APPROVAL SHEET

Title of Thesis: **Applying Ontologies and Semantic Web technologies to Environmental Sciences and Engineering**

Name of Candidate: **Viral Parekh**
Master of Science, 2005

Thesis and Abstract Approved:

_____

Dr. Jin-Ping (Jack) Gwo
Assistant Professor
Department of Civil and Environmental Engineering

_____

Dr. Timothy Finin
Professor
Department of Computer Science and Electrical Engineering

Date Approved: _____

CURRICULUM VITAE

Name: Viral Subodh Parekh.

Permanent Address: 4817 Grenville Square, Baltimore MD 21227.

Degree and Date to be conferred: M.S., 2005.

Date of Birth: 10th October, 1979.

Place of Birth: Mumbai (Bombay), India.

Collegiate Institutions attended:

- University of Mumbai, India

  Bachelor of Engineering (Computer Science)

  June 2001.

- University of Maryland, Baltimore County

  Master of Science (Computer Science)

  May 2005.

Major: Computer Science.

Professional Publications:

- V. Parekh, J. Gwo and T. Finin, "Ontology based Semantic Metadata for Geoscience Data", Proceedings of The 2004 International Conference of Information and Knowledge Engineering, Las Vegas, June, 2004.
- V. Parekh, J. Gwo and T. Finin, "Mining Domain Specific Text and Glossaries to Evaluate and Enrich Domain Ontologies", Proceedings of The 2004 International Conference of Information and Knowledge Engineering, Las Vegas, June, 2004.

ABSTRACT

Title of Thesis: Applying Ontologies and Semantic Web technologies to Environmental Sciences and Engineering

Viral Parekh, Master of Science, 2005

Thesis directed by: Dr. Jin-Ping (Jack) Gwo
Assistant Professor
Department of Civil and Environmental Engineering

Dr. Timothy Finin
Professor
Department of Computer Science and Electrical Engineering

The complexity and diversity of knowledge and terminology within environmental sciences and engineering is one of the key obstacles for successful interdisciplinary studies. Relevant data is difficult to locate and retrieve primarily due to varying formats, schemas and semantics. For example, for a typical modeling assignment a researcher needs to acquire knowledge of individual computational models, search, gather and analyze raw data, ensure the high quality of data, transform the data into formats compatible to the computation models that he or she is to use and then finally perform the modeling. This process takes several days to months.

To address these problems, we propose to use ontologies and emerging Semantic Web technologies. Ontologies provide shared domain models that are understandable to both humans as well as machines. We used the Web Ontology Language (OWL) to define ontologies with the objective of improving data sharing and integration. These ontologies define several domain concepts and describe a variety of domain models being used within environmental sciences and engineering. Metadata ontology is developed to define every facet of environmental datasets. Its aim is to provide

a conceptual schema for the dataset using the available domain ontologies. The overall goal is to achieve content based retrieval of datasets and integration of heterogeneous data. We demonstrate a few applications which use the developed ontologies to solve common environmental problems. Our results suggest that ontologies and Semantic Web technologies like RDF and OWL may provide the much needed semantics within these diverse domains of environmental sciences and engineering, and hence may serve as the building blocks for innovative solutions to existing problems.

APPLYING ONTOLOGIES AND SEMANTIC WEB TECHNOLOGIES TO ENVIRONMENTAL SCIENCES AND ENGINEERING

by
Viral Parekh

Thesis submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Master of Science
2005

Dedicated to
my beloved parents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Jack Gwo, my graduate advisor for guiding and motivating me from the very inception of this thesis. Apart from being a fantastic advisor, Dr. Jack Gwo has been a great mentor and a friend. I would also like to thank Dr. Tim Finin for his valuable suggestions, advice and feedback during the course of this thesis. I am also thankful to Dr. Yaacov Yesha for being a member of my MS thesis committee. I would also like to take this opportunity to thank the staff members of the Graduate School and the Department of Electrical and Computer Engineering for helping me with the administrative details concerning the thesis.

I would like to thank my family and friends for believing in me and encouraging me to perform to the best of my capabilities. My journey to USA with the aim of pursuing a Master's Degree in Computer Science would not have been successful without their support and inspiration.

TABLE OF CONTENTS

Chapter I

INTRODUCTION

A. Problem Description

In the post-9/11 era, environmental security becomes one of the paramount issues that require effective and efficient coordination and information sharing among government agencies and their industrial counterparts. For example, the discovery of lead contamination in the drinking water of Washington metro residences in early 2004 highlights the serious under-preparedness of industrial facilities to immediately address the problem and the lack of coordination among government agencies to identify possible solutions. It also reveals the ineffective use of the state-of-the-art information technologies that may greatly reduce the anxiety of local residences and increase the credibility of government agencies and the industry. In essence, it is a prime example of failure in information management and utilization leading to inadequate decision-making.

The effectiveness and efficiency of the usage of environmental data greatly depends on the domain-specific understanding and empirical experience of the users in various application domains, e.g., hydrology, environmental engineering and so on. The complexity and diversity of domain knowledge and terminology is one of the key obstacles for successful interdisciplinary studies. There is a vital need of an efficient mechanism for discovery and uniform integration of relevant datasets.

Huge volumes of geo-scientific and environmental sciences data are available and used by researchers, students, industries, government agencies, etc for different reasons such as data analysis, environmental modeling, etc. The data space for environmental problems is not very easy to define nor is the data easy to collect and interpret. There are several public data providers such as US

Government agencies like Environmental Protection Agency (EPA), United States Geological Survey (USGS), National Oceanic & Atmospheric Administration (NOAA), National Aeronautics and Space Administration (NASA), etc and other non-profit organizations like National Center for Atmospheric Research (NCAR). They produce a variety of data which is archived at various locations and distributed in different formats. The data can be daily weather forecast data, geological data about rocks or soils, geographical data, geochemical data, hydrological data, etc. The datasets are distributed and stored by various organizations making the task of locating and retrieving the relevant datasets very complex. The heterogeneity of data formats leads to data interoperability and data usability problems faced by the users of these datasets. The end users of these geoscience datasets could be researchers searching for relevant data to perform certain experiments or modeling tasks, people from industries looking for right data in order to facilitate decision making or even students in search of data for their class projects. Occasionally, data from different domains need to be used in order to perform certain modeling and analyses. For example, spatial data may need to be combined with chemical species data in order to perform certain hydrogeological studies. However, integrating and using these data sets can be very difficult. This is primarily due to the fact that each data set has different format, schema and semantics. Current methods for searching and retrieving data sets for use are extremely cumbersome. Heavy use of web technologies has led to this data being available on the web. Several modeling tools are also available online and this data is used by different tools to perform analysis. Information technologies are being widely used to address the above problems and provide efficient ways to support decision making and management of data [4, 5, and 7].

B. Our Approach

The Semantic Web provides a common framework that allows data to be shared and reused across applications, enterprises, and community boundaries. Its well-defined data semantics enable computer agents and humans to work in cooperation [3]. Recent efforts in the World Wide Web Consortium

(W3C) to implement Semantic Web [6] have spurred interest in the use of ontologies for information modeling and knowledge representation. Ontologies provide shared domain models that are understandable to both humans and machines. They describe a set of concepts and relationships between them. Ontologies provide a controlled vocabulary of terms that can collectively provide an abstract view of the domain [1, 2]. Such a shared understanding of the domain greatly facilitates querying of data and increases recall and precision. Semantic Web technologies and ontologies are being used to address data discovery, data interoperability, knowledge sharing and collaboration problems. Software agents can then be used to construct and provide dynamic services on the web.

We propose to use the emerging semantic web technologies and ontological framework to solve the data usability and information discovery problems for environmental sciences and engineering. In this research we develop several domain ontologies and demonstrate their usefulness through applications that highlight the role of ontologies in future intelligent environmental information systems. These technologies provide gains in information systems interoperability and promise more efficient data sharing and data integration. This will provide a basis to solve the data heterogeneity and data interoperability problem faced in environmental sciences and engineering. Our ultimate vision is to build intelligent and powerful environmental information systems by developing information infrastructures that may enable the deployment of efficient data sharing and integration mechanisms. The domain ontologies we developed describe concepts and relationships pertinent to environmental sciences and engineering. We have covered areas like geography, geology, geochemistry, hydrology to name a few. Apart from the domain ontologies, we developed another ontology especially to provide semantic metadata and conceptual schema for domain datasets. This machine understandable metadata will help software agents to reason about the content of the datasets using the domain ontologies and hence lead to better data discovery and data integration mechanisms.

To demonstrate the potential and need of using ontologies and Semantic Web technologies for environmental sciences and engineering, consider the following applications:

1. Say a research scientist wishes to model groundwater contamination for an area in Baltimore, MD. Without the infrastructure provided by the Semantic Web, this researcher needs to acquire knowledge of individual groundwater computational models, gather and analyze raw data, ensure the high quality of data, transform the data into formats compatible to the computation models that he or she is to use. He usually uses the models that he is familiar to or one among these that the thinks is most appropriate. He may also ask around and go to a training to learn to use a new model that is required or recommended by a regulatory agency. *In contrast, the proposed system which uses the emerging Semantic Web technologies will allow the scientist to query the knowledge base for different modeling programs that fit his requirements.* He either has prior knowledge of these data requirements or needs to obtain the knowledge from the training or to read the documentation of the computer model. *The model semantic metadata informs the scientist about the various kinds of data and measurements that are needed in order to perform this modeling.* As of today, there is not a single good way to collect data. Data sets are collected by the user himself or by another person. So, he will have to transform the data into the format that the models require or acquire the data from another person, mostly by communicating with the other person and by requesting that the other person gracefully documents the data and data collection process. He may also obtain the data from open literatures, but he still needs time to understand the contents of the datasets, making sense of them, and most importantly, make sure the quality of the datasets is good. This last step is the most difficult one if he always gets data from another group. *In contrast, the semantic knowledge base may provide him with a collection of the available datasets that could be useful based on the model requirements and dataset characteristics. The semantic metadata of the datasets would not only give the user a better idea about the content of the datasets but since it being machine understandable, software agents can now make sense of the different datasets and their content. The information within this metadata*

would provide details of every aspect of the dataset. This highlights the immense potential of the semantic approach. Equally important, in case the required datasets for running the computer model are not available, the system can query the knowledge base for other models which could be useful. These new models would be chosen by the system such that they can produce data required for the original chosen model. In this way, the research scientist can use a chain of models in order to achieve the original task. This is frequently done by scientists, but at present the chain of models is limited to those that the scientist has knowledge of. With the semantic modeling framework, the entire process may be simplified and automated using the semantic knowledge of models and datasets. In the present scenario, this process would require several days of searching on the web or phone calls to several agencies. Apart from this, the other problem would be interoperability of the various datasets from different models. Our use of ontologies and semantic web technologies provides a starting point, if not a complete solution for such problems.

2. Engineers sometimes need merely information to conduct preliminary studies, e.g., feasibility of restoring a heavy-metal contaminated industrial site, to support cost-benefit analysis and decision making. In these cases, an engineer needs to know soil composition details, groundwater details, various chemical specifications, geological details, etc for that particular area. The data gathering process is the most critical and time consuming process here which can take days to weeks. He will need to do a lot of house keeping, data quality analysis, organizing data into areas of particular concerns, select data that are needed for his tasks and finally do the data analysis. *With an ontology based system, several of these tasks can be automated. The engineer can perform ontology level searches against the knowledge base of semantic metadata for the environmental domain data.* This is much more powerful compared to traditional keyword based searches as the user can perform attribute level searches, yielding accurate results. For example, the engineer here would be able to search for specific datasets which meet certain geographic and temporal criteria, dataset format and software constraints, source of dataset, etc. The ontologies would

provide a conceptual view of the entire data. They would also facilitate integration of different datasets to allow uniform data analysis.

3. A geochemist may want to study how different molecules behave when they are mixed together. The knowledge base stores information about various chemical molecules and also various chemical reactions involving these molecules. There may be metadata of several chemical modeling programs present in the knowledge base too. Now the scientist may select molecules of his choice by querying the knowledge base. He may then see the several chemical reactions involving these molecules. He may proceed by specifying the concentration levels for his selected molecules. He may want to use this existing knowledge for several applications. The present conditions would require him to convert this knowledge in different formats depending on the need. However, with ontologies and semantic web technologies a uniform view of data is created. Heterogeneous vocabularies are made compatible via ontologies and multiple conceptual dimensions become queryable simultaneously.

C. Motivation

Environmental systems demand semantics for dataset descriptions and the actual data content. Lists of datasets maybe available with little known information apart from file names, certain identification and distribution details. Often times, there is not much available information that may be used to determine the usefulness of a particular dataset for a specific model before actually downloading and looking at the dataset. Semantic Metadata for scientific information is needed to help scientists and other users make decisions about the data available for their research.

FGDC (Federal Geographic Data Committee) Content Standard for Digital Geospatial Metadata [15] was developed in 1994 to describe all possible geospatial data. However, the standard is very complex with 334 different elements, 119 of which exist only to contain other elements making this

standard difficult to use. Moreover, the standard provides text based syntactic metadata with virtually no semantics and machine understandability when compared to the proposed ontology based semantic metadata. Although the standard is complex, it is quite comprehensive and it describes all possible facets of geospatial data. This standard has been developed by USGS and hence it is currently widely used to provide metadata for geospatial datasets.

We see the following motivating factors for using ontologies and semantic web technologies for environmental information systems:

- Ontologies can provide a shared, common vocabulary for environmental sciences and engineering. A semantic and standard knowledge base can be extremely useful to all environmental disciplines.
- Interoperability among heterogeneous datasets can be achieved by using shared machine understandable conceptual structures within ontologies.
- Ontologies can provide a conceptual schema for any dataset regardless of its format, structure, complexity or size.
- Ontologies can be used as a basis for content based discovery and retrieval of datasets
- Semantic Web technologies like RDF [35, 36] and OWL [34] are current W3C recommendations and future standards for data description and ontology encoding.
- The standard framework of RDF and OWL makes different ontologies and knowledge of other domains readily available to our environmental science domain.
- Ontologies can provide semantic descriptions for a variety of modeling programs used in this domain. Every aspect of the modeling program can be captured and stored as knowledge for that model.
- Ontologies support reusability; this allows same ontologies to be used for multiple applications.

- Ontologies are viewed as the most advanced knowledge representation model.
- Ontologies support inferencing, this allows new derived knowledge to be generated from existing knowledge.
- Ontologies provide the much needed reasoning power over existing knowledge through various user defined rules.

The goal of this research is to increase efficiency in data interoperability and data integration among heterogeneous environmental data sources using explicit, machine understandable ontologies encoded in new web standard languages, particularly the Semantic Web. Organizations like EPA and USGS maintain vocabularies describing the terminology in environment sciences. However, they need to develop standards for terms to provide a common understanding. With this thesis we develop a suite of domain-specific ontologies for environmental sciences and engineering. Particularly, we use glossaries, domain dictionaries, and emerging Semantic Web languages such as RDF and OWL to develop these ontologies. The objective is to provide semantic interoperability among heterogeneous environmental data sources by using these semantically rich ontologies, thereby facilitating and assisting in the detection, evaluation, and effort coordination that may lead to effective decision-making and resolution of imminent environmental problems.

In Chapter 2 we review some of the existing work related to our research. Here we see how different domains have adopted the ontology based approach to tackle a variety of problems. In Chapter 3 we present a detailed review of the technologies being used for ontology development. In this same chapter, we discuss our methodology for ontology development. We present several questions that we believe should be answered by the resulting ontologies. Later, in Chapter 4 we give a detailed description of the ontologies we developed, specifically **Environmental** ontology, **Molecule** ontology, **Metadata** ontology and the **Models** ontology. Chapter 5 discusses certain applications

which use the developed ontologies. In Chapter 6 we discuss certain aspects of this research. Finally we summarize and conclude in Chapter 7.

Chapter II

RELATED WORK

In this chapter we discuss similar or related approaches taken by researchers. We will see how ontology based approaches are being widely used with the emergence of Semantic Web.

GeoSemantic Web [10, 11] makes a sincere effort towards development of geographic ontologies for geospatial applications. Their aim is use this ontology based approach to associate geographically referenced data to any other non-spatial information related to geographic features expressed on the Semantic Web. Their ontologies will provide geographic references to well known locations and will also include geographic relations such as topological, direction and distance relationships. Seamless integration of geographic information with other information based on its semantic content regardless of its representation has also been utilized for GIS (Geographic Information Systems) [12, 13]. GeoSemantic Web supports a minimal translation of OpenGIS [14] official specification to the geographic ontology. There are also research efforts in creating ontologies [19, 20] based on different ISO standards of geographic information for web based simulation of hydrodynamic models.

Earth System Grid [16, 17] is an ongoing project in several national labs within US wherein the major goal is discovery and secure access to large datasets for earth sciences research. They have developed metadata schemas in the form of rich ontologies which are used to describe the datasets. The aim is to help scientists to efficiently search and retrieve information, manage data, record their observations, and perform other scientific tasks using the retrieved data.

Semantic Web for Earth and Environment Terminology (SWEET) is a major effort within NASA [18]. They are in the process of developing ontologies and a semantic framework for various earth science initiatives. Several ontologies have been developed covering concepts such as earth realm, physical substances, living elements, physical properties, units, numerical entity, temporal entity, spatial entity, phenomena and human activities. An ontology aided search tool is implemented which uses these ontologies to find alternative search terms.

In [21], the researchers have developed hydrologic ontologies and a few tools based on these ontologies to facilitate creation of semantic metadata specifications for hydrologic datasets. Several OWL ontologies are being used to extend standard metadata and hydrologic thesauri to specify metadata that conforms to the specific needs of a hydrologic information community. These ontologies are more or less based on Geographic Metadata Standards FGDC-STD-001-1998 which is a US standard and ISO 19115-2003 with its related ISO/TC 211 19100 series standards.

For the earthquake science community, the authors in [6] propose to develop a data semantics based system to improve interoperability among heterogeneous earthquake data. Again the approach adopted is ontology based system to annotate observation and hypothetical data.

In most of the above mentioned works, the domain being researched is diverse and the domain data is heterogeneous and the interpretations of data differ from resource to resource, and scientist to scientist. The environmental science domain also suffers from similar problems of data heterogeneity. We believe ontologies and semantic web technologies have the potential to provide the required semantic interoperability. And moreover as the semantic web grows, this will make the environmental metadata marked up in RDF and OWL available for use and integration with other data and ontologies.

Chapter III

ONTOLOGY DEVELOPMENT PROCESS

A. Technologies

The Semantic Web provides a common framework that allows data to be shared and reused across applications, enterprises, and community boundaries. Ontologies provide shared domain models that are necessary for realization of Semantic Web. They describe a set of concepts and relationships between them.

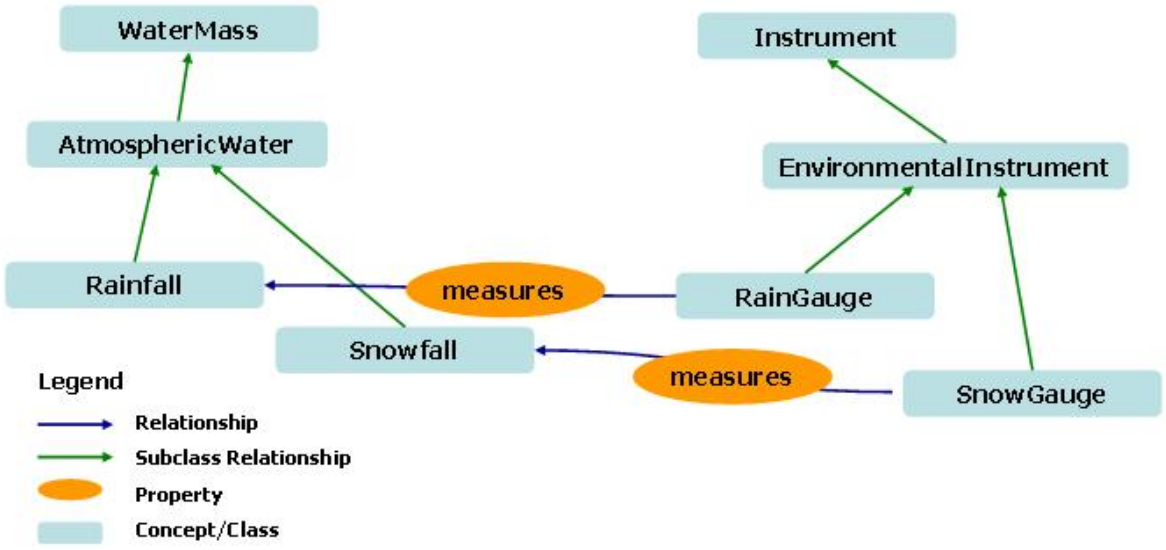

Figure 1. Simple ontology demonstrating concepts in Environmental Science

An example of a simple ontology might be the concepts of Rainfall, Snowfall, RainGauge, SnowGauge, AtmosphericWater and so forth that fall under the large domain of WaterMass, and relationships such as 'RainGauge measures Rainfall', 'SnowGauge measures Snowfall', 'Rainfall has a unit, value and date/time of measurement', 'Rainfall and Snowfall are types of AtmosphericWater'.

Figure 1 gives a schematic of the concepts and the relationships. This subsection of WaterMass ontology might be developed initially for a particular application, such as a weather prediction system. As such, it may be considered similar to a well-defined database schema. The advantage to an ontology is that it is an explicit, first-class description of the knowledge. So having been developed for one purpose, it can be published and reused for other purposes. For example, a modeler may use the Rainfall and Snowfall data as building blocks for climate modeling. Alternatively, the data can be used to suggest a precipitation pattern.

Ontologies serve as a means for establishing a conceptually concise basis for communicating knowledge across a large community. A key component of semantic web will be the annotation of web resources with metadata that describes their content, with ontologies providing a source of shared and precisely defined terms that can be used in this metadata. RDF (Resource Description Framework) [35, 36] and OWL (Web Ontology Language) [34, 38], both being W3C recommendations, are now widely used to encode ontologies and knowledge bases. Several techniques are being developed to automate or semi-automate construction of huge ontologies from domain specific corpus, glossaries/dictionaries, etc [31, 32, 33]. Simple ontologies like FOAF [9] & Dublin Core [8] in use today. Cyc is large, general purpose ontology with more than 100k terms [51]. WordNet Task Force [49], part of the Semantic Web Best Practices Working Group (SWBPD) [6], has an aim to deploy WordNet [12] and similarly structured lexica into RDF/OWL. Goldbeck et al. [48] present an OWL ontology for the National Cancer Institute (NCI) Thesaurus. These advancements, as a part of Semantic Web activity, will produce a rich set of lexical and domain vocabularies in the form of ontologies and knowledge bases in the near future. The following provides a brief review of the semantic web technologies, including RDF, RDFS, OWL and the tools developed around them, e.g., Protégé OWL plugin and Jena.

RDF (Resource Description Framework) [35, 36] provides a flexible graph based model which is used to describe and relate resources. An RDF document is an unordered collection of statements; each with a subject, predicate and object (triples). These statements describe properties of resources. Each resource and property can be identified by a unique URI (Uniform Resource Identifier) which allows metadata about the resource to be merged from several sources. RDF has a solid specification and it is widely being used in a number of standards. It provides a common framework for expressing information, so it can be exchanged between applications without loss of meaning. RDF Schema (**RDFS**) adds taxonomies for classes and properties. It allows expressing classes and their relationships (subclass), and defining properties and associating them with classes. It facilitates inferencing on the data based on the hierarchical relationships.

OWL (Web Ontology Language) [34, 38] provides extensive vocabulary along with formal semantics and facilitates machine interpretability. OWL is much more expressive than RDF or RDFS, allowing us to build more knowledge into the ontology. For example, cardinality constraints can be imposed on the properties of an OWL class. For example, we can restrict a Molecule class to have just one chemical formula whereas it can still have multiple names. Moreover, OWL is designed as a specific language to define and describe classes and properties within ontology. It has many predefined functionality built-in. For example, an ontology can import other ontologies, committing to all of their classes, properties and constraints. There are properties for asserting or denying the equivalence of individuals and classes, providing a way to relate information expressed in one ontology to another. These features, along with many others, are important for supporting ontology reuse, mapping and interoperability. Also, OWL is a standard and is supported by the standard organization W3C. We have used the Web Ontology Language OWL to define the ontologies and RDF is used to describe the actual instances within the knowledge base. As the semantic web grows, more and more ontologies will be available in OWL. There will be a wide variety of development tools available for integrating the different OWL ontologies and doing intelligent reasoning. OWL provides 3 sub-

languages with increasing levels of expressiveness and complexity. They are OWL Lite, OWL DL (Description Logics) and OWL Full. We have used mostly OWL Lite due to the reason that it has lower formal complexity than DL and Full. Also it fits our purpose well. More tool support is currently available for OWL Lite than others. A comprehensive list of OWL Lite language constructs is available at [38].

Protégé OWL plugin [39, 40] was mainly used as the development tool for ontologies. It is a widely used GUI ontology editor for the Semantic Web. It allows editing and visualizing classes and properties, and defining relationships among them. Protégé OWL plugin also provides description logic classifier which helps restructure the ontology. It also provides OWL validation. ezOWL plugin [41] was used for ontology visualization. This is another widely used visual ontology editor for OWL. Jena [42] provided the programming environment for automating the generation, storage and retrieval of knowledge base based on the developed ontologies. Jena is the most popular and widely used Java framework for Semantic Web applications. Its architecture centers on RDF graphs. It provides a rich API (Application Programming Interface) for RDF, RDFS and OWL. RDQL (RDF Query language) [43] is used as the query language to retrieve data from OWL and RDF knowledge bases. It provides a way of specifying a graph pattern that is matched against the RDF graph to yield a set of matches. We have also used the Jena 2 persistence subsystem that provides transparent persistence for RDF models through use of a back-end database engine. This enables faster insertion and deletion of data. MySQL [44] is used as the database backend as it is open source and fits our purpose of storage and retrieval.

B. Methodology

Ontology development is a very critical step in defining the domain knowledge and there is no single correct methodology to do so. It is typically harder as there are no good metrics to evaluate or test the

ontology being developed. Ontology Development Guide 101 [25] provides a good starting point to learn about ontologies in general and their development. This guide clearly points out how ontology development is different from designing classes and relations in object-oriented programming. In object oriented programming, design decisions are based on operational properties of classes whereas in ontology design they are based on structural properties of classes. As a result, a class structure and relations among classes in ontology are different from the structure for a similar domain in an object-oriented program.

Briefly speaking, ontology development includes:

1. Clearly defining the domain concepts as classes in the ontology
2. Determining the relationships (both taxonomic and non-taxonomic) among these concepts/classes
3. Defining the properties of the concepts/classes
4. Determining the domain and range of the defined properties
5. Defining various class level and property level restrictions if required
6. Finally, creating the knowledge base by identifying the various instances of the defined concepts

Ontology development is a naturally iterative process. A design fundamental is to make the concepts and relationships in the ontology depict the real world as closely as possible. The structure and contents of the ontology should be based upon the inherent knowledge of the discipline, rather than on how the domain knowledge is used. Compound concepts should be decomposed into their component parts, to make it easy to recombine concepts in new ways. Moreover, community involvement should guide the ontology development process.

Glossaries/dictionaries are widely being used to develop ontologies. We referred to a few during the process of ontology development. The USGS Learning Web [27] contains several glossaries related to environmental sciences and engineering. Books such as the Geological Dictionary [29] provided an excellent source of domain knowledge. The US EPA has an extensive Environmental Information System that facilitates the query and search of important environment related databases, in addition to such publications as Terms of Environment [28]. Oak Ridge Labs Environmental Sciences Division [50] proved to be another good resource of information. For metadata, the FGDC (Federal Geographic Data Committee) Content Standard for Digital Geospatial Metadata [15] provides an excellent source of constructing metadata ontologies in geosciences. Above all, there are a lot more knowledge sources, in much unstructured forms, out there in scientific and engineering literatures and on the web.

In general, sources of knowledge such as the above-mentioned are widely acceptable and used by the community. Hence these vocabularies provide excellent starting point for our domain ontology development. Domain expertise and knowledge was also always available through interactions with a domain expert involved in this research.

We started with a rough outline of the ontology with a few concepts and properties. We revised and refined it over several iterations. Over time, several ontologies for different sub-domains of the larger environmental science domain evolved. During this entire process, several modeling decisions were required. The important goals were that the developed ontologies be generic, extensible and maintainable, and readily available for the future Semantic Web world.

There are libraries of ontologies available on the web [23, 24]. Ontologies support reusability and hence we checked whether there existed any ontologies that can be useful to model our domain scope.

There were a few; however none of them addressed the direct problem. We reused or extended a few ontologies that prove helpful towards our application and domain.

We have followed a combination of top-down and bottom-up development process. We started by defining the salient concepts first and then we either generalized them or defined specific cases. As new concepts were defined, we also started organizing them into a hierarchical taxonomy. Various attributes of the concepts were identified and moreover relationships between these classes were defined. Identifying and defining the domain and range of the properties is another critical step in ontology development. Domain of a property defines the set of classes that can have this property while range defines the allowed values for this property. As classes are arranged in hierarchy, the properties are inherited from super-class to subclass. Cardinality of the properties was also determined, defining the number of values that the property can have. We followed this process iteratively and after several iterations we had a set of ontologies that were applicable and useful to our domain and applications. Once we developed the ontologies, our next goal was to identify the instances of the classes defined. This will help us build the knowledge base for our domain. Several datasets, online resources and vocabularies/glossaries were examined and automated mechanisms were developed in order to convert some of these resources into instances in our knowledge base. The coming chapters have a good discussion of the various ontologies and knowledge base that we have.

During the ontology development process, we determined the scope of the ontology and the range of applications that could benefit by using the ontologies. This helped us in taking the modeling decisions and be focused towards the goals in mind. The ontology in hand should be able to model the environmental science domain, especially the hydrology field. It should be able to describe metadata about the domain datasets as well as be able to model the content of these datasets, thereby provide a conceptual schema to the domain data. Also, another goal was to address the various modeling

programs and tools which are extensively being used to perform data analysis. It should be easy to use not only for domain experts but also for novice and intermediate level users.

We decided on a set of questions that the ontologies should be able to answer. These questions helped us define and limit the scope of the ontology in hand. Some of the important questions are:

1. What is the exact geographic location of this environmental entity or environmental instrument?
2. Is rock a type of porous medium? Is Basalt a type of igneous rock?
3. What are the rainfall measurements for this Rain Gauge during the month of March 2005?
4. What are the possible attributes and the different types of Soil?
5. Can we perform soil modeling on the chemical species present in the groundwater in this well located in Baltimore, MD? If yes, how?
6. What is the temporal and spatial extent for this dataset?
7. Give me all the identification information for this dataset including name, organization, contact information and any detailed publications available.
8. How do I retrieve and use this dataset?
9. What type of information does this dataset contain?
10. What is the format of this dataset, is it formatted text or a relational database or excel worksheet or a digital map or a remote-sensing image?
11. Can we track the provenance for this dataset in order to determine the trust level?
12. What are the types of Computational Models available in order to perform analyses of the climate data to predict weather patterns?
13. What are the chemical species found inside this sample of water? Do these chemicals react to form a particular compound, if not what are the possible outcomes?

These questions were prepared in discussion with domain expert and provided good starting points for ontology development. In trying to answer these questions, we came up with specific domain concepts and relationships. We started categorizing them in hierarchy. We also started investigating into the different ontologies required.

In trying to answer Questions 1-4, we came up with **Environmental ontology**. This ontology provides specific domain knowledge by describing concepts like *Rock, Soil, Rainfall, Groundwater, Well, Lake, River,* etc. In order to describe the above specific concepts, we came up with generic concepts like *WaterMass* and *PorousMedium* from which the above specific concepts can be sub-classed. It also utilizes an already existing geographic ontology for any geographic related descriptions. This ontology also provides definitions of different environmental instruments like *Rain Gauge, Well,* etc. Now in order to record the measurements by these instruments we defined a generic concept of *Quantity* and other specialized quantities like *PhysicalQuantity, ChemicalQuantity, DimensionalQuantity, HydrologicQuantity* and *GeographicQuantity*. In this way, the ontology makes provisions to record measurements by different instruments and links them to the domain concepts. In answering questions such as 2 and 4, we referred to relevant glossaries/dictionaries mentioned before to check the different types of rocks and soils. However, traditionally the glossaries/dictionaries do not have terms in taxonomy. Hence, in order to categorize the different rock types and soil types we had several iterations of knowledge transfer with domain expert. The properties of the different rocks and soils were taken into consideration and the hierarchy was finalized. Section A of Chapter 4 gives a detailed description of this ontology.

We created **Molecule ontology** to provide a chemical knowledge base of different molecules for use by applications and other ontologies. It helps the above environmental ontology to answer questions such as 5 and 13. We linked the *WaterMass* concept within environmental ontology to the *Molecule* concept within molecule ontology through appropriate attributes in order for allowing water to store

chemical species details present in it. Section B of Chapter 4 provides a detailed description for this ontology.

In order to provide solutions to questions like 6-11 we developed **Metadata ontology.** The main goal here is to provide relevant meta-information for environmental datasets. It answers who, what, why, where, when and how of every facet of the dataset. It has a concept called *DataExtent* with subclasses *SpatialExtent* and *TemporalExtent* in order to answer question 6 above regarding the spatial and temporal extent of the dataset. Concepts like *DataIdentification* and *DataDistribution* provide answers to questions like 7 and 8 by giving complete description of identification and distribution details pertaining to individual datasets. Here we identified the various possible attributes that a dataset can have and organized them in appropriate classes. In order to answer question 9 and provide a complete semantic description of the information within the dataset, this ontology links to the concepts defined in the above environmental ontology. In this way, by selecting concepts within the environmental domain ontology that best describe the content within the dataset, a semantic understanding for the dataset is generated. To accommodate such selections, we designed attributes capable of holding identifiers of the classes and relationships within the domain ontology. Concepts like *DataContentType* and *DataPresentationForm* answers question 10. *DataContentType* has a hierarchy of classes categorized under *StructuredDataContent* and *UnstructuredDataContent*, whereas *DataPresentationForm* provides an enumeration of values to determine whether the dataset is online or in hard copy form. Section C of Chapter 4 gives a thorough description of the Metadata ontology.

Models ontology was created to capture the semantics of the different models (computational, chemical, physical, etc) used for analysis of environmental and related datasets. It directly answers questions such as 12 above since it provides description of different models. It combines with the above environmental and molecule ontologies in order to provide a solution to question 13. In this

case, it describes a geochemical model. Environmental ontology provides *Soil* and *GroundWater* knowledge and Molecule ontology provides *Molecule* knowledge. This knowledge combined with knowledge of geochemical model provides a complete answer to question 13. In Section D of Chapter 4 we describe this ontology in more detail.

The overall goal of the system is to provide semantic interoperability among heterogeneous environmental data sources by using these semantically rich ontologies, thereby facilitating and assisting in the detection, evaluation, and effort coordination that may lead to effective decision-making and resolution of imminent environmental problems.

Chapter IV

ONTOLOGIES

A. Environmental Ontology

This ontology defines the core concepts and relationships in the environmental science domain. The following figures 2 and 3 illustrate the different classes defined and show their hierarchy. As the scope of this ontology is huge, its development will continue with more inputs from the community. We do not guarantee this ontology to be complete. At present more importance is given to hydrologic sub-domain of environmental science. Few other concepts are defined as placeholders for future development.

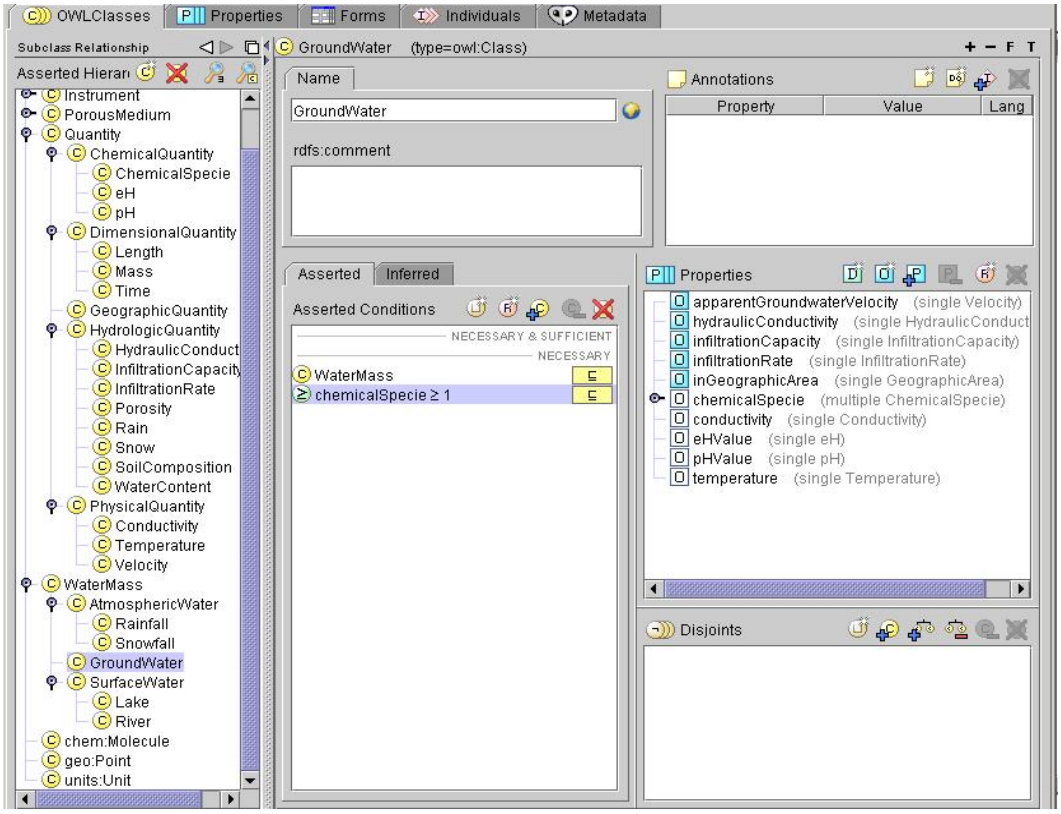

Figure 2. Environmental Ontology loaded in Protégé

1. Important Concepts

Instrument: This class can represent any kind of instrument.

EnvironmentalInstrument: This class represents an instrument which can measure environmental related quantities like rainfall, snowfall, air, gas, temperature, water quality, humidity, etc.

RainGauge: An instrument which measures rainfall.

Well: This class represents a well not just as a boring in the earth through which water can be obtained but also as an instrument that can be used to measure groundwater specifications.

WaterMass: This generic class serves as a super class for all types of water masses.

AtmosphericWater: This class provides a container for atmospheric water types like rainfall, snowfall, etc.

Rainfall: This class represents rainfall water by storing the physical as well as chemical properties of water along with the rainfall value.

Snowfall: This class represents snowfall by storing the physical as well as chemical properties of snow along with the snowfall value.

GroundWater: This class holds several chemical and physical attribute values related to ground water.

SurfaceWater: This class provides a container for surface water types like lake, river, sea, etc.

Lake: This class represents a lake.

River: This class represents a river.

Quantity: This represents a generic quantity class with simple attributes such as *hasUnit* and *hasValue*.

ChemicalQuantity: It serves as a container class for specific chemical quantities.

ChemicalSpecie: This class is used to store a chemical molecule and its concentration. It links to the *Molecule* ontology described in Section B of this chapter.

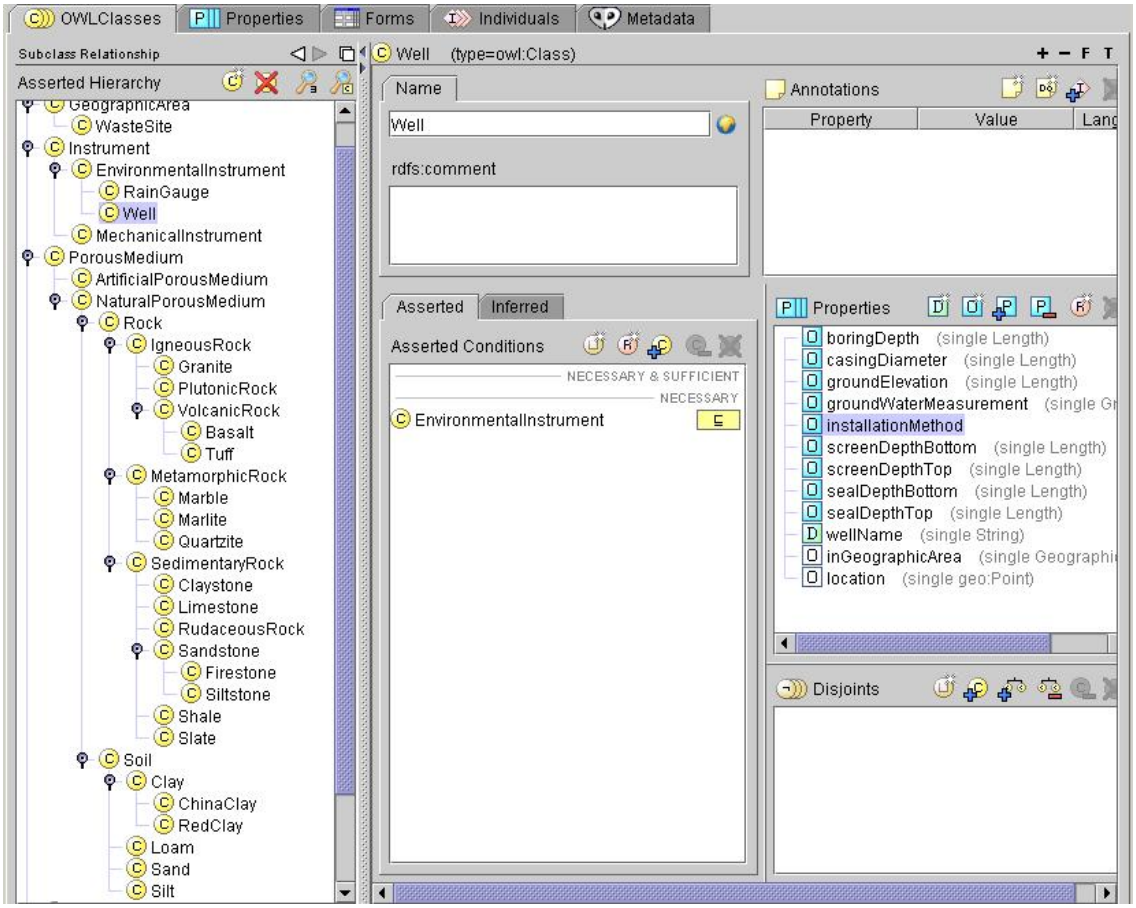

Figure 3. Environmental Ontology loaded in Protégé

eH: This class represents the eH measure which is the potential created by oxidation-reduction reactions.

pH: This class represents pH which is a relative measure of the acidity or alkanity of water.

DimensionalQuantity: This class provides a container for dimensional quantities like length, mass, time and others which fall in this category.

Length: This class represents a length quantity.

Mass: This class represents a mass quantity.

Time: This class represents a time quantity.

GeographicQuantity: This class provides a placeholder for geographic related quantities.

HydrologicQuantity: This class provides a placeholder for hydrology related quantities.

HydraulicConductivity: This class represents a measure for hydraulic conductivity which is a measure of the rate at which water can move through a permeable medium.

InfiltrationCapacity: This class represents a measure of infiltration capacity which is the maximum rate at which water can enter soil.

InfiltrationRate: This class represents a measure of infiltration rate which is the amount of water that enters the soil surface in a specified time interval.

Porosity: This class represents porosity measure which is normally a percentage of void space in a volume of substance.

Rain: This class represents Rain quantity.

Snow: This class represents Snow quantity.

SoilComposition: This class represents a measure indicating the composition of soil.

WaterContent: This class provides a measure of water content in the specified medium.

PhysicalQuantity: This class provides a placeholder for physical quantities.

Conductivity: It represents a measure of the ability to conduct electric current.

Temperature: This class represents a measure for temperature.

Velocity: This class represents a measure for velocity.

GeographicArea: This class is defined as set of boundary points which are identified by their latitude and longitude.

WasteSite: This class is defined as a specific subclass of *GeographicArea* and is used to represent a waste site.

PorousMedium: This class is used to represent all possible porous mediums.

NaturalPorousMedium: This class represents porous mediums formed by a natural process.

ArtificialPorousMedium: This class represents porous mediums formed by an artificial process.

Rock: Rock can be defined as a mass of mineral matter. This class represents a generic Rock type. We have defined several specific Rock types such as *MetamorphicRock, Quartzite, Marlite, Marble, SedimentaryRock, Limestone, Sandstone, Firestone, Siltstone, RudaceousRock, Shale, Slate,*

IgneousRock, Granite, VolcanicRock, Basalt, Tuff, PlutonicRock and *Claystone*. Many more may be added as needed by domain experts.

Soil: This class represents generic Soil class. We have defined several specific Soil types such as *Clay, ChinaClay, RedClay, Sand, Silt* and *Loam*. As earlier, more Soil types can be defined by domain experts.

2. Important Properties

areaName: This property provides the name of the *GeographicArea* site.

boundaryPoint: This property identifies a boundary point for instances of *GeographicArea* class. It takes values of type *Point* which in turn has attributes for storing latitude, longitude and altitude. Multiple *boudaryPoint* values can together give an exact description of a geographic area.

boringDepth: This property indicates a measure of depth of the *Well*.

casingDiameter: This property indicates a measure of diameter of the *Well*.

chemicalSpecie: This property has domain as *WaterMass* and range as *ChemicalSpecie* quantity, and hence it provides a way of specifying chemical molecules present in water.

conductivity: This property has domain as *WaterMass* and range as *Conductivity* quantity, and hence it provides a way of specifying conductivity for water.

eHValue: This property has domain as *WaterMass* and range as *eH* quantity, and hence it provides a way of specifying eH value for water.

pHValue: This property has domain as *WaterMass* and range as *pH* quantity, and hence it provides a way of specifying pH value for water.

porosity: This property has domain as *PorousMedium* and range as *Porosity* quantity, and hence it represents the porosity measure for porous mediums.

groundElevation: This property has domain as *Well* and range as *Length* quantity, and it provides a way for specifying the ground elevation for the well.

groundWaterMeasurement: This class has domain as *Well* and range as *GroundWater*. It provides a way for specifying ground water measurements for the well.

hasUnit: This property has domain as *Quantity* and range as *Unit*. A brief description of the Unit ontology follows later in this section. The purpose of this property is to provide an accurate representation of the unit for the specified quantity. This can also help in case units need to be changed.

hasValue: This property has domain as *Quantity* and takes values of type *float*. The value combined with unit provides a correct measure description for the *Quantity* class.

hasMolecule: This property has domain *ChemicalSpecie* and range as *Molecule* class defined in the Molecule ontology which is described in Section B of this chapter. Its purpose is to provide a way for classes in this environmental ontology to utilize the Molecule ontology in order to provide correct descriptions of chemical species.

hydraulicConductivity: This property has domain as *GroundWater* and range as *HydraulicConductivity* quantity, and hence it provides a way of specifying hydraulic conductivity for ground water.

infiltrationCapacity: This property has domain as *GroundWater* and range as *InfiltrationCapacity* quantity, and hence it provides a way of specifying infiltration capacity for ground water.

infiltrationRate: This property has domain as *GroundWater* and range as *InfiltrationRate* quantity, and hence it provides a way of specifying infiltration rate for ground water.

inGeographicArea: This property is used by several classes such as *EnvironmentalInstrument*, *GroundWater* and *SurfaceWater*. It takes values of type *GeographicArea*. Its purpose is to provide a mechanism for these classes to describe their geographic location.

location: This property provides a way for specifying geographic locations for *EnvironmentalInstrument*.

measurementDate: This property provides a way for specifying the date of measurement.

rainfallMeasurement: This property has domain as *RainGauge* and range as *Rainfall*, and hence it provides a way for rain gauge to link to Rainfall class.

rainfallValue: This property has domain *Rainfall* and range as *Rain* quantity, and hence it provides a way for Rainfall class to record the rain measurement.

screenDepthTop, screenDepthBottom, sealDepthTop, sealDepthBottom: These are properties for the *Well* class and they take values of type *Length* quantity.

snowfallValue: This property has domain *Snowfall* and range as *Snow* quantity, and hence it provides a way for Snowfall class to record the snow measurement.

temperature: This property has domain *WaterMass* and range as *Temperature* quantity, and hence it provides a way for specifying the temperature for water.

This ontology makes use of several other ontologies. A brief description of the geographic and units ontology being used follows later in this section.

We used groundwater and well data from NABIR FRC website [31], Oak Ridge Labs in order to generate a knowledge base of wells with groundwater measurements. We chose a test site with several wells. The data was parsed by a Java program and mapped to corresponding properties and classes within the ontology. A knowledge base of more than 30 wells was created this way.

Here we give an example of a well with important attributes defined:

```
<!-- well description -->
<rdf:Description
rdf:about="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#TPB32">
   <env:wellName>TPB32</env:wellName>
    <env:boringDepth rdf:nodeID="A7311"/>
    <env:groundElevation rdf:nodeID="A7398"/>
    <env:groundWaterMeasurement rdf:nodeID="A3676"/>
    <env:location rdf:nodeID="A2561"/>
    <rdf:type
rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#Well"/>
</rdf:Description>
```

```
<!-- boring depth -->
<rdf:Description rdf:nodeID="A7311">
    <env:hasValue>5.1</env:hasValue>
    <env:hasUnit
rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#meter"/>
    <rdf:type
rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#Length"/>
</rdf:Description>

<!-- ground elevation -->
<rdf:Description rdf:nodeID="A7398">
    <env:hasValue>1007.29</env:hasValue>
    <env:hasUnit
rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#meter"/>
    <rdf:type
rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#Length"/>
</rdf:Description>

<!-- location -->
<rdf:Description rdf:nodeID="A2561">
    <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#Point"/>
    <geo:latitude>35.977493</geo:latitude>
    <geo:longitude>84.27267</geo:longitude>
</rdf:Description>

<!-- groundwater measurement -->
<rdf:Description rdf:nodeID="A3676">
    <env:chemicalSpecie rdf:nodeID="A3677"/>
    <env:chemicalSpecie rdf:nodeID="A3678"/>
    <env:eHValue rdf:nodeID="A3680"/>
    <env:temperature rdf:nodeID="A3682"/>
   <rdf:type
rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#GroundWater
"/>
</rdf:Description>

<rdf:Description rdf:nodeID="A3677">
    <env:hasMolecule
    rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/molecule.owl#Cl"/>
    <env:hasUnit
    rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#mg_per_liter"/>
    <env:hasValue>332.5</env:hasValue>
<rdf:type
rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#ChemicalSpe
cie"/>
</rdf:Description>

<rdf:Description rdf:nodeID="A3678">
    <env:hasMolecule
    rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/molecule.owl#NO3"/>
    <env:hasValue>865.0</env:hasValue>
    <env:hasUnit
rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#mg_per_liter"/>
    <rdf:type
rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#ChemicalSpe
cie"/>
</rdf:Description>

<rdf:Description rdf:nodeID="A3680">
    <env:hasValue>111.0</env:hasValue>
    <env:hasUnit
rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#milliVolt"/>
```

```
    <rdf:type
rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#eH"/>
</rdf:Description>

<rdf:Description rdf:nodeID="A3682">
    <env:hasValue>70.3</env:hasValue>
    <env:hasUnit
rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#fahrenheit"/>
    <rdf:type
rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/environment.owl#Temperature
"/>
</rdf:Description>
```

3. Geographic Ontology

We have used a very minimalistic RDF vocabulary which describes Points with latitude, longitude, and altitude properties from the WGS84 reference datum specification. This ontology and its detailed description is available at [45]. It describes a *Point* concept which has only one 'lat', only one 'long', and only one 'alt'.

An example is:

```
 <rdf:Description rdf:nodeID="A4257">
    <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#Point"/>
    <geo:latitude>35.977695</geo:latitude>
    <geo:longitude>84.27152</geo:longitude>
 </rdf:Description>
```

As seen at present we are using anonymous nodes to identify geographic locations. However, in the future a namespace should be defined which can identify each and every location.

4. Units Ontology

We have used the Units ontology which is developed as a part of SWEET (Semantic Web for Earth and Environment Terminology) ontologies [18, 47]. They define Units using Unidata's UDUnits [46]. The ontology includes conversion factors between various units. Prefixed units such as km are defined as a special case of m with appropriate conversion factor. Several characterizing classes are defined such as *Unit, BaseUnit, DerivedUnit, UnitDerivedByRaisingToPower, SimpleUnit,*

ComplexUnit, Prefix, BaseUnitOrUnitDerivedWithoutChangingOfDimension, UnitDerivedByScaling, PrefixOrUnit, UnitDerivedByShifting and *UnitDerivedWithoutChangingOfDimension*. Several units are defined as appropriate instances of these above classes. A few examples are minute, hour, meter, degree, Newton, kilogram_meterSquare_perSecondSquare, coulomb, volt, pascal_perSecond, etc.

A derived unit like perMeter is represented as follows:

```
<units:UnitDerivedByRaisingToPower rdf:ID="perMeter">
       <units:derivedFromUnit rdf:resource="#meter"/>
       <units:hasPower rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
       -1</units:hasPower>
       <units:hasSymbol rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
       1/m</units:hasSymbol>
</units:UnitDerivedByRaisingToPower>
```

And a Complex unit like is represented as follows:

```
<units:ComplexUnit rdf:ID="watt_perMeterSquare_perSteradian">
       <units:productOf
       rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#watt"/>
       <units:productOf
       rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#perMeterSquare"/>
       <units:productOf
       rdf:resource="http://sweet.jpl.nasa.gov/ontology/units.owl#perSteradian"/>
</units:ComplexUnit>
```

The *hasUnit* attribute of the *Quantity* class within the above environmental ontology takes values of type *Unit* which is the super-class of all the specific Unit classes.

B. Molecule Ontology

The aim of molecule ontology is to provide a knowledge base of all kinds chemical molecules and their properties. Figure 4 below shows a snapshot of Protégé with this ontology loaded.

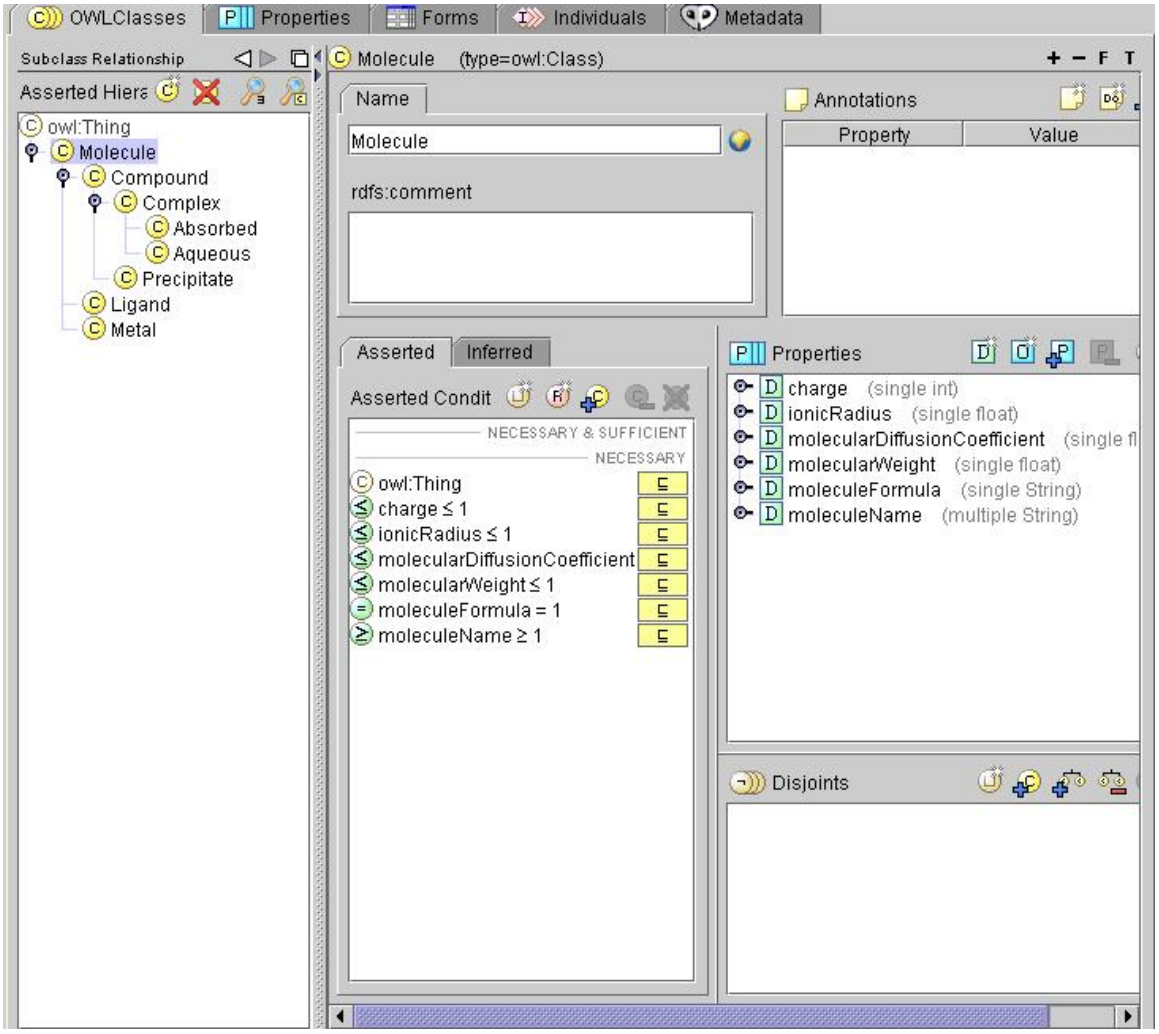

Figure 4. Molecule Ontology loaded in Protégé

1. Important Concepts

Molecule: This class represents a generic molecule.

Metal: A metal can be viewed as chemical elements with special characteristics such as conductivity, luster, ductility and opacity. This class is used to represent such chemical elements. Metals like Aluminum, Iron, and Manganese are good instances for this class.

Ligand: A ligand can be said as a molecule that binds to the surface of another molecule. This class represents such molecules. Examples are molecules like Chloride, Bromide, Ammonia, etc.

Compound: A compound can be referred as a molecule formed by chemical union of 2 or more elements. This ontology defines several specific types of compounds like **Complex, Precipitate, Aqueous** and **Absorbed**. Each of these classes has its own characteristics which make them distinct

2. Important Properties

ionicRadius: It is used to represent radius of an ion. It takes values of type float.

molecularWeight: It is used to represent molecular weight of the molecule. It also takes values of type float.

charge: It represents the charge of the molecule. It takes values of type integer.

molecularDiffusionCoefficient: It represents the molecular diffusion coefficient of the molecule. It takes values of type float.

moleculeName: It represents name of the molecule.

moleculeFormula: It provide the chemical formula used to uniquely identify the molecule.

We used the geochem program dataset [26] to generate the knowledge base of molecules. This dataset lists several different types of molecules with their properties. The process was automated using Java programs which would read in the dataset files and parse the text. The fields within the dataset were processed and mapped to the classes and properties defined in this ontology. In this way we generated a knowledge base of more than 100 molecule instances categorized to the appropriate subclass of Molecule. We have used the chemical formula of the molecule as a part of the URI to uniquely identify each molecule.

A metal like Aluminum has the following definition within the knowledge base:

```
<rdf:Description
rdf:about="http://www.cs.umbc.edu/~viral1/ontologies/molecule.owl#Al">
       <chem:charge>3</chem:charge>
       <chem:ionicRadius>0.51</chem:ionicRadius>
       <chem:molecularDiffusionCoefficient>3.46E-6
       </chem:molecularDiffusionCoefficient>
```

```
        <chem:molecularWeight>27.4</chem:molecularWeight>
        <chem:moleculeFormula>Al</chem:moleculeFormula>
        <chem:moleculeName>Aluminum</chem:moleculeName>
        <rdf:type
        rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/molecule.owl#Metal"/
        >
</rdf:Description>
```

C. Metadata Ontology

This ontology provides a standard vocabulary of terms useful for describing a domain dataset. The objective of this ontology is to provide metadata for the dataset as well as to provide a semantic understanding of the data content within the dataset. The ontology defines a set of elements which will be used for the purpose of documentation of the dataset. It answers who, what, why, where, when and how of every facet of the dataset. The ultimate goal is to provide a basis for an efficient mechanism of content based retrieval of datasets. Semantic understanding is achieved by mapping the dataset to concepts defined in the environmental science domain ontology. This mapping provides ontology based conceptual schema for the dataset. Figure 5 shows the Role of Metadata ontology.

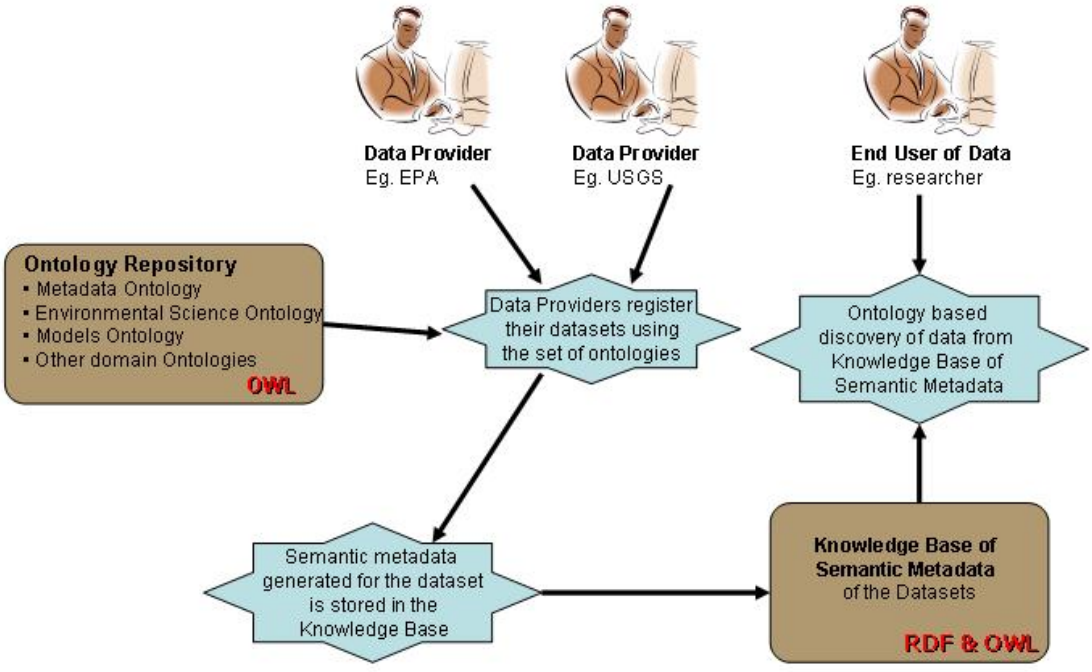

Figure 5. Role of Metadata ontology

Various data providers register their datasets using the Metadata ontology and also select domain concepts which best describe the content within the dataset. The knowledge base repository stores the RDF semantic metadata for these registered datasets. In this way, the Metadata ontology will provide a common model for search across various data sources.

Certain concepts in this ontology are based on FGDC (Federal Geographic Data Committee) standard [15]. FGDC standard is very complex and it is text based. We claim our representation to be much simpler yet resourceful, semantically rich and machine understandable as it is based on domain rich ontologies which are encoded in OWL.

1. Important Concepts and Properties

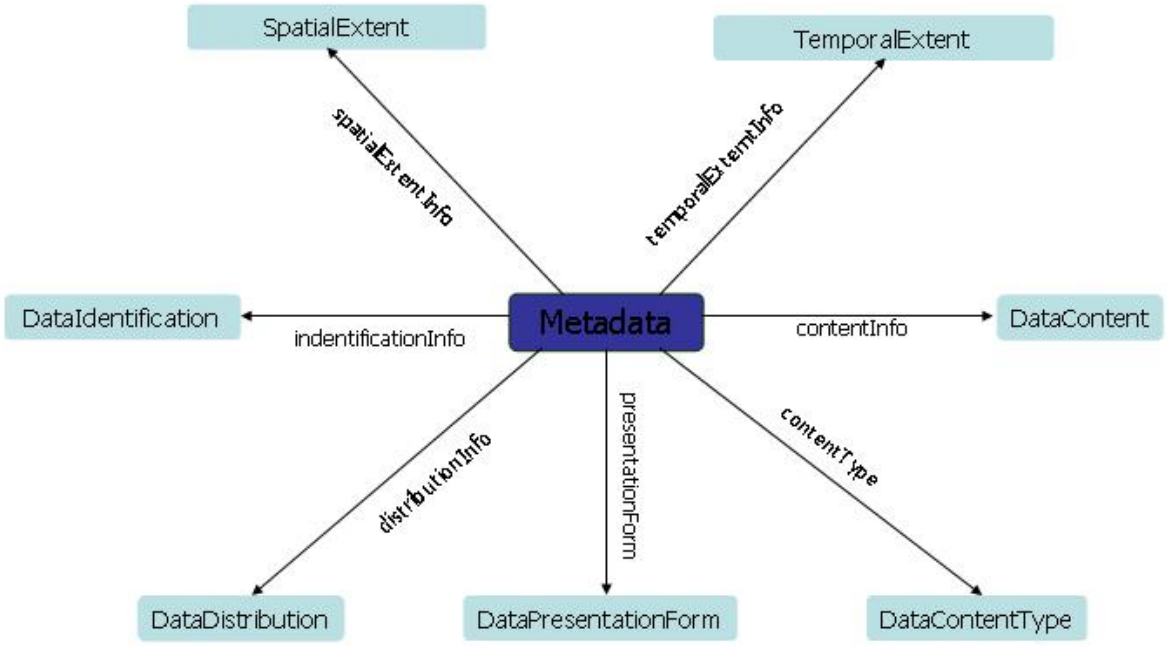

Figure 6. Metadata ontology

Figure 6 above gives an overview of the metadata ontology and how different classes are related to each other. As can be seen, several classes constitute the metadata ontology. A brief description of the different classes involved in this ontology follows:

Metadata: Metadata class is the principal component of the ontology and it links to other classes in the ontology through its attributes as can be seen in the figure 6 above. For each data set that is registered with our system, a corresponding instance of metadata class is created and stored in the knowledge base.

DataIdentification: This class allows the provider to specify basic identification information about the dataset. The important attributes of this class are:

- *title, description, publication, note*
- *creator, participant, pointOfContact*
- *creationDate, lastModificationDate*
- *status, maintenanceFrequency*
- *isPartOf, isDerivedFrom*

Several attributes of this class are sub-properties of Dublin Core [7] metadata element set which provides a standard for information resource description. Certain attributes such as *status* and *maintenanceFrequency* have an enumeration of allowable values. Properties such as *isPartOf* and *isDerivedFrom* are used to track the provenance of the dataset. They take values as other Metadata instances and are defined as transitive properties. In this way, we can track the lineage of the dataset which can be very useful. Provenance can help build trust in the data and allow reuse of a larger number of datasets. Attributes such as *creator, participant* and *publication* use **Person** and **Publication** ontologies developed by the ebiquity research group [52] of UMBC. The Person class has attributes such as *firstName, middleName, lastName, street, city, state, country, postalCode,*

company, phoneNumber, emailAddress, fax, etc. The Publication class has attributes such as *title, publishedOn, description, keyword, version, author, firstAuthor, softCopyURI, softCopyFormat, softCopySize, abstract, edition, chapter, series, pages, volume, number, note, address, organization, journal, bookTitle, institution, publisher, etc.*

SpatialExtent: This class gives information about the geographic area covered by the dataset. It permits the data provider to specify the bounding coordinates of coverage of the dataset in terms of latitude and longitude values in the order western-most, eastern-most, northern-most, and southern-most. The important attributes are *eastBoundLongitude, northBoundLatitude, southBoundLatitude* and *westBoundLongitude.*

TemporalExtent: This class provides a means for describing the temporal information corresponding to the dataset. It is possible to specify a single date/time or a specific time period. The important properties are *beginDate, endDate* and just *date* in case it is a single date.

DataContent: This is a pivotal class in this ontology and is responsible for mapping the dataset to the domain concepts defined in the Environmental Science and other domain ontologies. This linkage generates a semantic conceptual schema for the dataset. The data provider selects the concepts from the domain ontology that best describes the dataset. This selection is stored in the DataContent class allowing the metadata ontology to provide not only metadata about the dataset but also semantic description of the data content within the dataset. This linkage is stored in terms of URIs of the concepts and relationships from the domain ontology. The 2 important properties of this class are:

- *hasConcept* – stores the URI of the domain concept which describes content within this dataset
- *hasRelation* – stores the URI of the domain relation which describes the relations within this dataset

Multiple values of the above 2 properties will best describe the data content.

DataContentType: This class provides information regarding the type of dataset. It indicates whether the dataset has structured content in the form of relational database, excel files, markup data, etc or unstructured content like text files, images, maps, etc.

DataPresentationForm: Information about form of dataset, i.e. whether it is digital or exists in hardcopy is provided using this class in the ontology. This class provides an enumeration of values to use.

DataDistribution: Information about the distributor of the dataset and the digital transfer options for obtaining this dataset from the concerned organization can be provided using this class. It also has provisions to specify any legal disclaimer and any use or access constraints associated with the dataset. The important properties are *accessConstraints, distributionFormat, distributor, legalDisclaimer, transferOptions* and *useConstraints.*

Thus, each dataset that is registered with our Metadata OWL ontology has content based semantic description associated with it apart from the metadata information about identification, spatial, extent, distribution and presentation form. This semantic description is independent of the dataset format and is generated using the environmental science domain specific ontologies. This approach allows the end users of data to search for relevant datasets based on their semantic content and metadata rather than just simple keywords.

Figure 7 below gives a snapshot of this ontology loaded in Protégé.

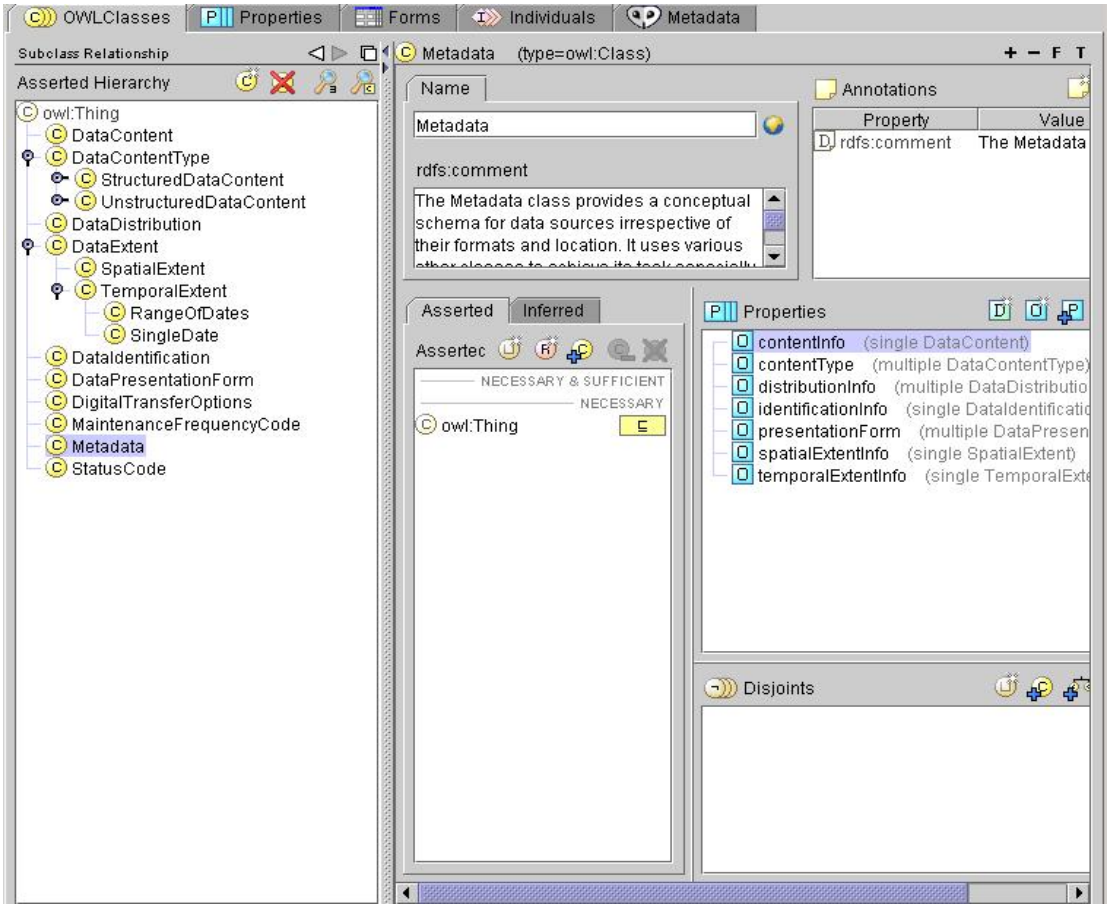

Figure 7. Metadata Ontology loaded in Protégé

D. Models Ontology

This ontology serves the purpose of providing a representation of various models being used. A model can viewed as a simplified description of a complex entity or process. This ontology should appeal to modelers and model users, as well as anyone engaged in practice of hydrology, civil engineering, environmental science, agricultural engineering, climatology and related domains. It should provide a comprehensive account of the various domain models and tools currently being used. We have taken a generic approach and defined several different domain models.

Figure 8 below depicts the variety of model classes that we envision. New Model classes can be easily defined.

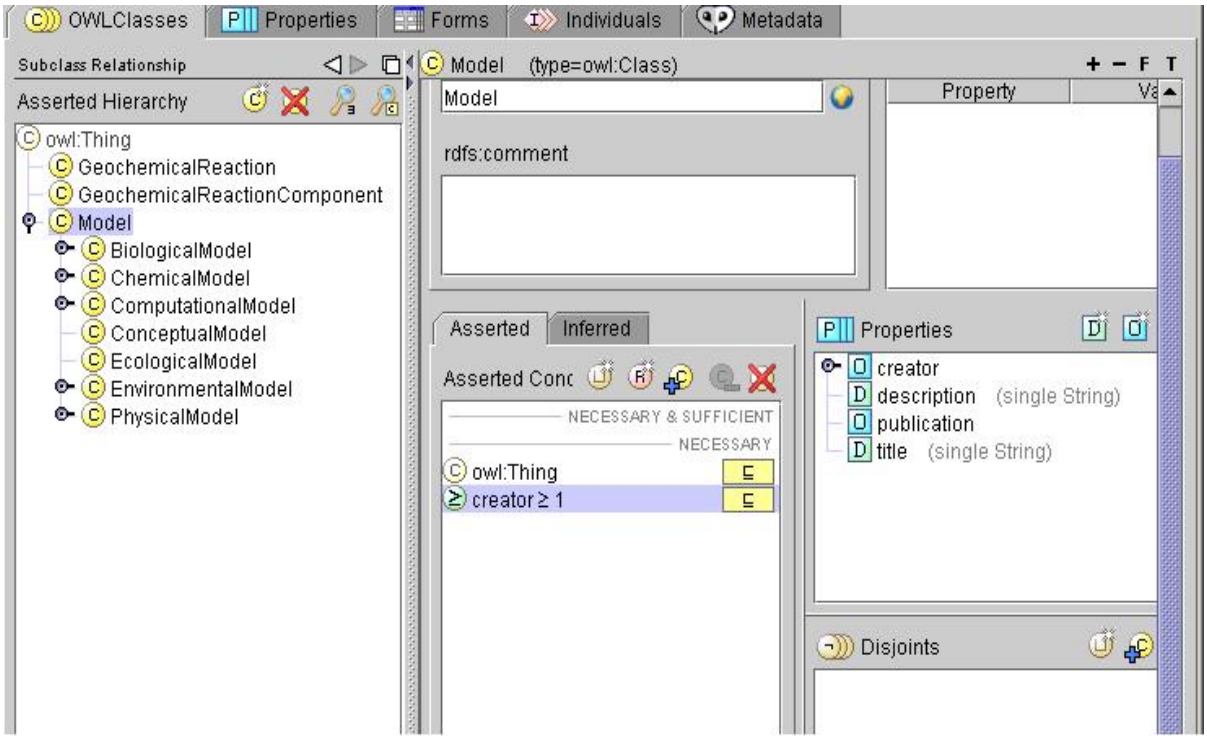

Figure 8. Models Ontology loaded in Protégé

This ontology is not fully developed. We have developed various place holder classes so that domain expert modelers can extend it and define new models as well as update existing models within this ontology. The goal is that this ontology should be able to provide model run descriptions, input scenarios, identification of input datasets, model run configuration, model documentation and a description of the tools/software required to run the model. The aim is to automate the entire process of choosing a model, searching for datasets and running the model. These semantic model descriptions would also help to link the various models along their data requirements in the sense that the output of say models A and B can be used as input for model C. This would provide much relief to researchers as currently this process is the most time consuming process for any modeling task.

1. Important Concepts and Properties

The *Model* class represents a generic model. This class will serve as a super class for specific model types. As shown several different domain models are defined such as *PhysicalModel, AnalogModel, ChemicalModel, GeochemicalModel, ComputationalModel, MathematicalModel, BiologicalModel, StatisticalModel, GeneticModel, EnvironmentalModel, HydrologicalModel, ErosionModel, GeographicalModel, GISModel., ConceptualModel* and *EcologicalModel*. They are appropriately categorized according to their characteristics.

A list of few Dublin Core style properties that are used by these Model classes are *description, title, creator and publication.* Again Person and Publication ontologies described in Section A are used for applicable properties.

Domain expert ontology developers can further extend these existing classes and properties to serve their own purpose. As discussed before, we would like to see this ontology providing not just metadata about the model, but giving a semantic overview of the entire process flow for the model. This will enable users to understand the model better and also assist them to execute the model.

To illustrate this ontology, the geochem [26] program model can be described as follows:

```
<models:GeochemicalModel rdf:ID="geochem">
       <models:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
       >Geochem</models:title>
       <models:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
       >A computer program for the calculation of chemical equilibria in soild
       solutions and other natural water systems</models:description>
       <models:creator
       rdf:resource="http://ebiquity.umbc.edu/v2.1/ontology/person.owl#GarrisSposit
       o" />
       <models:publication
       rdf:resource="http://ebiquity.umbc.edu/v2.1/ontology/publication.owl#id547"/
       >
</models:GeochemicalModel>
```

There would be several other supporting classes which will store the knowledge provided by these models. We have demonstrated this by creating classes and properties for the geochem program model as follows: *GeochemicalReaction, GeochemicalReactionComponent, hasReactant, hasProduct, reactionConstant, backwardRateCoefficient, forwardRateCoefficient, stoichiometricCoefficient, numberReactants, numberProducts*. These ontology elements serve the purpose of storing chemical reactions.

The *GeochemicalReaction* ontology class and the other supporting ontology elements would in fact use instances of the *Molecule* class defined in Section B of this chapter. We again used the geochem program dataset [26] to generate the instances of GeochemicalReaction. This dataset provides description of various chemical reactions. The process was automated using Java programs which would read in the dataset files and parse the text. The required fields within the dataset were processed and mapped to the classes and properties defined in this ontology. In this way we generated a knowledge base of geochemical reactions.

A geochemical reaction involving one Calcium and two Chlorine molecules combining to form aqueous CalciumChloride can be described as follows:

```
<rdf:Description
rdf:about="http://www.cs.umbc.edu/~viral1/ontologies/models.owl#RCa2Cl_a">
       <models:backwardRateCoefficient>100</models:backwardRateCoefficient>
       <models:forwardRateCoefficient>10</models:forwardRateCoefficient>
       <models:hasProduct rdf:nodeID="A3343"/>
       <models:hasReactant rdf:nodeID="A7898"/>
       <models:hasReactant rdf:nodeID="A8151"/>
       <models:numberProducts>1</models:numberProducts>
       <models:numberReactants>2</models:numberReactants>
       <models:reactionConstant>0.01</models:reactionConstant>
       <rdf:type
       rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/models.owl#Geochemic
       alReaction"/>
</rdf:Description>

<rdf:Description rdf:nodeID="A3343">
```

```
        <models:molecule
        rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/molecule.owl#Ca2Cl_a
        "/>
        <models:stoichiometricCoefficient>1</models:stoichiometricCoefficient>
        <rdf:type
        rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/models.owl#Geochemic
        alReactionComponent"/>
</rdf:Description>

<rdf:Description rdf:nodeID="A7898">
        <models:molecule
        rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/molecule.owl#Ca"/>
        <models:stoichiometricCoefficient>1</models:stoichiometricCoefficient>
        <rdf:type
        rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/models.owl#Geochemic
        alReactionComponent"/>
</rdf:Description>

<rdf:Description rdf:nodeID="A8151">
        <models:molecule
        rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/molecule.owl#Cl"/>
        <models:stoichiometricCoefficient>2</models:stoichiometricCoefficient>
        <rdf:type
        rdf:resource="http://www.cs.umbc.edu/~viral1/ontologies/models.owl#Geochemic
        alReactionComponent"/>
</rdf:Description>
```

As seen, the geochemical reaction references molecules present in the Molecule knowledge base. This demonstrates how different ontologies can interact and integrate to provide a coherent view of the knowledge.

Chapter V

APPLICATIONS

To demonstrate the applications of the ontologies developed in this research, we selected two typical applications in the geochemical community and the groundwater hydrology communities. Geochemists frequently use geochemistry models and existing thermodynamic databases to calculate the speciation of mixture of chemical species. The purpose ranges from experimental design to determining the level of toxicity of groundwater pollution. The geochemistry application, put in the settings of contaminated waste site (in this case, a waste site of the Manhattan Project legacy of the US Department of Energy), allows the ontologies and knowledge base to be used for the groundwater hydrology community.

Consider a geochemist wanting to do modeling of chemical species for soil samples. We have developed a web based chemical modeling environment which uses the **Molecule** and **Models** ontologies described in previous chapter. It allows the user to search and retrieve a set of molecules (metals, ligands, compounds, complexes, etc) from the molecule knowledge base through a simple search interface as shown in Figure 8. The geochemist determines the chemical molecules present in the soil sample and makes his selection from the knowledge base accordingly. He can either search the molecules using their molecular formula or the molecule name. Once the selection of molecules is done, the geochemist can view all the chemical reactions (Figure 9) which have a subset of his molecule selection set as their reactants. These reactions are retrieved from the knowledge base of Models ontology. Now the geochemist has the option to stop here. He/She may use the information obtained from the query to conduct a chemical speciation calculation offline. Better yet, she or he may continue with chemical modeling using a geochemistry program available at the same web site such as GEOCHEM [26] on his or selection of chemical molecules and reactions.

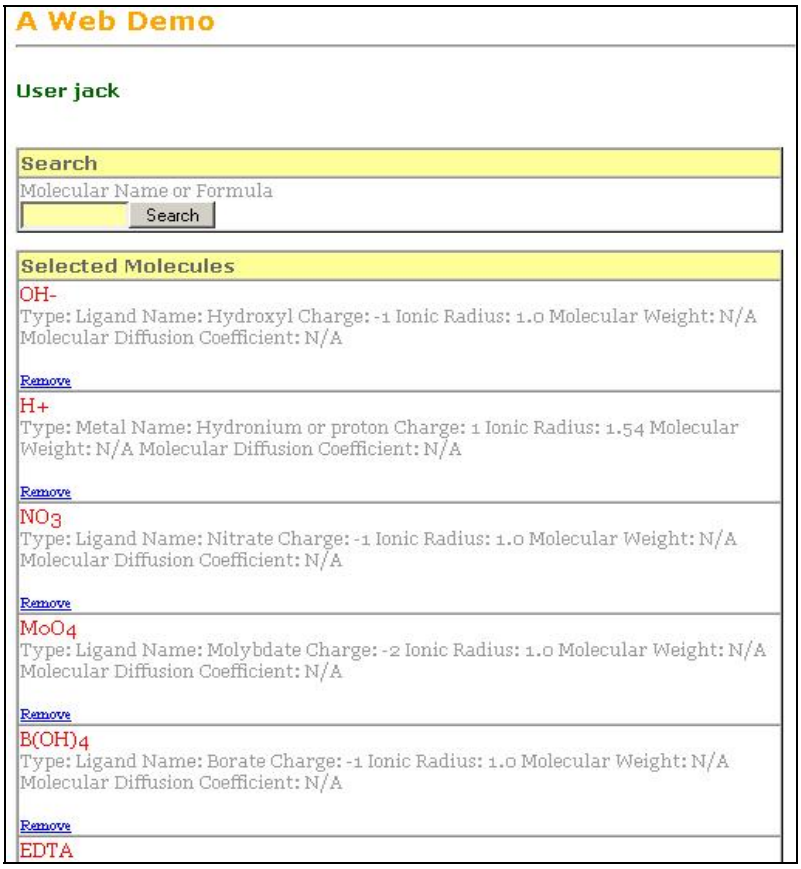

Figure 8. Snapshot of Web Application showing Molecule search interface

GEOCHEM is a computer program for predicting the chemical speciation in soil systems. The equilibria that can be calculated by this program are complexation, oxidation-reduction, precipitation, cation exchange, and metal ion adsorption. The input file for the program is generated by the application based on the geochemist's selection. The geochem modeling program now executes on the web server and the results are presented to the geochemist. The geochemist can now download the output file for future reference. This application shows how a simple ontology like Molecule ontology and the molecule instances can solve a geochemist's modeling requirements and help him perform the modeling task. We also developed a web based interface for experts to enrich the knowledge base by creating new instances of the Molecule concept in Molecule ontology. It is also possible to create new chemical reactions using the available molecule knowledge base.

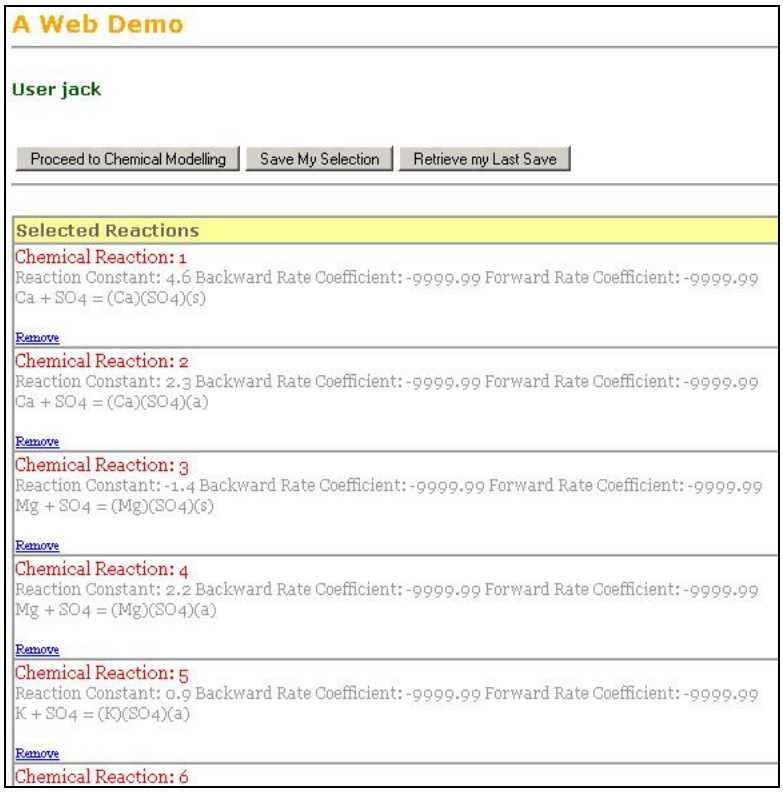

Figure 9. Snapshot of Web Application showing chemical reactions for selected molecules

Following the speciation calculations above, the geochemist may proceed or another environmental scientist may use his or her results to study spatial distributions of pollutants in the wells of a waste site. This next application also uses the **Molecule** ontology and GEOCHEM with the assistance of the **Environmental** ontology and a knowledge base of wells derived from [31]. The idea is to study spatial distributions of pollutants and their reaction products at the US DOE NABIR Field Research Center [30]. The web application presents a schematic of the field site and allows a scientist to view and select any well as shown in Figure 10. The Environment ontology describes the Well concept and the related properties necessary for this application. On selection of a well, semantic metadata of that well is retrieved from the knowledge base. The scientist can view the well specifications details, groundwater physical properties and groundwater chemical properties for the well as shown in Figure 11. The groundwater chemical properties displays the chemical species present in the well groundwater. The scientist may like to perform geochemical modeling on these chemical species and

study their behavior. Such studies are of great interest as a large amount of the radioactive wastes may be absorbed on the solid surface of the subsurface rocks. The measurements as stored in the knowledge base represent only the amount of these wastes in the groundwater. The amount of the pollutants in the solid phase has a much greater impact to the environment in the long term attenuations of the pollutants as well as the restoration of the waste sites. In this respect the scientist can retrieve the molecule specifications for the chemical species by querying the Molecule knowledge base. Now similar to the previous application, the scientist can query for chemical reactions for these molecules and/or perform GEOCHEM modeling on the molecule selection. Here we demonstrate how a user started with a semantic view of wells and its specifications and used knowledge from multiple ontologies and knowledge bases to integrate data and perform geochemical modeling task.

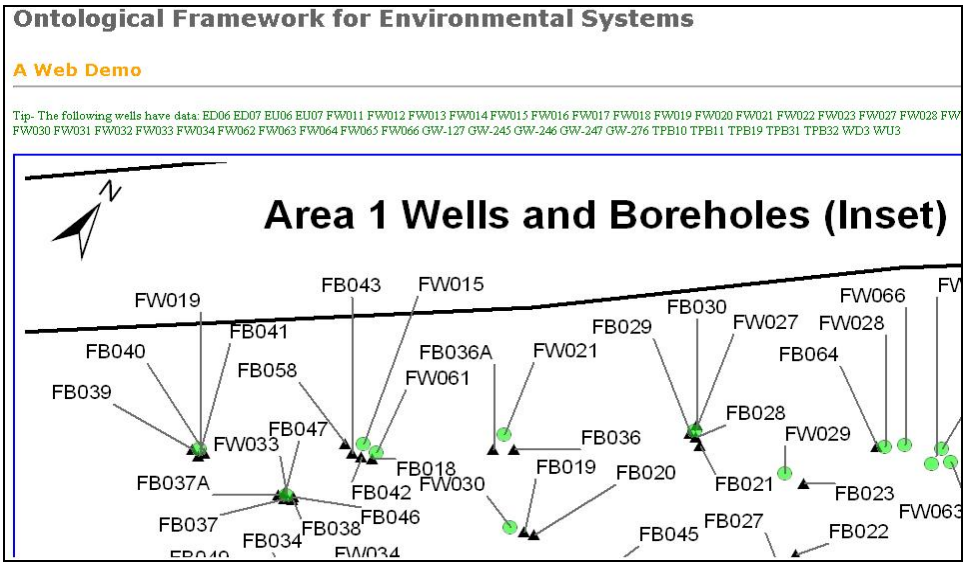

Figure 10. Snapshot of Web Application showing Wells in field site

Figure 11. Snapshot of Web Application showing Well data retrieved from knowledge base

These applications demonstrate how simple ontologies like Molecule ontology and Environmental Ontology can provide semantic description for data like molecules and wells in this case. This machine understandable description was used by GEOCHEM model to execute and perform relatively complex tasks that would usually take a researcher twice to three times of efforts to accomplish.

Chapter VI

DISCUSSION

Ontologies provide an abstract conceptualization of information to be represented and machine-interpretable definitions of basic concepts in specific domain and the relations among them. As these ontologies evolve and become a part of global environmental information system, several complex applications can be facilitated which have always been a far fetched dream. Not only will ontologies provide semantic interoperability among heterogeneous data, but it will also act as a guiding tool for many useful applications. Discovery of datasets based on their content and metadata will be possible as our Metadata ontology provides a complete semantic description of the datasets. Moreover, it provides a common conceptual model for datasets across multiple data sources. Also using the Models ontology, searching and choosing the right computational model along with locating the relevant datasets which form the input for the chosen model can be facilitated. With more intelligent reasoning, a composition of a sequence of models can be determined in order to perform a task. For example, an environmental scientist wants to perform hydrological modeling of a waste site and he/she selects an appropriate model that fit the requirements. The system may then suggest the user the required data and the available datasets useful for this task. In this case, it can also advice the user to use a GIS model in order to prepare the required geographic dataset and use a geochemical model to prepare the chemical composition dataset required for the original hydrological model to execute. All this reasoning is possible due to the available semantic description of the model. The chain can continue and the system can now help the user to determine the input requirements for the GIS and the geochemical models. In this way, a complete description of the different models to run and the different datasets to use is now available to the scientist. At present stage, this complete task can take days to months. This ontology based system will provide a big improvement and relief for the users.

Although the Semantic Web and ontologies are growing with time, more realistic applications are needed for its adoption by a broader community. W3C is promoting Semantic Web [6] and publishing new standards for the same reason. Numerous research efforts are currently underway at universities and corporate research organizations such as HP and IBM. Semantic Web is a vision through which information will have machine understandable semantics that will support automation, integration and reuse across applications. The success of Semantic Web will require new automated procedures to generate the RDF metadata for information. Moreover, for environmental science community this would be a bigger requirement. This is because traditionally environmental scientists are not information technologists and we cannot expect them to understand or write RDF/OWL files. We demonstrated a few applications of our ontologies. However, more complex and realistic scenarios need to be tackled for earlier adoption of these emerging technologies by the environmental science and other scientific communities.

We have provided a starting point for intelligent environmental information systems by developing these ontologies and showing their power and usefulness. However, more efforts will need to be taken in the direction of standardizing these ontologies. The standard government bodies like EPA, USGS, NASA, etc need to come together and develop the common standards for such ontologies. This will help spread the word on use of these knowledge structures. These organizations hold the power of promoting these standards on users and publishers of the datasets. Hence proactive actions and research needs to be undertaken by these organizations in order to understand the usefulness and applications of ontology based systems.

In the Semantic Web world, each resource is identified by its URI. It is through URIs that different pieces of information can be linked together. However, unlike the actual Web, RDF URIs can refer to any identifiable thing, including things that may not be directly retrievable on the Web. More precisely, a URI reference (URI + fragment identifier at the end) identifies a RDF resource. Like

http://www.cs.umbc.edu/~viral1/ontologies/molecule.owl#Fe+2 identifies Iron (cuprous). RDF Primer [36] provides more detailed information about URIs and its usefulness to Semantic Web. As the ontologies evolve and with more community inputs, URI references for the concepts and relationships defined in our developed ontologies will get standardized.

As seen in Section C of Chapter 4, for the Metadata ontology to provide semantic description of the dataset content, it links to the domain ontologies by providing the URIs of the concepts within these ontologies. We could have used classes as property values here, however then the ontologies would have been OWL Full. This would have added to the complexity level of the ontologies. There is current research [37] under W3C to search for alternate ways of representing classes as property values. This would help in describing data more easily.

The current Semantic Web standards have little or no provisions to incorporate uncertainty information. This situation is frequently encountered in our domain like when scientists or engineers record measurements of data or when a user inputs data to a modeling program and he is unsure about the exact values. This is requirement for other scientific communities and a solution to this problem would be advantageous.

We believe fast and efficient strategies for ontology development are much needed today. Massive effort is needed from the domain expert in order to construct ontologies manually, especially in case where the application domain is large such as ours. There is a need for semi-automatic approach in ontology building which will help the domain expert in constructing extensive domain ontologies efficiently. There is ongoing research in this field. [33] aims to convert dictionary to a graph structure where each node is a headword from the dictionary and arcs between nodes represent the use of other headwords for the definition of one particular node i.e. headword. Their approach uses algebraic extraction techniques to output a set of related terms. In [31], the authors mine the WWW and enrich

the ontology based on the comparison between statistical information of word usage in the corpus and structure of the ontology itself. In [32] the authors use hierarchical conceptual clustering, dictionary parsing, and association rule mining in order to infer relationships among concepts. With more research in this direction, ready to use ontology learning tools will be available in future. These tools will be able to use the latest techniques such as text mining to discover new concepts and infer relationships between these concepts using glossaries/dictionaries and domain texts. This will greatly help in the advancements of the Semantic Web.

Chapter VII

CONCLUSION

The objective of this research is to design and deliver information infrastructures that enable the deployment of efficient data sharing and integration mechanisms for environmental sciences and engineering. The solution we proposed is using ontologies and Semantic Web technologies like RDF and OWL. In its conventional form an ontology accounts for the representation of shared concepts and relationships in a domain facilitating communication among people and applications systems. The vision of these emerging semantic standards is an infrastructure where machines can understand and reason about data. This will lead to efficient and automated mechanisms for information sharing and integration.

Finding the right data for our chosen domain is often difficult for humans, and impossible for intelligent computational agents. The Semantic Web supports a future of intelligent agents sharing knowledge and thereby supporting efficient, fast and reliable decision making. Environmental decision-making problems are solved typically by combining multiple, heterogeneous datasets. This requires search, retrieval and integration of different types of information which may be using different schemas and formats. The ability to locate, obtain and use the datasets as easily and seamlessly is crucial for research scientists, which remains challenging today. The time required for this process must be shortened and the process needs to be automated. In this research, we described and demonstrated how ontologies and semantic web technologies will largely empower the required discovery and semantic interoperability among the heterogeneous data sources.

We developed a collection of ontologies for the environmental sciences and engineering disciplines. This research also developed a model that facilitates standard semantic description of domain

datasets. The complexity and diversity of the domain knowledge and terminology is captured and represented through ontologies. Through various applications we demonstrated how this ontological approach benefits environmental sciences and engineering and provides building blocks for solutions to several problems being faced by the domain users due to lack of semantics. As Semantic Web grows and new ontologies arise, data sharing across multiple domains will become a reality. Machine processable semantic metadata described using RDF and OWL will be widespread and more robust and powerful tools will be available to process, visualize, navigate and reason this metadata. As ontologies support reusability, this will allow same ontologies to be used for multiple applications.

In conclusion, with the infrastructure of domain ontologies and semantic knowledge an intelligent environmental information system will evolve. This system will have the potential of data integration, ease of data discovery, facilitate searching and planning for inputs to computational models and guidance in their execution, support semantic analysis of data, etc. This will lead to effective decision-making and resolution of imminent environmental problems.

REFERENCES

[1] T. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. *In International Journal of Human-Computer studies, 1995.*

[2] M. Ushold and M. Gruninger. Ontologies: Principles, methods and applications. *In The Knowledge Engineering Review, 1996.*

[3] T. Berners-Lee, J. Hendler and O. Lassila. The Semantic Web. *In Scientific American, May 2001*.

[4] A. Rousseau et al. Information technologies in a wider perspective: integrating management functions across the urban–rural interface. *Environmental Modelling & Software; Apr2005, Vol. 20 Issue 4, p443, 13p.*

[5] L. Ceccaroni and U. Cortés. OntoWEDSS: augmenting environmental decision-support systems with ontologies. *Environmental Modelling & Software; Sep2004, Vol. 19 Issue 9, p785, 13p.*

[6] Semantic Web Best Practices Working and Deployment Working Group.
Url: http://www.w3.org/2001/sw/BestPractices/.

[7] M. Purvis et al. A multi-agent system for the integration of distributed environmental information. *Environmental Modelling & Software; Jul2003, Vol. 18 Issue 6, p565, 8p.*

[8] D. Beckett, E. Miller and D. Brickley (2002). Expressing Simple Dublin Core in RDF/XML.
Url: http://dublincore.org/documents/dcmes-xml/.

[9] FOAF – Friend of a Friend. Url: http://www.foaf-project.org/.

[10] K. Hiramatsu and F. Reitsma (2004). GeoReferencing the Semantic Web: ontology based markup of geographically referenced information. *Joint EuroSDR/EuroGeographics workshop on Ontologies and Schema Translation Services, Paris, France, 15th - 16th April 2004.*

[11] GeoSemantic Web. Url: http://www.mindswap.org/2004/geo/geoStuff.shtml

[12] F. Fonseca and M. Egenhofer (1999). Ontology-Driven Geographic Information Systems. *7th*

ACM Symposium on Advances in Geographic Information Systems, Kansas City, MO.

[13] F. Fonseca, M. Egenhofer, P. Agouris and G. Câmara (2002). Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS 6(3): 231-257.*

[14] The Open Geospatial Consortium, Inc. (OGC). Url: http://www.opengeospatial.org/.

[15] FGDC Metadata. Url: http://www.fgdc.gov/metadata/metadata.html

[16] L. Pouchard et al. Exploring Ontologies in ESG. *Symposium on Cluster Computing and the Grid (CCGrid 2003), Tokyo, Japan, May 12-15, 2003.*

[17] L. Pouchard, L. Cinquini and G. Strand. The Earth System Grid Discovery and Semantic Web Technologies. *Semantic Web Technologies for Searching and Retrieving Scientific Data, ISWCII, Sanibel Island, FL, October 20, 2003.*

[18] R. Raskin and M. Pan. Semantic Web for Earth and Environmental Terminology (SWEET). *Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, Florida, USA, 2003*

[19] Ontologies for Geographic Information Standards provided by ISO.
Url: http://loki.cae.drexel.edu/~wbs/ontology/list.htm.

[20] Saiful, A. S., L. Bermudez, S. Fellah, B. Beran and M. Piasecki (2004). Implementation of the Geographic Information - Metadata (ISO 19115:2003) Norm using the Web Ontology Language (OWL). *Submitted to Transactions in GIS.*

[21] L. Bermudez. and M. Piasecki (2004). Achieving Semantic Interoperability with Hydrologic Ontologies for the Web. *In Proceedings of AWRA's 2004 Spring Specialty Conference Geographic Information Systems (GIS) and Water Resources III.*

[22] A. Chen et al. Interoperability and Semantics for Heterogeneous Earthquake Science Data. *Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, Florida, USA, 2003.*

[23] SchemaWeb – RDF Schemas Directory. Url: http://www.schemaweb.info/.

[24] Protégé: Ontology Library. Url: http://protege.stanford.edu/ontologies/ontologies.html

[25] N. Noy and D. McGuinness. Ontology Development Guide 101: A guide to creating your first ontology.

[26] G. Sposito and S. Mattigod. Geochem: A Computer Program for the Calculation of Chemical Equilibria in Soil Solutions and other Natural Water Systems. *The Kearney Foundation of Soil Science, University of California, 1980.*

[27] USGS Learning Web - Glossary.

Url: http://interactive2.usgs.gov/learningweb/explorer/geoglossary.htm

[28] EPA - Terms of Environment. Url: http://www.epa.gov/OCEPAterms/

[29] D. G. A. Whitten and J. R. V. Brooks. Dictionary of Geology. *Penguin Books, Pg 493.*

[30] NABIR Field Research Center. Url: http://public.ornl.gov/nabirfrc/

[31] NABIR Field Research Center – Well Data. http://public.ornl.gov/nabirfrc/frcdata1.cfm

[31] A. Faatz and R. Steinmetz (2002). Ontology enrichment with texts from the WWW. *Semantic Web Mining 2nd Workshop at ECML/PKDD-2002, 20th August 2002, Helsinki, Finland.*

[32] A. Maedche and S. Staab. Ontology Learning for the Semantic Web. *IEEE Intelligent Systems, Special Issue on the Semantic Web, 16(2), 2001.*

[33] J. Jannink and G. Wiederhold (1999). Ontology maintenance with an algebraic methodology: A case study. *In. Proceedings of AAAI workshop on Ontology Management, July 1999.*

[34] W3C: OWL Web Ontology Language Semantics and Abstract Syntax.

URL: http://www.w3.org/TR/owl-semantics/.

[35] W3C: Resource Description Framework (RDF) Concepts and Abstract Syntax.

URL: http://www.w3.org/TR/rdf-concepts/.

[36] F. Manola and E. Miller. RDF Primer. URL: http://www.w3.org/TR/rdf-primer/.

[37] N. Noy. Representing Classes As Property Values on the Semantic Web.

Url: http://www.w3.org/TR/swbp-classes-as-values/

[38] W3C: OWL Web Ontology language Overview.

Url: http://www.w3.org/TR/owl-features/.

[39] The Protégé Ontology Editor and Knowledge Acquisition System.

Url: http://protege.stanford.edu/index.html.

[40] Protégé OWL plugin – Ontology Editor for the Semantic Web.

Url: http://protege.stanford.edu/plugins/owl/.

[41] ezOWL Plugin for Protégé. Url: http://iweb.etri.re.kr/ezowl/.

[42] J. J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne and K. Wilkinson. Jena: Implementing the Semantic Web Recommendations. *Proceedings of the thirteen international conference on World Wide Web, 2004.*

[43] A. Seaborne. RDQL – A Query Language for RDF.

Url: http://www.w3.org/Submission/RDQL/.

[44] MySQL: The World's most popular Open Source Database. Url: http://www.mysql.com/.

[45] RDFIG Geo vocab workspace. Url: http://www.w3.org/2003/01/geo/.

[46] UDUNITS – A library for manipulating units of physical quantities.

Url: http://my.unidata.ucar.edu/content/software/udunits/index.html

[47] SWEET Ontologies. Url: http://sweet.jpl.nasa.gov/ontology/.

[48] J. Goldbeck, G. Fragoso, F. Hartel, J. Hendler, B. Parsia and J. Oberthaler. The National Cancer Institute's Thesaurus and Ontology. *Journal of Web Semantics, 1(1), Dec 2003.*

[49] Semantic Web Best Practices: WordNet Task Force.

Url: http://www.w3.org/2001/sw/BestPractices/WNET/tf.

[50] Environmental Sciences Division, Oak Ridge National Labs. Url: http://www.esd.ornl.gov/.

[51] Cyc in OWL: http://www.cyc.com/2003/04/01/cyc.

[52] The Ebiquity Research Group. URL: http://ebiquity.umbc.edu/