Semantic Web Services for Ocean Knowledge Management

Syed SR. Abidi, Ali Daniyal, Ashraf Abusharek, Samina R. Abidi

***Abstract*—**We present a web-services based e-research platform to support scientific research about oceans and marine life. The featured Platform for Ocean Knowledge Management (POKM) employs a services-oriented approach to offer a suite of web services to support the handling, filtering and visualization of large-scale marine-life and ocean data sets. Taking a knowledge management approach, we have employed semantic web technologies to model both the domain knowledge and the service descriptions. This allows a seamless interoperability between the different data-sets and services. The web-services are supported by Canada's CANARIE high bandwidth network that enables the transfer of large volumes of data from multiple sites and the execution of the various web services. In this paper, we present an overview of the architecture and functionality of POKM.

***Keywords*—**Knowledge Management, Services-Oriented Architectures, Web Services.

I. INTRODUCTION

THE ocean scientific community is quite adept at generating large volumes of data on the physical parameters of the ocean and the behavioral characteristics of marine life. But, one of the major challenges faced by research community as a whole, is linking the relatively sparse observations on animal movement with the voluminous output from ocean models, and then using aggregated data with simulation models to make predictions about (a) changes in animal movement on time scales of days to decades; (b) coastal flooding due to changes in certain ocean parameters; (c) fish colonies and stocks; and (d) time-varying physical structure of the oceans. The study of such complex multi-faceted scientific questions demands innovative computing solutions—solutions that transcend beyond new ways of capturing eco-system data to dedicated network-enabled, technology-driven knowledge management info- and infrastructures to manage and access, process, analyze, visualize share the data/knowledge owned by the research community.

We present an innovative E-Science platform— termed *Platform for Ocean Knowledge Management (POKM)*—that is built using innovative web-enabled services, services-oriented architecture, semantic web, workflow management and data visualization technologies. POKM offers a suite of web services [7] that allow oceanographic researchers to (a) handle large volumes of ocean and marine life data; (b) access, share, integrate and operationalize the data, models and knowledge resources available at multiple sites; (c) collaborate in joint scientific research experiments by sharing resources, results, expertise and models; and (d) form a broad, *virtual community* of national and international researchers, marine resource managers, policy makers and climate change specialists. POKM exploits the high-bandwidth offered by CANARIE for multi-site integration of high-volume data for data modeling and simulations at distributed user sites, whilst broadcasting the high-dimensional results to a wide range of users.

In this paper we present the high-level technical architecture of POKM and the e-Research technologies developed. We also provide an overview of the operation of key data handling web services to support oceanographic research.

II. POKM TECHNICAL ARCHITECTURE

The technical architecture of the POKM is inspired by recent developments in the realm of Semantic Web and Semantic Web Services [2]. The POKM architecture is modeled along a services-oriented architecture that exposes a range of task-specific web services accessible through a web 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, such as a data fetching and integration service that allows to fetch ocean data with respect to geographical and temporal criterion and then integrates the ocean data with marine animal tracks.

POKM presents a highly distributed resource environment in which researchers/users distributed across multiple nodes (where some nodes are connected with CANARIE) are able to collaborate through a suite of web services. The distributed nature of POKM is inherent in the (a) distribution of the data, model and knowledge resources across multiple nodes; (b) diversity of users that are situated at different nodes; and (c) distribution of OTN partners and data contributors across another set of nodes. POKM, leveraging CANARIE, therefore serves as the glue that connects distributed data, models and

---

Raza Abidi is with the Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada. (phone: 1-902-4942129; e-mail: sraza@ cs.dal.ca).

knowledge resources to users who are in turn connected to services. The four main components of the POKM environment are (see Fig. 1):

1. *POKM* technical infrastructure which realizes the SOA architectural components, including semantic web services, enterprise service bus, UDDI and a workflow management system.

2. *OTN* as the main data resource backbone –i.e. the provider of data, analytical models, users and knowledge.

3. *CANARIE* as the high bandwidth infrastructure to enable the transfer/sharing of high-volume data and models at high-speeds between multiple users across the world.

4. *Users* who will use the POKM services and contribute to POKM's data, models and knowledge resources.

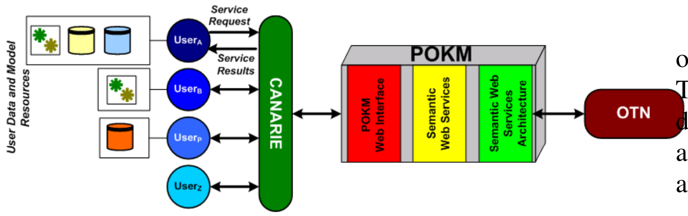


Fig. 1 POKM Environment

Fig. 2 illustrates the functional architecture of POKM, highlighting its functional layers, application components, data/information/knowledge sources, user-interfaces and message flow between software components. The POKM software architecture features five distinct functional layers—*Presentation Layer*, *Collaboration Layer*, *Service Composition Layer*, *Service Layer* and *Ontology Layer*. These functional layers are orchestrated by *POKM Service Bus* that coordinates the overall workings of POKM.

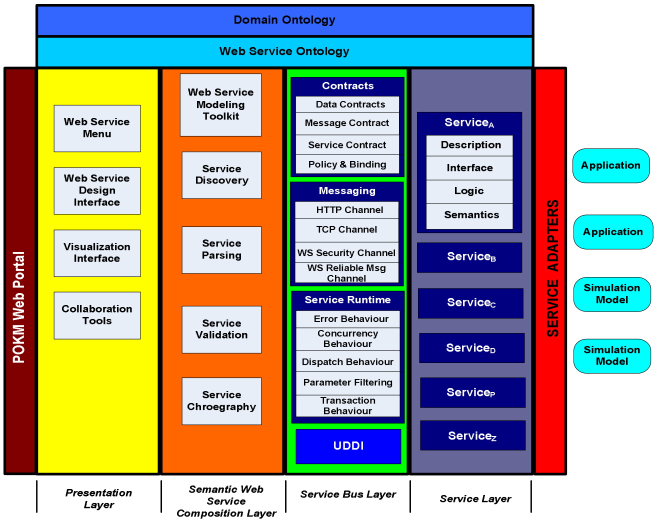


Fig. 2 POKM functional architecture.

A. POKM Portal and Presentation Layer

The POKM Portal provides direct access to all services provided by the POKM SOA. The pages on the POKM portal are designed as *portlets* that provide graphical interface to the POKM web services. User authentication at the portal is through an OpenSSO identity server that authenticates both the POKM users and POKM web services.

At the POKM portal, each service-specific portlet provides graphical interface to both display service-related information and provide access to specific service-related functionalities. Our design approach is to separate the display logic with the processing logic so that portlet designs are generic. The messaging between the POKM portal and the backend processing services is based on SOAP, whereas the content of the messages is encoded as XML payloads. The POKM portal serves as the interface to the presentation layer which offers users a self-explanatory and intuitive Web-based interface to access the back-end POKM services.

B. Ontology Layer

The ontology layer will constitute of two ontologies—(i) oceanography (domain) ontology and (ii) service ontology. The ontology layer serves as the semantic interpretation of the domain concepts and the semantic description of the services, and hence functions as the glue integrating the various POKM architectural components and services.

The *Oceanography Ontology* is developed in the OWL [3] language. The oceanography ontology provides a high-level abstraction of the oceanography and marine biology domains, and provides a conceptual and functional relationship between intra-domain and inter-domain concepts. We used the Model-based Incremental Knowledge Engineering approach for ontology engineering. The oceanography ontology is used for the following functions: (a) semantic integration of the data and knowledge artifacts; (b) alignment of different concepts and linguistic terms used by the diverse research community; (c) normalizing the new service composition specifications provided by users; (d) functional constraints governing the dynamic composition of user-specific services. The role of the domain ontology as the binding glue is quite central because the framework aims to integrate multiple knowledge resources and bring together researchers from different backgrounds.

The *Service Ontology* provides a functional description of the advertisement and specifications of a service. It is developed using OWL-S [4]. For each service we have provided an OWL-S profile in terms of the standard IOPR (*Input-Output- Precondition-Result*) specifications. To coordinate the proper execution of a service, we define *Preconditions* that need to hold and a set of *Postconditions* that need to be achieved after the service finishes execution (*Result*), and finally we define the *Input* and *Output* describing the I/O functional descriptions.

C. Service C*omposition Layer*

The service composition layer allows the composition of new composite services by using and fine-tuning existing atomic services. Although, POKM offers a wide-range of pre-defined services and the provision to add third-party services, we anticipate that there will be times when the existing services are insufficient to meet a user's more complex or unique needs. In this situation, POKM allows users to dynamically compose a problem-specific service by providing a high-level description of the task's objective.

The functionality of the service composition layer is to provide scientists the functionality to systematically integrate multiple task-specific services, as per a specific experimental workflow, to compose a complex task-specific service. The layer provides mechanisms for (a) the selection of multiple atomic services; (b) establishing interoperability between the two connecting services at the input and output interfaces—this is achieved by reconciling the semantic types of the input-output variables of two interconnecting services; (c) composing an execution workflow in a high-level workflow language, such as BPEL [5]; and (d) publishing the experimental workflow with the accompanying services as a composite web service. The architecture of the Service Composition Layer is as follows (see figure 3):

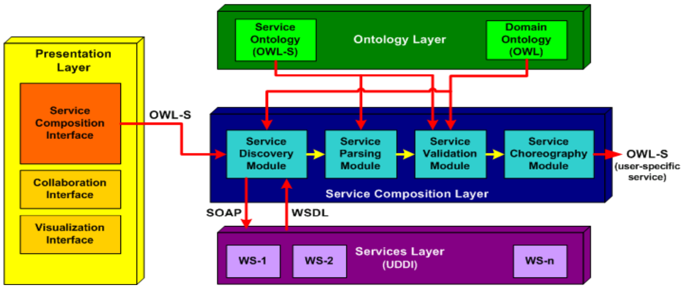


Fig. 3 Architecture of the Service Composition layer

D. Services Layer

The Services Layer hosts the various POKM web *services*—the services are classified into five main categories. The services layer uses a standard UDDI functionality to register and advertise a service. Each service is represented as a tuple of {Interface, Description, Logic, Semantics}, as shown in Fig. 4. Semantic expressiveness has been achieved by using domain and service ontologies that provide a conceptualization of the knowledge and the specification of the service, respectively.

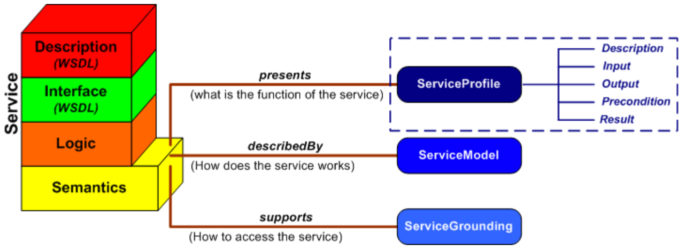


Fig. 4 The structure of a Semantic Web Service

Technically speaking, each service has the following components:

- *Service Profile* provides a concise representation of service capabilities through the advertising of the functionalities description
- *Service Model* gives a detailed description of how the service operates, specifically describing the various transformations (i.e. the processes) that the service undertakes
- *Service Grounding* provides the details on how the service will interoperate with other service—i.e. mapping the messages (according to the format and input/output specification provided in the process model) to the syntactic WSDL compliant form.

E. POKM (Enterprise) Service Bus

The POKM Service Bus (PSB) constitutes the middle tier of the POKM architecture. The PSB manages the SOA principles, such as exploiting meta-data describing interaction endpoints as well as the domain models used to describe the capabilities of those endpoints; supporting configuration of links that bridge between capabilities demanded by service requestors and those offered by service providers; dynamically matching requestors with providers and in the process establishing and enacting contracts between those interaction endpoints.

The PSB has been developed using GlasshFish ESB with the following functional components:

- *Glassfish Application Server*: The POKM system has been modeled as a set of web applications that consist of a portal server, portlets, supporting applications, web services and business processes.
- *UDDI Server*: This component provides a registry for Web Services that are registered with POKM—the web services could be internal. The UDDI provides standard mechanisms for publishing and querying of Web Services descriptions.
- *Service Execution Engine*: This component provides runtime for deployment and execution of POKM Services that are WSDL and SOAP compliant.
- *Process Execution Engine***:** This component provides a runtime environment for deploying and executing BPEL defined composite processes.
- *OpenSSO*: This component provides an identity server for both portal level and web services level authentication and authorization based on the single-sign-on framework.

III. POKM IN ACTION

In this section we demonstrate the working of the different services of POKM with respect to scientific experiments. We will demonstrate the functionalities of the data fetching services and the data visualization services.

A. Data Fetching Services

POKM provides access to two types of data—oceanographic data and marine animal detection data. We have developed a service that crawls through a data source and notes the data points available at each geographical position—these data points are represented as data tiles on the globe. Fig. 5 shows the data tiles for the OBIS [6] data resource.

Fig. 5 Screen shot of the POKM system displaying Tiling based Metadata Visualization of OBIS detection data source.

1) Marine Animal Detection Data (MADD) Fetching Service

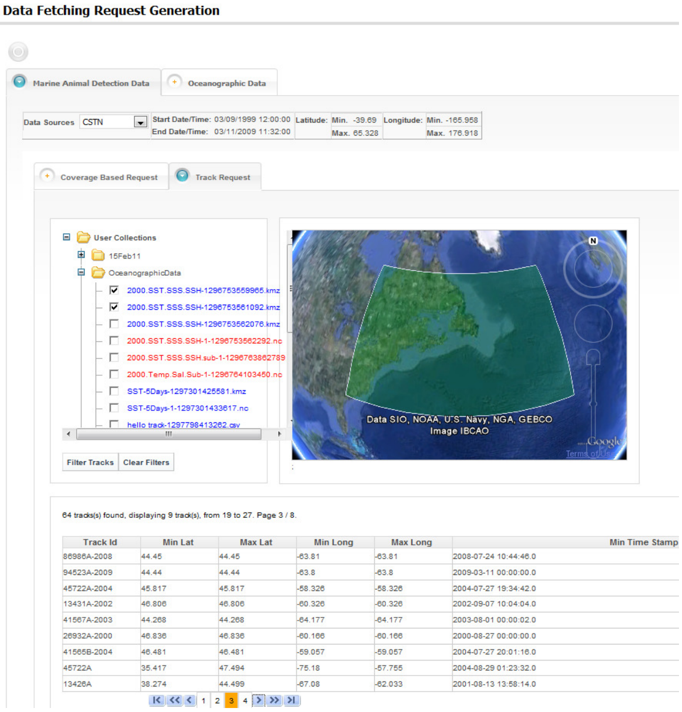
Marine animal data can be retrieved based on two methods:

Coverage of a geographical region—i.e. all marine animal detections within the specified region are retrieved. In an interactive manner, the user can interactively specify the region by manipulating the region selection window on the map (see Fig. 6). In addition, users can also specify constraints on their data query by selecting features and desirable values of those features. The retrieved data can be transformed into a KML file so that it can be visualized as a time-varying animation on Google Earth.

Track of an animal—i.e. the entire track of animal is retrieved showing the position of the animal. The data fetching portlet features a collection browser (shown on the left of the map), which can be used to specify coverage of interest for filtering tracks based on the coverage of already retrieved ocean data. The list of tracks can then be further filtered based on keyword search on different features of the tracks listed in the table below the map.

2) Ocean Data Fetching Service

Ocean data retrieval is based on the coverage based approach—the user specifies the region and the ocean data layer (as shown in Fig. 7). The ocean data portlet features a collection browser (on the left of the map) which can be used to select already generated data files to define coverage of interest for the oceanographic data. User can then apply filter on the oceanographic data shown below the map based on this coverage.

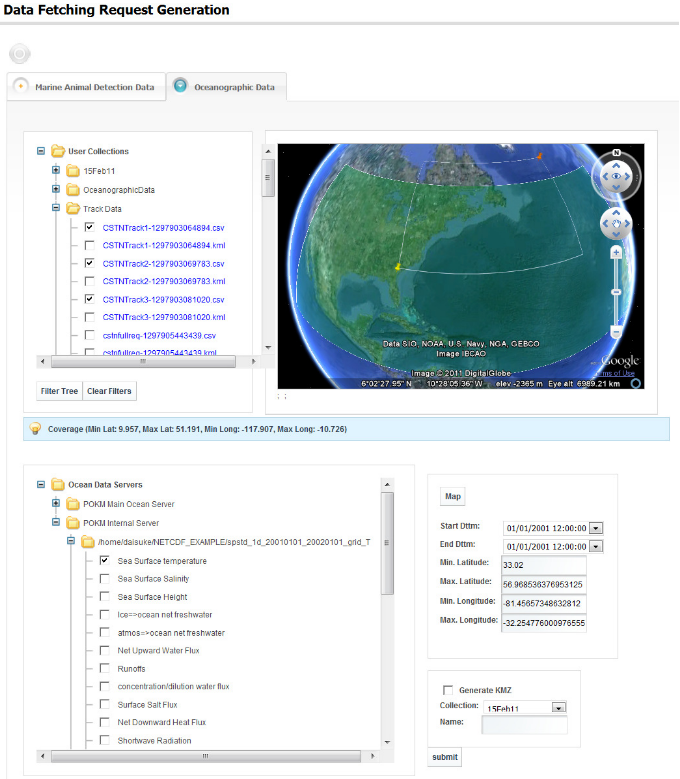


Data Sources: CSTN
Start Date/Time: 03/09/1999 12:00:00
End Date/Time: 03/11/2009 11:32:00
Latitude: Min. -39.69 Max. 65.328
Longitude: Min. -165.958 Max. 176.918

64 tracks(s) found, displaying 9 track(s), from 19 to 27. Page 3 / 8.

Track Id	Min Lat	Max Lat	Min Long	Max Long	Min Time Stamp
86986A-2008	44.45	44.45	-63.81	-63.81	2008-07-24 10:44:46.0
94523A-2009	44.44	44.44	-63.8	-63.8	2009-03-11 00:00:00.0
45722A-2004	45.817	45.817	-58.326	-58.326	2004-07-27 19:34:42.0
13431A-2002	46.806	46.806	-60.326	-60.326	2002-09-07 10:04:04.0
41567A-2003	44.268	44.268	-64.177	-64.177	2003-08-01 00:00:02.0
26932A-2000	46.836	46.836	-60.166	-60.166	2000-08-27 00:00:00.0
41565B-2004	46.481	46.481	-59.057	-59.057	2004-07-27 20:01:16.0
45722A	35.417	47.494	-75.18	-57.755	2004-08-29 01:23:32.0
13426A	38.274	44.499	-67.08	-62.033	2001-08-13 13:58:14.0

Fig. 6 A screen shot of the Data Fetching Portlet showing Track Fetching Request Submission tab.



Coverage (Min Lat: 9.957, Max Lat: 51.191, Min Long: -117.907, Max Long: -10.726)

Start Dttm: 01/01/2001 12:00:00
End Dttm: 01/01/2001 12:00:00
Min. Latitude: 33.02
Max. Latitude: 56.968536376953125
Min. Longitude: -81.45657348632812
Max. Longitude: -32.254776000976555
☐ Generate KMZ
Collection: 15Feb11
Name:

Fig. 7 A screen shot of the Data Fetching Portlet showing Oceanographic Data Request Submission tab.

B. Visualization Service

POKM provides a suite of data visualization services, categorized as globe-based visualization and plot-based visualization services. Functionally speaking, we have developed a visualization workbench that allows scientists to (a) select data/results to be visualized; (b) set-up different visualization scenarios with the same data-set; (c) set-up different visualizations for different data-sets with the intention of comparing the different data-sets; and (d) saving the visualization settings for future use.

The globe-based visualizations plot the data as static and animations on a Google Earth interface. We have implemented dedicated scripts to control the time bar so that the scientist has the ability to visualize the data with different timing setting.

Now to visualize the data, the scientist is provided with globe-based visualizations of the retrieved data. Fig. 8 shows a visualization of an animal track, whereas

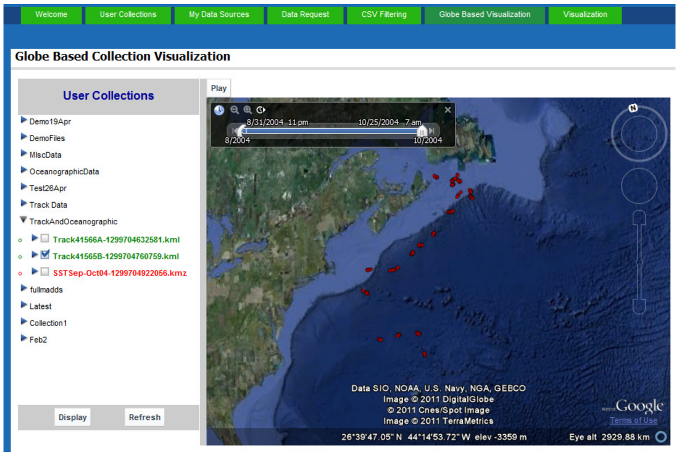


Fig. 8: Screen shot of a Globe Based Visualization portlet displaying an animal track in red dots on the globe

POKM also offers a suite of plots—2D, 3D, 4D, Mesh and Contour plots—that the scientist can use to analyze the data (See Fig. 9). The key feature of POKM is that it allows the scientist to specify the plot parameters in an interactive manner, and these setting can be re-used to develop multiple plots with minor modifications.

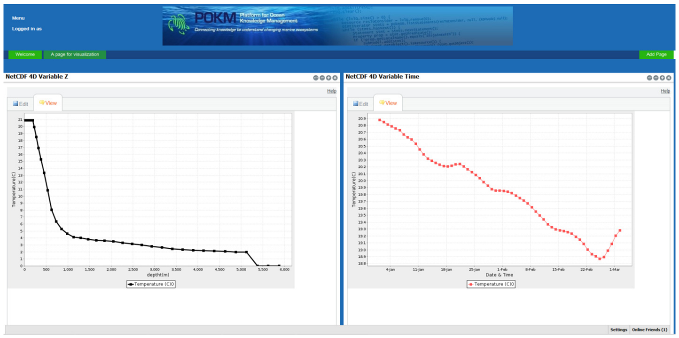


Fig. 9: A screen shot of the POKM system showing two line plots, side by side, generated to observe the behavior of same oceanographic layer over different dimensions. In this example a 4D oceanographic variable, namely Temperature, is observed by fixing the spatial coordinates. In the line plot on the left the variable is observed over different depth levels by fixing the time stamp, whereas in the plot on right the variable is observed over a time interval by fixing the depth level.

IV. CONCLUDING REMARKS

We have presented an e-Science platform—POKM—that showcases a unique synergy of semantic web, services oriented architectures, web services and visualization technologies. 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. We highlighted the technical infrastructure of POKM which is modeled along a services-oriented architecture that exposes a range of task-specific web services accessible through a web-based portal. To enables 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.

In future, we will be applying POKM to study a variety of eco-system phenomenon, such as the migration patterns of marine animals (especially leatherback turtles). In addition, we are looking at using POKM to study the effects of coastal flooding.

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 all the members of the POKM team, especially the marine biology, ocean data modeling and document management teams.

REFERENCES

[1] T. Gruber, “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”, in *International Journal of Human-Computer Studies*, vol. 43, 1995, pp. 5-6.

[2] L. Cabral et. al., “Approaches to Semantic Web Services: an Overview and Comparisons”, in *Lecture Notes in Computer Science*, vol. 3053/2004, 2004, pp. 225-239.

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

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

[5] Business Process Execution Language for Web Services, viewed May 13, 2011, <http://public.dhe.ibm.com/software/dw/specs/ws-bpel/ws-bpel.pdf>

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

[7] K. Gottschalk et. al., “Introduction to Web Services architecture”, in *IBM Systems Journal*, vol. 41 (2), 2002, pp. 170-177