An Ontology Framework for Modeling Ocean Data and E-Science Semantic Web Services

Samina R. Abidi, Syed SR. Abidi, Mei Kwan, & Ali Daniyal

Manuscript
Received: *8,Sep., 2011*
Revised: *19,Nov., 2011*
Accepted: *7, Jan., 2012*
Published: *15,Sep., 2012*

Keywords
Knowledge Management, Oceans, Ontology, Services-Oriented Architectures, Web Services

***Abstract*— To investigate the underlying characteristics of oceans and relating them with behavioral changes in marine life, we have developed an E-science platform—termed Platform for Ocean Knowledge Management (POKM.) POKM employs a services-oriented architecture and delivers a suite of web services to support collaborative research projects amongst a community of environmental, marine life and oceanographic scientists across the world. To establish interoperability between the concepts from different domains and the data generated by different methods at different geographical sites, we present a knowledge management approach that entails the modeling of the domain concepts and the functionalities of the services in terms of domain and service ontologies, respectively. We present details of the domain and service ontologies at the heart of our E-science platform and demonstrate their application towards ocean knowledge management.**

1. Introduction

The affects of global environment changes are impacting our ecosystem and our oceans are affected in ways that need to be better understood to protect and conserve our coast lines and the marine lives in the oceans. The physical characteristics of our oceans and the marine life within it are experiencing unprecedented changes that need to be understood in order for scientists to respond to the negative effects of environment change. To understand the impact of the changing ecosystem, ocean and marine life scientists are studying a range of physical ocean parameters in tandem with the behavior of marine life. One of the major challenges faced by the scientific community is to retrieve specialized, multi-modal data from global data repositories, then to link the relatively sparse observations on marine life with highly voluminous ocean data, and finally to derive insights from the integrated data through a series of scientific experiments and multi-faceted visualizations. The dispersion of ocean and marine life data across the world, in different formats, demands a collaborative approach for data collection, procurement and operationalization—this approach needs to be supplemented with existing knowledge artifacts (reports, scientific studies, research publications) in order to establish interoperability between the different data sources and also to make sense of the scientific experimental results [1]. So, we need both a collaborative platform to engage researchers from around the world and a scientific platform that provides state-of-the art tools to understand with ocean and marine life data [2-3].

Taking a unique knowledge management approach we have developed an E-science platform—termed as *Platform for Ocean Knowledge Management (POKM)*—to support the oceanographic research community [4]. The functional portfolio of POKM includes a suite of services to (a) *enable* the selection and sharing of multi-modal data collected from different geographic sites, (b) *perform* analytics and simulations, using complex simulation models, to understand various phenomenon such as behavior of marine animals, affects of oceanographic parameters of temperature, salinity, etc; (c) *visualize* multiple data layers at a geographic location and simulation results of models via various globe-based, 2D and 3D plots and animations; (d) *publish* simulation models for use by the entire community of scientists; (e) *interconnect* two different research communities so that they can seamlessly interact and share data, scientific models, experiment results, knowledge resources and expertise without the usual impediments of terminology mismatch, conceptual variances, data heterogeneity and knowledge misalignments and misinterpretations; (f) *catalogue* experiment-specific data and knowledge so that it can be used for future experiments and analytics; and (g) *enable* researchers to design and execute complex experiments by composing specialized experimental workflows—an experiment workflow may entail a systematic arrangement of multiple services, such as data/knowledge collection, simulation models, analytics and visualization—that are suited for their scientific tasks. It may be noted that POKM is supported by the CANARIE network (Canada's high bandwidth network) that allows the transfer of high-volumes of ocean data and to facilitate collaboration between eco-scientists across the world to conduct multi-site scientific experiments.

In this paper we present the functional design of POKM, focusing on its knowledge management components. We discuss the knowledge modeling research

Samina R. Abidi, Syed SR. Abidi , Mei Kwan, & Ali Daniyal are with Faculty of Computer Science, Dalhousie University, Halifax, Canada

Corresponding Author: sabidi@cs.dal.ca
Samina R. Abidi (sabidi@cs.dal.ca) ;
Syed SR. Abidi (sraza@cs.dal.ca)

leading to the definition of two core ontologies—a domain ontology modeling the ocean knowledge and data resources and a service ontology modeling the web service descriptions. We present details of these ontologies and demonstrate how they are applied to support the data management services for the E-Science platform—i.e POKM.

2. POKM Functional Architecture

The design of POKM showcases a unique synergy of semantic web, services oriented architectures, web services and visualization technologies [5-7]. POKM takes a unique knowledge management approach by exploiting semantic web technologies to semantically describe the data, scientific models, knowledge artifacts and web services. This not only allows the seamless interoperability between complex data streams originating from different sources, but it also enables the selection and integration of fine-grained, problem-specific data from large repositories by simply specifying the data needs for the task at hand. POKM pursues a high-level abstraction of ocean and marine science domains to establish a high-level conceptual interoperability between the two domains. This is achieved by developing a rich domain ontology that captures concepts from both domains and inter-relates them to establish conceptual, terminological and data interoperability. To define the functional aspects of the e-science services we have developed a services ontology that provides a semantic description of knowledge-centric e-research services. These semantic descriptions of the e-science services are used to both establish correlations between domain and functional concepts that are the basis for data and knowledge sharing, cataloguing and visualization.

The POKM infrastructure is modeled along a services-oriented architecture that exposes a range of task-specific web services accessible through a web-based portal. POKM's underlying design philosophy is to exploit the web as a services platform to deliver knowledge-centric services for the oceanographic research community—a knowledge-centric service is deemed as a specific function that can help oceanographic researchers conduct their scientific work in a collaborative and knowledge-intensive environment. POKM offers a distributed resource environment in which researchers across multiple nodes are able to collaborate through a suite of knowledge-centric services. Fig. 1 shows the layered functional architecture of POKM.

A. Knowledge Resource Layer

The knowledge resource layer constitutes:

- Data repositories for ocean data, marine life data and simulated data generated through various simulations. At present, POKM is directly connected with the Ocean Tracking Network (OTN) to access their data repositories. In addition, POKM is also able to procure public-domain data from *Ocean Biogeographic Information System* (OBIS) [8]
- Domain-specific knowledge represented as research papers and technical reports. A content management system is used to index the knowledge and to form knowledge clusters of topic-specific articles.
- Domain-specific information represented as images, movies, audio and posters.
- Simulation models shared by the researchers for public use. POKM, therefore, provides a library of specialized simulation models.

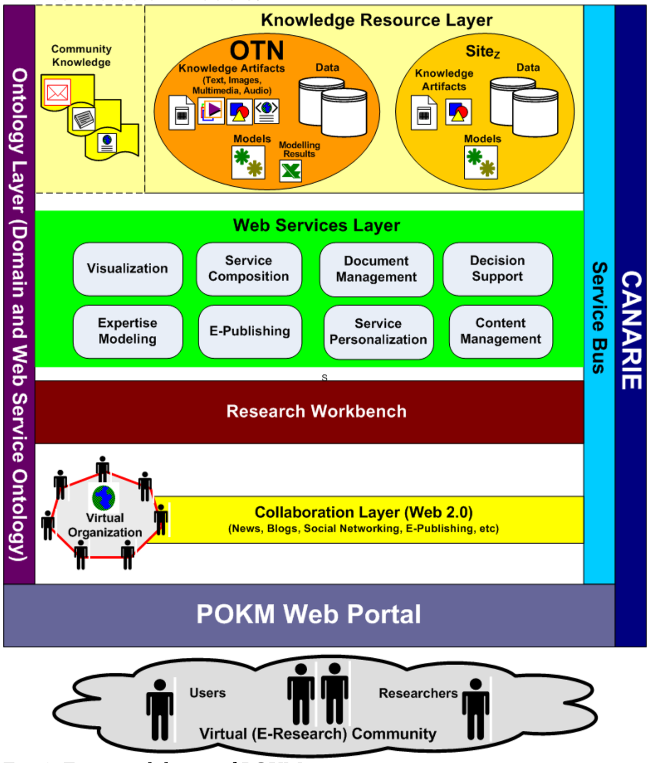

Fig. 1: Functional design of POKM

B. Ontology Layer

The ontology layer serves as the semantic interpretation of the domain concepts and the semantic description of the services. It functions as the glue integrating the various POKM architectural components and services. We have developed two ontologies—(i) domain ontology and (ii) service ontology.

The *Domain Ontology,* developed using OWL [9], provides a high-level abstraction of the oceanography and marine biology domains, and establishes conceptual and functional relationships between intra-domain and inter-domain concepts. The oceanography ontology is used for: (a) semantic integration of the heterogeneous data-sets, especially when new data-sets, models and services are added to POKM; (b) aligning different domain concepts and linguistic terms used by the diverse research community; and (c) normalizing the functional specifications of services to domain concepts and then connecting to data sources.

The *Service Ontology* provides a semantic model for representing the functional description of a service. We used OWL-S [10] to develop a semantic model that captures three aspects of a service; namely, Profile, Model and Grounding. The key feature of the service ontology is that it offers a semantic structure that allows profile and model to provide abstract description of a service using domain concepts specified in the domain ontology. The link between the abstract description and concrete implementation is captured via Web Services Description Language (WSDL) grounding—i.e. mapping concrete notions in the service's WSDL file with abstract concepts in the service ontology.

C. Services Layer

The services layer hosts a range of e-research services. At present, POKM offers eight services with the provision of additional services to be added later. The services layer provides standard Universal Description, Discovery and Integration (UDDI) functionality to register and advertise a service. The interface of a service is described using WSDL, and business processes are modeled using Business Process Execution Language (BPEL) which are executed by the POKM Enterprise Service Bus (ESB).

The service layer distinguishes the granularity of services—(i) *Atomic services* serve a specific purpose, such as fetch data, transform data, document management, decision support, etc; and (ii) *Composite services* that represent a systematic orchestration of a set of atomic services that are combined by the user to perform a complex process (represented as a BPEL business process. At the service layer, each service is described as a tuple of *{Interface, Description, Logic, Semantics}*. The service design allows for semantic expressiveness of a service through the domain and service ontologies that provide a conceptualization of the knowledge and the specification of the service, respectively. The services are divided into the following high-level functional groups: (a) data management services; (b) user management services; (c) visualization services; and (d) modeling services. Each group comprises a suite of services pertaining to their respective functional tasks.

D. Research Workbench Layer

The research workbench layer allows researchers to design and perform experiments using resources and services provided by POKM. The research workbench layer offers a range of capabilities as follows:

Data fetching capabilities comprise a set of data fetching services that enable researchers to specify their data request and then fetch the data from the respective source. The data fetching services retrieve data from (a) databases—this is usually for marine animal tracking and detection data; and (b) data files—this is typically for ocean data represented in netCDF format. The data request can be provided in three ways: (i) using a graphical query generator that uses metadata and semantic descriptions present in the domain ontology to help the researcher construct a formal query; (ii) specifying an SQL query; and (iii) using a visual interface (such as Google Earth (GE) plug-in) to specify the region for which data is required. In this case, the user draws a bounding box—i.e. the longitude and latitude—for the region of interest and then specifies the ocean or animal parameters to be retrieved for that region.

Data transformation capabilities, a set of services that enable the automatic transformation of data from one format to another, either to use the data in a specific model or to visualize it using a specific visualization tool. For instance, a subset of ocean data present in a netCDF file is transformed to KML format to be view in GE plug-in.

Data normalization capabilities, a set of services that enable the normalization of data headers to the standard terminology specified in the domain ontology. Such services are used when researchers need to (a) aggregate data from different data sources; (b) add a new data source; and (c) bind the data to a service description.

Experiment management capabilities enable a researcher to design an experiment and coordinate the resources necessary for conducting it. An experiment constitutes data, the scientific models and algorithms needed for the experiment, and the experiment's workflow. Furthermore, the results of such an experiment need to be presented, shared, visualized and/or stored for further use (any combination). These capabilities retain the user's workspace by managing and maintaining (a) the user's data requests and handling the storage and caching of the retrieved data in the user's workspace for subsequent use; (b) the experiment workflow as specified by the user; and (c) the results generated by experiments so that it can be either reused for additional experiments, presented and visualized for analysis purposes or shared with other users.

Visualization capabilities enable scientists to derive more meaningful insights from the data and the experimental results. POKM has developed a dedicated data visualization framework that allows scientists to visualize and interact with multiple layers of data as time- and location-varying animations, globe-based views and a range of multi-dimensional plots. At present, the visualization capabilities of POKM offer the following functionalities: (i) visualizing marine animal tracks, such as the migration patterns of leatherback turtle, as an animation using GE plug-in; (ii) visualizing the parameters of the ocean—such as salinity, temperature, current—for a bounded region for a specific time period. A time-varying animation is presented using GE plug-in; (iii) visualizing an integration of marine animal data with ocean data. For instance, visualizing animal movement in the backdrop of ocean temperature at the corresponding time and location. This is a time-varying animation visualized using GE plug-in; (d) visualizing graphical 2D and 3D plots—such as time series, contour and mesh plots, etc—of ocean parameters or simulation results; (e) visualizing a combination of time-varying animations and viewing a 2D/3D plot of a related ocean parameter in the same session; and (f) visualizing 2D images of experimental results, such as the output of an filtering model depicting the predicted locations of a leatherback turtle over a time period as an image. The above visualization functionalities are available through the POKM portal and are rendered through individual portlets.

E. Collaboration Layer

The functionality of the collaboration layer is to facilitate collaboration between researchers around specialized research topics. The key feature of the collaboration layer is that it allows the formation of specialized communities of interest, whereby within a specialized community the participating researchers can (a) share data, models and literature; (b) use web 2.0 tools for collaborative networking, discussion forums, experiment-specific blogs and share pertinent news and announcement; (c) jointly develop complex experiments; (d) publish simulation models for use by the community; and (e) customize the view of the community web space by adding different portlets. At this stage, we have created a Leather Back Turtle community comprising oceanographers and marine life researchers.

3. Ontology Based Domain Modeling

The domain ontology in POKM serves as the formal semantic description of the concepts and relationships pertaining to the Marine Biology and the Oceanography domain. POKM provides a core ontology that contains concepts necessary for modeling Marine Animal Detection Data (MADD), Oceanography Data, data transformations and interfaces of the Web Services in POKM. The taxonomic hierarchy of the domain ontology constitutes 20 highest level classes; 15 of these classes are further decomposed into sub-classes at the lower levels of hierarchy. The domain ontology is developed in OWL and captures a range of ocean and marine life concepts grouped into six main classes (see Fig. 2).

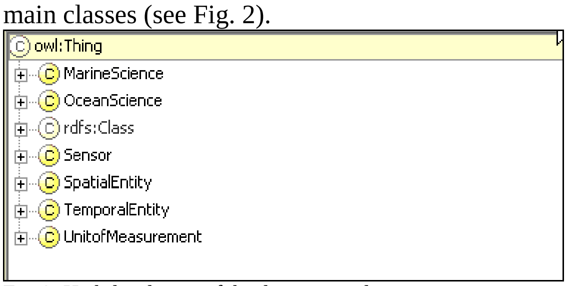

Fig. 2: High-level view of the domain ontology.

A. Modeling Marine Sciences

There are six upper level classes related to marine sciences—i.e. MarineOrganism, AnimalDetail, Taxonomy,TaxonID,MarineLifeData,MarineLifeData Collection, ataSource, DataFormat.

MarineOrganism represents all marine animals, plants and plankton via classes MarineAnimal, MarinePlant and Plankton respectively. There are four main subclasses of MarineAnimal: Fish, MarineMammal, Reptile and SeaBird. MarinePlant has two main sub-classes: Algae and Seagrasses. Plankton has three sub-classes representing three functional groups of planktons: Bacterioplankton, Phytoplankton and Zooplankton.

AnimalDetail represents all the necessary information to build a marine animal profile. It has five main sub-classes: Age, LifeStage, MovementBehavior, Sex, TagID. Age represents the age of the animal. LifeStage represents the current stage of the animal in life, e.g. adult, juvenile, sub-adult etc. MovementBehavior represents various movement behaviors of marine animal that are captured by its sub-classes: BehavioralSwitching, Dispersal, Diving, Drift, Foraging, Migrating and MovementPattern.

Taxonomy represents nine main taxonomic ranks used to categorize marine organisms as follows: Class, Family, Genus, Kingdom, Order, Phylum, ScientificName, ScientificNameAuthor and Species.

TaxonID describes an organism in terms of the above mentioned nine taxonomic ranks.

MarineLifeData represents various aspects of the data about the marine organisms. These include temporal data represented by sub-classes: DayCollected, MonthCollected, YearCollected, DateLastModified and TimeStampCollected, which has two sub-classes of its own: EndTimeStampCollected and StartTimeStampCollected. The class MarineLifeData is also used to represent concepts related to the cache of the marine data that is represented using sub-classes such as CacheID, RecordLastCached, BasisOfRecord and ResourceID. This class also represents other aspects of marine life data using sub-classes: Depth, DepthPrecision, Temperature and TimeZone.

MarineLifeDataCollection is a class the properties of which are used to capture all the data represented by class MarineLifeData.

B. Modeling Ocean Sciences

The classes to model ocean sciences include: OceanRegion, OceanParameter, SatelliteInformation, Instrument, Measure, MovementModel, ModelAttribute, FileType,

OceanRegion represents all ocean regions categorized by five main sub-classes: ArcticOcean, AtlanticOcean, IndianOcean, PacificOcean and SouthernOcean. Each of these classes are further sub-divided into sub-classes representing sub regions of the each ocean region.

SatelliteInformation represents the satellite used to monitor the oceans, represented in terms of nine sub-classes: SatelliteID, Altitude, BestSignalStrength, FrequencyOfTransmission, ElapsedTime, NumOfMessagesRecieved, NumOfSuccessfulPlausibleChecks, QualityIndicator and SensorChannel.

Instrument represents all the instruments used for the observation of oceans and to measure various parameters, such as: temperature, salinity and density of the ocean water, ocean currents, depth, pressure, etc. These instruments are represented as the following sub-classes: ADCP, Argos, ArgoFloat, CTD, ElectronicTag, Glider, GlobalPositioningSystem, Satellite and SubmersibleRadiometer.

Measure represents all the spatial and temporal measures of the regions used in the domain of Ocean Sciences, and are modelled as two main sub-classes SPATIALMEASURE and TEMPORALMEASURE respectively. The sub-class SPATIALMEASURE has further sub-classes: HEIGHT, LATITUDE, LONGITUDE AND SPATIALRESOLUTION representing the respective spatial measures of the relevant ocean region. TEMPORALMEASURE has two sub-classes: TIMEINTERVAL and TIMERESOLUTION, representing the respective temporal measures.

MOVEMENTMODEL represents various models used to estimate the migrating and foraging behaviors of marine organisms and their movement parameters such as determining the next positioning estimate of an animal after a period of missing data. These models are represented as sub-classes: FIRSTPASSAGETIME, FRACTALANALYSIS, GEOLOCATIONMODEL, KERNELANALYSIS, STATESPACEMODEL.

MODELATTRIBULE represents all the attributes of a movement model, represented as sub-classes: HIERARCHY, LINEARITY, OBSERVATIONERROR, OUTPUT, STATISTICALESTIMATIONERROR, STATISTICALFRAMEWORK, STOCHASTICITY and TIME.

UNIT represents all the units used to measure geophysical parameters describing an ocean. It has nine sub-classes: DENSITYUNIT, DEPTHUNIT, LIGHTLEVELUNIT, SALINITYUNIT, SPATIALRESOLUTIONUNIT, SPATIALUNIT, TEMPERATUREUNIT, TIMEUNIT, VELOCITYUNIT.

C. Relationships Between Classes

The purpose of the domain ontology is to inter-relate the domains of Marine Sciences and Ocean Sciences. There are seventy seven object properties and six datatype properties. We describe only the salient properties are described in this section.

The class MARINEANIMAL (sub-class of MARINEORGANISM) is related to respective sub-classes of the class MARINELIFEDATA through properties has_age, has_sex, has_life_stage, has_movement_behavior and has_TagID. In addition it is also related to class OCEANREGION through property has_geographic_area. Thus, this property relate the domains of marine sciences and ocean sciences.

The class OCEANPARAMETER is related to class Unit through property has_unit. This property is given hasValue restriction, to restrict the filler of the property to a specific instance of the class UNIT. For example AirTemperature, which is an OceanParameter has_unit Degree Celsius, which is an instance to class UNIT.

The class MARINELIFEDATACOLLECTION is related to respective sub-classes of class MARINELIFEDATA through properties has_basis_of_record, has_cache_ID, has_date _last_modified, has_day_collected, has_depth, has_depth_precision, has_latitude, has_longitude, has_month_collected, has_record_last_cached, has_record_ID, has_taxon_ID, has_temperature, has_time_of_display_collected, has_time_zone_collected and has_year_collected. Each one of these properties is a functional property.

The class MOVEMENTMODEL is related to respective sub-classes of class MODELATTRIBUTE through properties: has_hierarchical, has_input_data, has_linearity, has_observation_error, has_output, has_statistical_ estimation_method, has_statistical_framework, has_stochasticity and has_time_value.

Each OCEANREGION is related to various OCEANPARAMETERS through properties: has_density, has_flow_velocity, has_salinity, has_sea_surface_elevation, has_water_depth, has_water_mass and water_temperature. Class OCEANREGION is also related to respective sub-classes of class MARINELIFE through sub-classes has_marine_animal, has_marine_plant and has_plankton. Note that these three sub-classes relate the ocean sciences domain with marine sciences domain.

The class TAXONID is related with respective sub-classes of class TAXONOMY, in order to capture the identification features of each of the marine species. These properties are: has_class, has_family, has_genus, has_kingdom, has_order, has_phylum, has_sceintific_name, had_scientific_name_author and has_species. Each one of these properties is a functional property.

4. Ontology Based Service Modeling

A. Capturing Semantics of Web Services using OWL-S

To enable the system to capture semantic description of Web Services, POKM employs OWL-S ontology in conjunction with the Domain Ontology. OWL-S provides a semantic model for capturing three aspects of a Web Service: namely; Profile, Process Model and Grounding. Since we employ UDDI as a registry for web services, the Profile part of the OWL-S ontology is not instantiated in POKM, to avoid duplication. POKM only employs the Process Model to capture IOPE of the Web Services and Grounding to capture semantics of the concrete WSDL description of web services.

The Process Model in OWL-S provides classes to capture the notions of Parameters, Inputs, Outputs, Preconditions and un/conditional Effects of a process. The property parameterType is used to relate a parameter with its semantic type in terms of an OWL class. An Atomic Process in OWL-S may have multiple inputs and multiple outputs, however when modelling semantics of an operation of a web service as an OWL-S Atomic Process, they are mapped to the parts of the input and output messages. This mapping is captured via the WSDLGrounding class in the Grounding part of OWL-S.

Each Atomic Process in OWL-S supports exactly one WSDLGrounding that in turn provides a unique WSDLAtomicProcessGrounding. This class facilitates capturing mapping between the message parts of the web service operation and the parameters of the OWL-S Atomic Process. Fig. 3 illustrates how OWL-S is employed in POKM to describe semantics of web services.

B. Web Services Deployment and Publishing Life-Cycle

One of the main functionalities of the POKM is to allow users to enrich the pool of web services accessible via POKM. The main intention behind this functionality is to enable scientists to publish their scientific models as web services on the POKM, so that these models can be used in more complex processes and can be shared with other users. To achieve standardization and interoperability all the scientific models are exposed as web services, described by WSDL and compliant with the SOAP messaging protocol. POKM supports three different scenarios of publishing web services on the system: namely, (i) deploying WAR file on the POKM ESB, (ii) publishing scientific models as R-scripts, and (iii) publishing WSDL file on POKM UDDI.

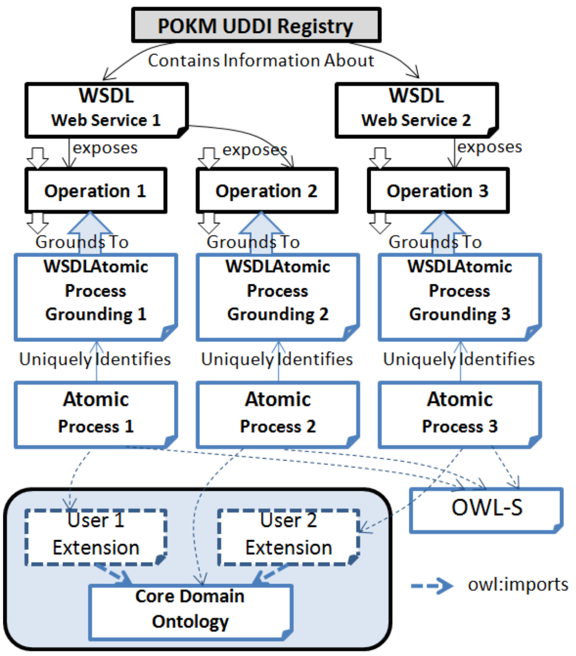

Fig. 3: Semantic Modeling of Web Services in POKM using OWL-S

5. Ontologies in Action: Ocean Data Management

One of the functionalities of POKM is to seamlessly connect multiple data sources to obtain Marine Animal Detection Data (MADD). Modeling of the domain concepts in the domain ontology enables the integration of the data sources by mapping both schemas and attributes based on concepts as opposed to labels. Furthermore, the domain ontology allows the use of domain concepts to generate abstract data fetching queries, in terms of standard operations, that can provide access to heterogeneous data sources. In this section, we illustrate the data management features of POKM supported by the domain and service ontologies.

A. Semantic Description of Data Sources

POKM accesses MADD sources that are implemented as relational databases. Since each data source is implemented as a different relational structure, in order to query the datasets in a uniform manner it is required to have a description of the corresponding schema in a common vocabulary. To address this problem we map columns of the tables in the relational structure to concepts in the domain ontology. This allows POKM to entertain data queries made in terms of concepts in the domain ontology against selected data sources, irrespective of the underlying schema, and present the fetched data in ontological terms—standard concepts as opposed to disparate labels. Presenting data obtained from different data sources in a common vocabulary helps reducing ambiguities in data interpretation at the end-user level and allows automatic alignment and aggregation of multiple data sets.

B. Uniform Access to Data Sources

To enable uniform access to heterogeneous MADD sources via POKM, we have defined two core standard operations for querying these data sources: namely, getCoverage and getDetections. To facilitate this querying mechanism, a Data Retrieval Service (DRS) is implemented in POKM for each MADD source. The getCoverage operation provides spatial coverage in terms of rectangular region defined by top left and bottom right geographical coordinates in terms of latitude and longitude, as well as, temporal coverage in terms of minimum and maximum values of time stamps associated with detection records. The getDetections operation returns detections records as Comma Separated Vectors (CSV) for spatial and temporal coverage provided as arguments. In order to manipulate these CSV formatted records, automatically, for each DRS a mapping between columns of the header of the CSV generated by getDetections operations and concepts of the domain ontology.

C. Data Fetching Scenario

We discuss below how we make use of the knowledge, stored in the ontological model, about the MADD sources and DRS, to fetch MADD from different resources. We have developed a Meta Animal Data Retrieval Service (MADRS), in order to dynamically discover newly added DRS in the POKM system and to provide a uniform mechanism of invoking the registered DRS.

The MADRS operates on the domain ontology, capturing semantics of MADD sources, to discover DRS described in the ontology. It provides an operation named getDataSources to enable POKM Portal or a third party application to discover MADD sources that are accessible via POKM.

By exploiting the standardized signature of the getCoverage operation, MADRS is capable of providing coverage of any of the underlying MADD sources. Similarly an operation is provided to request detection records from a particular MADD source by providing spatial and temporal coverage. MADRS is also capable of automatically transforming detection data from CSV to KML. Note that this is possible due to the mapping between the CSV headers and concepts in the domain ontology stored against every DRS. This allows POKM end-users to generate data fetching requests from any selected MADD source and visualize the results without any human intervention.

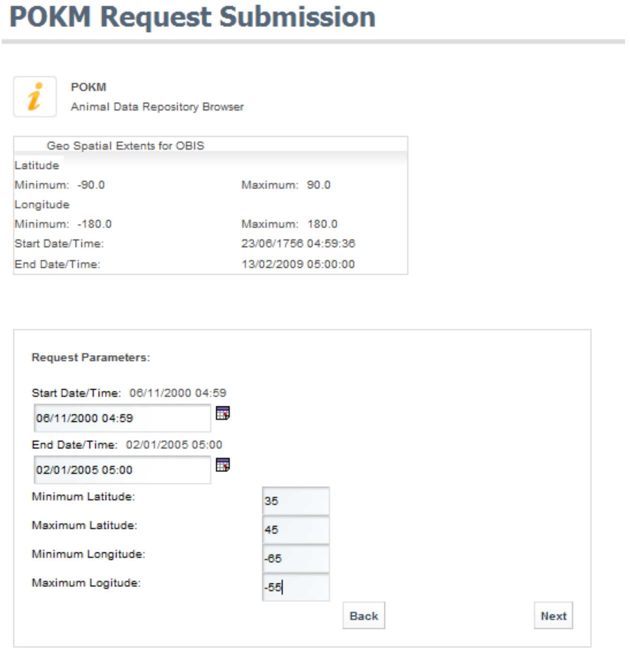

Fig. 4: Portlet for Requesting MADD Data on the POKM Portal employs MADRS for discovering MADD sources and obtaining corresponding coverage

Fig. 4 and Fig. 5 provides snapshots of the two portlets in the POKM portal that make use of the MADRS in order to request detection data from a selected MADD source and visualize it.

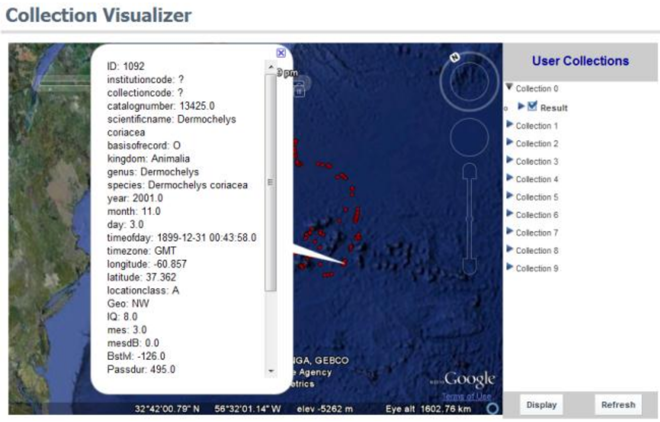

Fig. 5: The Collection Visualizer portlet makes use of the MADRS to obtain requested data in renderable format. The portlet uses GE Plugin to render KML files generated using MADRS.

6. Concluding Remarks

POKM offers an E-science platform targeting ocean sciences researchers. The key feature of POKM is the underlying knowledge layer—manifested in terms of the semantics of ocean and marine life concepts, captured in the domain ontology and descriptions of the services captured in the service ontology—that enables the seamless integration and interoperability at the data, services and user levels. The domain ontology presented here is scalable to include new concepts and relationships between concepts. The services ontology is coupled with interesting semantic-based methods to discover compatible and relevant services pertaining to a specific user task. In this paper, we have demonstrated the potential of applying knowledge management methods, specifically the use of OWL-based ontologies to semantically describe the domain and service concepts and to operationalize these concepts in terms of e-science services.

The POKM project is currently in operation and used by a community of ocean and marine life scientists in Canada. These scientists are using POKM to study the migration patterns of leatherback turtles across the Nova Scotian shelf in the Atlantic ocean.

Acknowledgment

This project is funded by CANARIE Canada under the Network Enabled Program (NEP). The authors acknowledge the support of CANARIE Canada to develop POKM. The authors acknowledge the support of members of the POKM team, especially the document management, marine biology and ocean data modeling teams.

References

[1] C. Hine, New infrastructures for knowledge production: Understanding e-science, London: Information science publishing, (2007).

[2] N. Bos, A. Zimmerman, J. Olson, J. Yew, J. Yerkie, & E. Dahl, "From shared databases to communities of practice: A taxonomy of collaboratories", in *Journal of Computer-Mediated Communication*, vol. 12 (2), 2007, pp. 652-672.

[3] J. N. Cummings & S. Kiesler, "Collaborative research across disciplinary and organizational boundaries," in *Social Studies of Science*, vol. 35 (5), 2005, pp. 703-722.

[4] NW. Jankowski, "Exploring e-Science: An Introduction", in *Journal of Computer-Mediated Communication*, vol. 12(2), 2007, pp. 549-562.

[5] D. De Roure & JA. Hendler, "E-Science: The grid and the Semantic Web", in IEEE Intelligent Systems, 2004.

[6] E. Laure, F. Hemmer, &F. Prelz, "Middleware for the next generation grid infrastructure", in Computingin High Energy Physics and Nuclear Physics, 2004

[7] C. Lin, S. Lu, &X. Fei, "A Reference Architecture for Scientific Workflow Management Systems and the VIEW SOA Solution", in IEEE Trans. On Services Computing, vol. 2(1), 2009, pp. 79-92.

[8] OBIS. Ocean Biogeographic Information System, viewed May 13, 2011, <http://www.iobis.org/>

[9] OWL: Web Ontology Language, reviewed on May 13, 2011, <http://www.w3.org/TR/owl-features/>.

[10] OWL-S: Semantic Markup for web services, viewed May 13, 2011, <http://www.w3.org/Submission/OWL-S/ >