineering was one of the major drivers for many to join the Ontolog Forum, it is likely that additional individuals will choose to participate in the UBL Ontology Project as the training becomes available.

2.6 Methodology

A number of alternatives were considered by the team:

- Guarino & Welty's OntoClean/Methontology methodology
- Noy and McGuinness' Ontology 101
- Use of a Wiki to support the initial capture of concepts

A modified form of the Ontology 101 guidelines were chosen. The Ontology 101 process is less complicated than the Methontology/OntoClean methodology, which tries to take a more formal engineering approach. The team decided to adopt a subset of the methodology and generalize it to avoid becoming tool specific:

- Step 1. Determine the domain and scope of the ontology
- Step 2. Consider reusing existing ontologies
- Step 3. Enumerate important terms in the ontology
- Step 4. Define the classes and the class hierarchy
- Step 5. Define the properties of classes
- Step 6. Define the additional properties related to or necessary for properties (i.e., cardinality, bidirectionality/inverse, etc.)
- Step 7. Create instances
- Step 8: Create axioms/rules

3.0 Project Status by Activity Area

3.1 Technology Selection (ongoing)

The following technologies have been considered for use in the project:

- Established Ontologies
 - DOLCE (a Descriptive Ontology for Linguistic and Cognitive Engineering)
 - Mikrokosmos
 - Open CYC
 - SENSUS
 - SUMO (Suggested Upper Merged Ontology)
- Representation Languages
 - Common Logic
 - DAML+OIL / DAML-S
 - ISO-IEC-11179 classification standard
 - KIF / SKIF
 - LISP/LOOM
 - OWL
 - PAL (Protoge Axiomatic Language)
 - RDF/S
 - UML
 - XML DTD and Schema
- Tools
 - Chimaera
 - IODE from OntologyWorks
 - JESS (Java Expert System Shell, based on CLIPS)
 - MS Enterprise Architect
 - OilEd
 - Ontolingua / OKBC API
 - Prolog engines (XSB, Amzi!, binProlog)
 - Protégé
 - Protégé OntoViz plug-in
 - SNARK
 - Teknowledge DAML generator
 - Use ebXML Registry as an ontology/terminology server
 - UML-ORM

SUMO has been selected as the Upper Ontology to be used for the initial work. There are no significant IP issues to deal with and it has been ported to Protoge. Further, we have the commitment of Adam Pease to make changes if our UBL formalization work points to problems with SUMO.

OpenCYC appears to be the leading upper ontology alternative.

Consensus has not been reached regarding the choice of representation languages and tools. Consensus seems to be emerging around the use of KIF and Protégé, but there still appears to be some interest in doing portions of the work using less formal representation schemes. Mike Deconta is leading deliberations in this area.

3.2 Determining Domain and Scope (ongoing)

The team has decided to focus, initially, on a small number of UBL doctypes:

- Purchase Order, led by Bill McCarthy.
- Invoices, led by Adam Pease.
- Shipping Documents, led by John Yunker.

The team is relying on use cases to define functional and semantic scope. The development of use cases has been underway for a few weeks and there is considerable work left to do in this area. Three use cases have been selected for definition:

- Automated reconciliation of Purchase Orders and Invoices, being worked by Mike Daconta and Bill McCarthy.
- Identifying the proper structure of an address from context, being worked by Sue Probert and Peter Yim. For example, what is needed to present the proper address formats when a purchase order from a UK buyer is placed with a supplier in Japan?
- Mapping between different standards and representation languages, being worked by Sue Probert, Peter Yim, and John Yunker. Thus far, the participants noted a functional difference between mapping UBL to two other standards and using UBL to map different standards. Both functional models are being considered.

The team is finding that discussion of the use cases is triggering more general discussions of project scope, which is to be expected of the iterative project management strategy which was chosen. While the decision to limit scope to UBL (as opposed to including all XML-based standards efforts) still holds, early analysis indicates that the UBL and Core Component data architectures, in reality, comprise approximately four different ontologies that will need to be formalized. For the sake of convenience, these potentially multiple ontologies are still being referred to by the singular phrase “UBL Ontology”.

3.3 Considering the Reuse of Existing Ontologies (complete)

The team has decided to start their work by mapping UBL structures (concepts) to an upper ontology. Mapping to an upper ontology is seen as a starting point for formalization. It is expected to reduce the number of concepts which need to be defined and thereby avoid duplication of effort.

As was described in the Technology Selection section, SUMO has been selected as the upper ontology to be leveraged for this project.

3.4 Enumeration of important terms in the ontology (commencing)

Activities in this area, to date, could best be described as exploratory.

The team commenced analysis of the existing UBL models, aspects of the Core Component architecture, and Context Drivers. For example, in March, there was considerable discussion around the modeling of roles in the UBL trading cycle.

An initial mapping between UBL Invoice structures and SUMO has been completed.

Ontolog UBL Ontology Project Status Report – 2003-04-28

With regard to Purchase Orders, a preliminary mapping between xCBL and the SUMO Business Process ontology has been completed. From that exercise, it has been concluded that the relevant categories of the SUMO ontology are likely to be Financial and E-Commerce.

Adam Pease has articulated a set of proposed guidelines that detail how to map UBL constructs to SUMO concept classes.

The team has recently started to deal with the appropriateness / need to map elements of the Core Component architecture to SUMO.

The conversation is starting to shift to mechanisms for engaging others in this task so that the work can be better distributed.