Social and Personal Factors in Semantic Infusion Projects

Patrick West[1] (westp@rpi.edu), **Peter Fox**[1] (pfox@cs.rpi.edu), **Deborah McGuinness**[1,2] (dlm@cs.rpi.edu), **Stephan Zednik**[1] (zednis@rpi.edu) ([1]Rensselaer Polytechnic Institute 110 8th St., Troy, NY, 12180 United States), ([2]McGuinness Associates 4 Shaker Bay Rd., Latham, NY, 12110 United States)

Abstract

As part of our semantic data framework activities across multiple, diverse disciplines we required the involvement of domain scientists, computer scientists, software engineers, data managers, and often, social scientists. This involvement from a cross-section of disciplines turns out to be a social exercise as much as it is a technical and methodical activity. Each member of the team is used to different modes of working, expectations, vocabularies, levels of participation, and incentive and reward systems. We will examine how both roles and personal responsibilities play in the development of semantic infusion projects, and how an iterative development cycle can contribute to the successful completion of such a project.

Knowledge Infusion

Interdisciplinary support: Virtual Solar-Terrestrial Observatory (VSTO) science domains are solar physics, space physics, and solar-terrestrial physics. Major communities include those interested in solar images from the Mauna Loa Solar Observatory (MLSO), and the NSF-funded Coupled Energetics and Dynamics of Atmospheric Regions (CEDAR). The BCO-DMO project (Biological and Chemical Oceanographic Data Management Office) at WHOI (Woods Hole Oceanographic Institution) domain encompasses the ocean sciences. These collections provide a good focus for virtual observatory work since the datasets are of significant scientific value to a set of researchers and capture the challenges inherent in complex, diverse scientific data.

Use Cases: Use cases are used to indicate a specific capability that drives both what knowledge is to be represented and used and also what software and interfaces are built for the user and to the underlying data. It contains functional and non-functional requirements, success and failure scenarios, the vocabulary used within the research domain of the user developing the use case (typically a research/domain expert) and domain specific research practices.

Ontologies: Ontologies created in OWL define concepts, relations, terms, etc…, in order to utilize their precise formal definitions for semantic search and interoperability. Use case sentences are examined to identify initial concepts, and relations between them. Hierarchies became apparent, and important properties, as well as restrictions on the values for certain concepts within a given context. Initial ontologies are as close to the vocabulary of the user as possible. With additional iterations, new concepts can be introduced by the ontologist, other ontologies utilized.

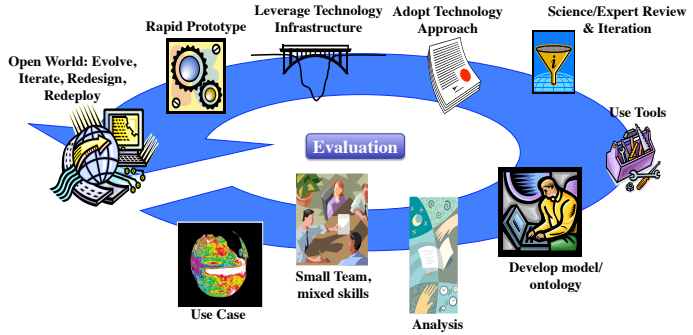
Experience

Collaboration: Extensive engagement of the end user; the domain scientists, students, and professionals; enables explicit semantic interoperability. In order to provide a scientific infrastructure that is usable and extensible, VSTO and BCO-DMO required contributions concerning semantic integration, and knowledge representation while requiring depth in each of the science areas. In developing and analyzing use cases within the iterative approach, a small team made up of a facilitator (knows iterative methodology), domain experts (knows resources, data, applications, tools), ontology modelers (to extract objects/relations), software engineers (architecture and technology), and a scribe (capturing everything discussed), should be utilized.

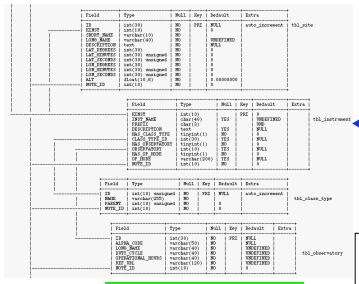
Social: An implication of adding semantics is the strong need and role for domain literate members of the team to develop and then vet the knowledge encoding. This turns out to be a social exercise as much as it is a technical and methodical activity since the team comprises people from multiple disciplines, and can even include social scientists. Social considerations also affect how projects are sustained.

Personal: Personalities are important to consider. The facilitator must be aware of and accommodate many dimensions of participation if the infusion activity is to be successful. Team members are used to different modes of working, vocabularies, and incentive and reward systems. One result of this is that it is very difficult to conduct the infusion simply as a remote or virtual activity unless it is substantially founded on face-to-face meetings where personal and social forces come into play.

Iterative Design and Development

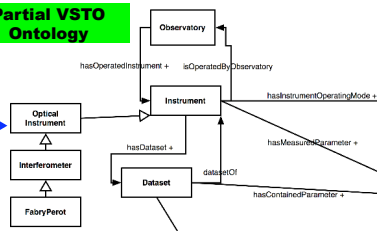


Catalogs/Ontology Integration



CEDAR Catalog Relationship Diagram

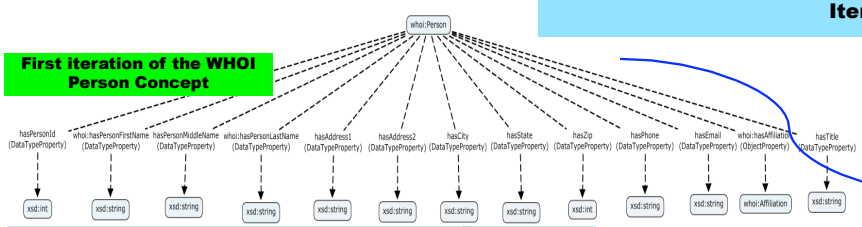
Partial VSTO Ontology



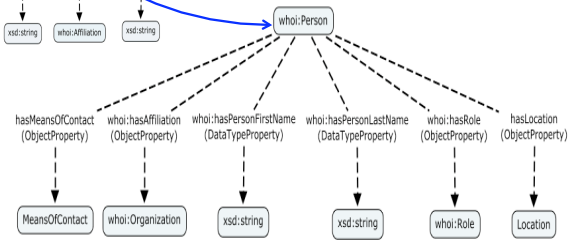
You can see here how many of the tables and relationships in the CEDAR catalog relate directly to classes and properties in the VSTO ontology. The work in CEDAR has helped to define the ontology for VSTO, and the work on the ontology for VSTO has helped to define a better Catalog for CEDAR.

Iterative Ontology Development and Understanding

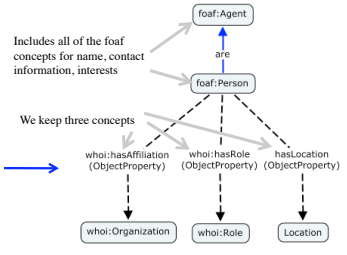
Example where the iterative process helped to develop an understanding by WHOI domain experts ontologies and translating their concepts into an ontology and the ontology developers to understand the specific domain vocabulary. Successive iterations helped to expand and simplify concepts and incorporate already existing ontologies. Similar in instrument, platform, parameter ontology development.



First iteration of the WHOI Person Concept



Second iteration of the WHOI Person Concept



Third iteration of the WHOI Person Concept

Sponsors:

National Science Foundation

Glossary:

VSTO - Virtual Solar-Terrestrial Observatory
MLSO – Mauna Loa Solar Observatory
CEDAR - Coupled Energetics and Dynamics of Atmospheric Research
WHOI – Woods Hole Oceanographic Institution
BCO-DMO – Biological and Chemical Oceanographic Data Management Office

Acknowledgments:

Cyndy Chandler, Dickey Allison, Robert Groman and Christine Hammond of WHOI for their considerable contributions.