The Digital Ocean

N. M. Patrikalakis[1] nmp@mit.edu
S. L. Abrams[2] stephen_abrams@harvard.edu
J. G. Bellingham[1] belling@mit.edu
W. Cho[1] wjcho@mit.edu
K. P. Mihanetzis[1] kmihanet@alum.mit.edu
A. R. Robinson[2] robinson@pacific.harvard.edu
H. Schmidt[1] henrik@keel.mit.edu
P. C. H. Wariyapola[1] gitano@mit.edu

[1]Massachusetts Institute of Technology
Cambridge, MA 02139, USA

[2] Harvard University
Cambridge, MA 02138, USA

Abstract

The ocean is fundamentally important to many areas of modern society and thus improved knowledge of the ocean is essential. Ocean scientists have made remarkable progress in observation technology, modeling and assimilation in physical oceanography, acoustics, and biology. To some extent, such advances have been confined to each discipline. Therefore a great demand has arisen for a modern distributed computing and networking infrastructure within which we bring together advanced modeling, observation tools and field estimation methods. This paper describes a knowledge network of distributed heterogeneous data and software resources for multidisciplinary ocean research.

1 Introduction

The ocean is central to many areas of modern society. Our food supply, transportation, national security, and recreational activities depend critically upon the state of the ocean. Improved knowledge of the ocean is thus essential in order to forecast possible environmental disasters, control the impact of human activity in climate changes, or respond to sea-borne aggressions. Over the last decades, ocean scientists have made remarkable progress both in the ability to observe the ocean and in the methodology of field estimation. Sophisticated techniques have been developed by physical oceanographers in order to dynamically combine computations and data through assimilation, and to model the interaction of phenomena on different scales through nested computational domains. Sampling strategies and biological modeling are also undergoing constant improvement through adequate and adaptive coupling with the above models. To some extent, the advances that have been made in observation technology, modeling and assimilation in physical oceanography, acoustics, biology, have been *confined to each discipline.* Therefore a great demand has arisen for a *modern distributed computing and networking infrastructure* within which we bring together advanced modeling, observation tools and field estimation methods in the above disciplines. We are currently developing a knowledge network (named POSEIDON: `http://czms.mit.edu/poseidon/`) of distributed heterogeneous data and software resources, which will allow seamless search, exchange, analysis, and visualization of resources, and enable widely distributed computing such as efficient forecasting and adaptive sampling of the ocean. The distributed computing infrastructure will remove many obstacles to multidisciplinary ocean research, improve the timely distribution and dissemination of results, in turn increasing scientific productivity and benefit both the scientific community and society.

This paper briefly describes our recent work on the POSEIDON system development, which includes: development of a Littoral Ocean Observing and Predicting System (LOOPS); construction of a network-based architecture; executive system for dependency graph-based workflow management; and ontology and metadata creation for distributed geophysical resources.

2 Review of the State-of-the-Art

The wide availability of high-performance networks, tools and interfaces for access, has made possible the construction of distributed information systems. A system of this type uses individual data and software servers, located on diverse computer hosts, as the building-blocks for a unified application. A number of projects are underway to research various aspects of

0-7695-0643-7/00 $10.00 © 2000 IEEE

network communications for distributed systems, producing experimental systems focusing on high performance parallelism (Legion) [1], client/server architectures for service libraries (Ninf) [2], and vertically integrated virtual metacomputers (Globus) [3]. Much recent research has been undertaken towards the development of software-based *middleware,* which manages network communications efficiently between distributed clients and servers. One example, which we use for the POSEIDON system, is the Common Object Request Broker Architecture (CORBA) [4]. CORBA is centered on the Object Request Broker (ORB), which manages all communications between CORBA-compliant objects, regardless of implementation language, operating system, or hardware platform, using Interface Definition Language (IDL), and derived services.

Middleware products merely provide communication services across networks, making no assumptions concerning the problem domain of the clients and servers using them. Integrated domain-specific information systems have been developed in the areas of numerical processing, such as NetSolve, which provides transparent access to a pool of distributed computational servers [5]. To specifically manage meteorological and other geophysical data, several Internet-based search and retrieval systems have been developed. Examples are a central clearing-house for NASA real-time satellite data (EOSDIS) [6]; a system for the dissemination of meteorological data, data standards, and processing software (Unidata) [7]; the Distributed Oceanographic Data System (DODS) for dissemination of oceanographic data with appropriate translators [8]; collaborative study of geo-scientific data (OASIS) [9]; providing a Web-based catalog of environmental datasets (MEL) [10]; and DIENST-based [11] data integration and visualization for coastal zone management (Thetis) [12].

3 Littoral Ocean Observing and Predicting System (LOOPS)

With the advent of new sensors, storage technologies, and especially widespread access to the Internet, the potential exists for a new era of ocean science investigation where scientists, students, and government officials have frictionless access to oceanographic data, simulation results, and software. Realistic field estimates – including real-time nowcasts and forecasts as well as simulations – are now feasible due to the advent of Ocean Observing and Prediction Systems (OOPS), in which a set of coupled interdisciplinary models are linked to an observational network, consisting of various sensors mounted on a variety of platforms, via data assimilation schemes – see Figure 1.

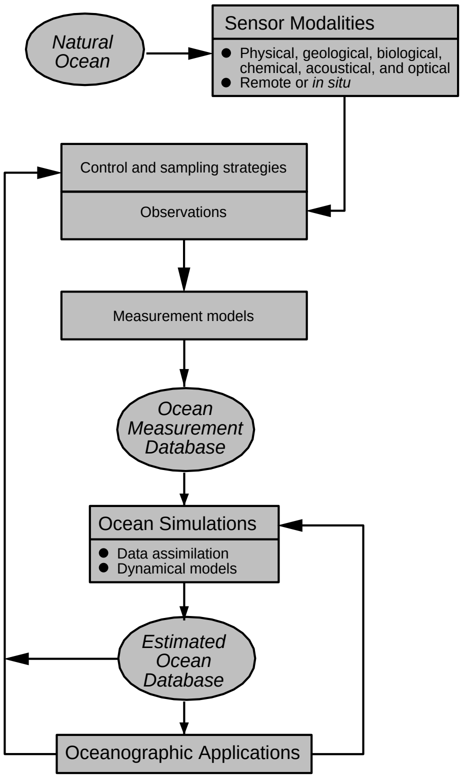


Figure 1: High-level OOPS architecture

A major project in this area is LOOPS: Littoral Ocean Observing and Prediction System [13]. Recent real-time experiences of HOPS (Harvard Ocean Prediction System) in the LOOPS illustrate the diversity of situations and applications of the system. For example, the distribution of real-time products via the WWW is shown in Figures 3-(a),(b). Figure 3-(a): This is from the operational Web page of Rapid Response 97, a NATO real-time exercise. The figure shows a forecast of surface temperature for 18 September 1997 with overlying velocity vectors for the Ionian Basin of the Mediterranean Sea. The same quantities are shown for Massachusetts Bay in Figure 3-(b) for the LOOPS/AFMIS (Advanced Fisheries Management Information System) Massachusetts Bay Sea Trial of August–October 1998 (MBST-98). This figure was available in real-time from the MBST-98 Web site. Figure 4-(a): A whale was found dead in Cape Cod Bay on 21 April 1999 at the position noted by the star. The whale was known to be alive off Race Point on 15 April. This figure was created to present to the National Marine Fisheries Service Criminal Investigation Division the possible locations of ship strikes of a

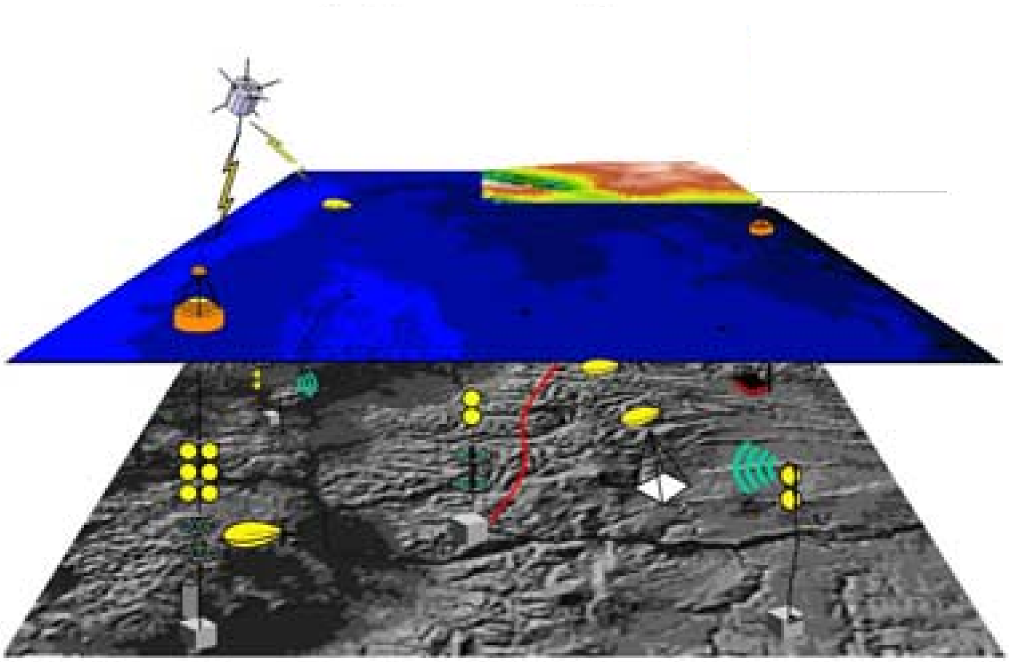
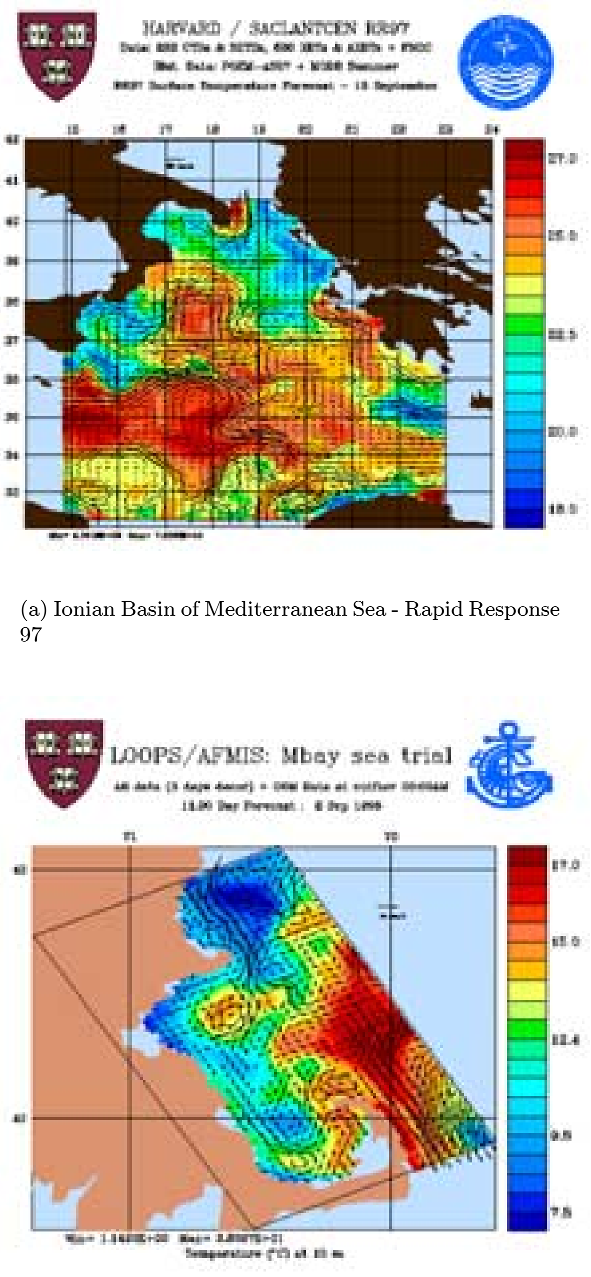


Figure 2: AOSN: Autonomous Ocean Sampling Networks

whale given an unknown time of impact. The ellipses mark the possible locations for a ship strike at an individual date and time which would lead to transport of the whale carcass by the hindcast currents to within 2.5 km of the location at which the whale was found. The gray lines indicate the shipping lanes to Boston and the Atlantic Ocean from the Cape Cod Canal (not shown). Figure 4-(b): The predicted drift of floating debris from the crash of Egypt Air Flight 990 is illustrated here. Given an estimated position of impact, the path of floating debris within an initial locus of 10 km was predicted forward in time. This information was made available to the National Transportation Safety Board (NTSB) during the search period.

The Massachusetts Bay Sea Trial (MBST-98, see also Figure 3-(b)) utilized interactive multi-scale adaptive sampling for three research vessels and two fleets of Autonomous Underwater Vehicles (AUVs) and was performed in collaboration with AFMIS-NASA and the Autonomous Ocean Sampling Network (AOSN-ONR) programs. The objective of AOSN [14] is to create and demonstrate the next-generation robotic oceanographic survey system - see also Figure 2. This is being accomplished by: (1) Creating small, high performance mobile platforms capable of several month deployments. Both propeller-driven, fast survey vehicles, and buoyancy-driven glider vehicles have been developed. (2) Creating an infrastructure that supports controlling, recovering data from, and managing the energy of, remotely deployed mobile platforms. Elements include moorings, docking stations, acoustic communications, two-way satellite communications, and the Internet. (3) Demonstrate these capabilities in science-driven field experiments. (4) Develop operational techniques that make most effective use of these new assets, including adaptive sampling strategies. Simultaneous synoptic physical and biological data sets were obtained over a range of scales. The multi-scale sampling strategies were based on: (1) ocean field forecasts with the HOPS assimilating all prior data (regions of most active or interesting dynamics) and (2) forecasts of error variances and of dominant eigendecompositions of error covariances, using Error Subspace Statistical Estimation (ESSE) [15]. All of the sampling patterns of the platforms and sensors were designed and made available in real-time, assimilating yesterday's data today for tomorrow's forecast and sampling. This data was assimilated using standard

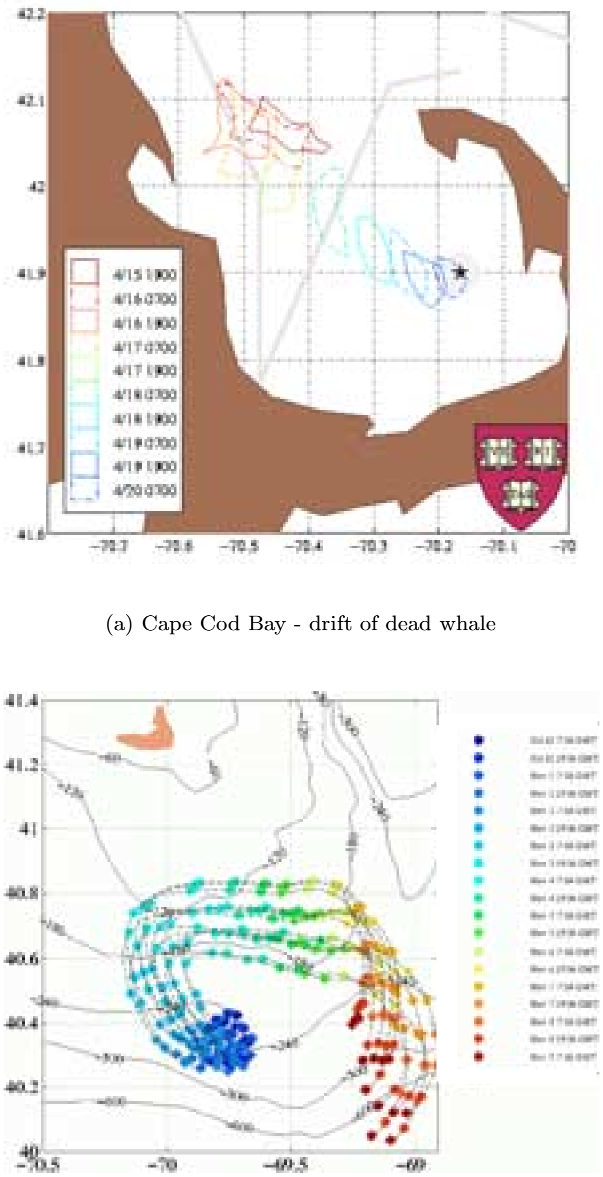
0-7695-0643-7/00 $10.00 © 2000 IEEE



(a) Ionian Basin of Mediterranean Sea - Rapid Response 97

(b) Massachusetts Bay - LOOPS/AFMIS Mass. Bay Sea Trial 1998

Figure 3: HOPS predictions



(a) Cape Cod Bay - drift of dead whale

(b) New England Bight - path of floating debris

Figure 4: HOPS predictions

optimal interpolation (OI) and ESSE. Real time forecasts of fields and error covariance eigendecompositions were provided. Several dynamical interactions among the circulation, productivity and ecosystem systems were found. These accomplishments have resulted in a combined and compatible physical, biological, and chemical multi-scale data set applicable to interactive process studies and data assimilation, adaptive sampling, and predictive skill Observational System Simulation Experiments (OSSEs).

4 Poseidon Architecture

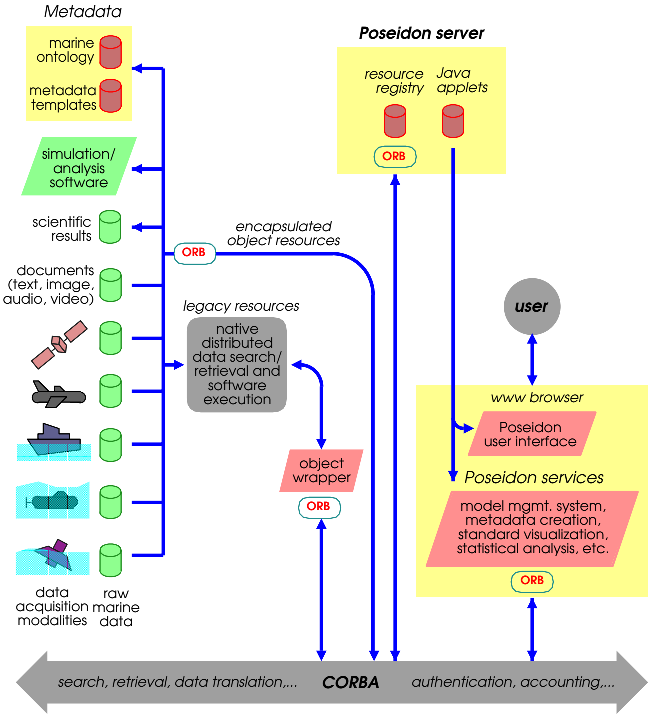

Metadata
marine ontology
metadata templates
simulation/ analysis software
scientific results
documents (text, image, audio, video)
data acquisition modalities
raw marine data
ORB
encapsulated object resources
legacy resources
native distributed data search/ retrieval and software execution
object wrapper
ORB
Poseidon server
resource registry
Java applets
ORB
user
www browser
Poseidon user interface
Poseidon services
model mgmt. system, metadata creation, standard visualization, statistical analysis, etc.
ORB
search, retrieval, data translation,...
CORBA
authentication, accounting,...

Figure 5: Poseidon system architecture

The software infrastructure for LOOPS will be provided by the Poseidon system, which is conceived as a network of distributed data and software resources, communicating with one another across a Common Object Request Broker Architecture (CORBA: `http://www.org.omg/`) backplane, see Figure 5. The standard user's environment will be provided by means of a standard Web browser, to which the Poseidon User Interface and Model Management System will be automatically downloaded as Java^{TM} (`http://www.javasoft.com/`) applets from a Poseidon server. Initially, a single centralized server is envisioned. As the system utilization grows, however, geographically dispersed multiple instances of the server may be established, with the necessary provisions for the concomitant problems of data consistency. The Poseidon server also maintains a resource registry, which identifies the resources that are available for metadata searches and for data access. Resources include marine ontology, metadata templates, measured ocean data (and in fact the platforms or robots collecting such data), documents (text, image, audio, video), scientific simulation/analysis software and simulation data (scientific results). Newly developed resources can be constructed directly as CORBA-compliant objects. Legacy resources, on the other hand, will require an object wrapper, which is composed of a CORBA front-end and a back-end that supports the resource-specific communication protocol.

5 Model Management System

The heart of the Poseidon architecture is the Model Management System (MMS) [16], which is responsible for: (1) creation of a Graphical User Interface (GUI) that allows a user to build an information workflow; (2) validation of the workflow; and (3) management of the execution of the workflow. In the early stages of the Poseidon system development, a number of tests were conducted to test the linking of remote/heterogeneous applications. In order to produce a full-scale example of the application linking, two different applications were wrapped with CORBA layers to become Poseidon objects: (1) an application from Harvard University's HOPS for physical prediction of the ocean; (2) an application from MIT, Department of Ocean Engineering for acoustic tomography. These applications were invoked from a remote applet (Figure 6) that supplied them with a set of initialization parameters and asked for a two dimensional, color contour representation of the acoustic field intensity in a vertical cross-section of the ocean in Massachusetts Bay. In this way, linking of heterogeneous codes written in different programming languages, executed on remote platforms running under different operating systems, and invoked by remote users, was demonstrated.

A Poseidon workflow is constructed as a dataflow dependency graph. An example of a dependency graph is shown in Figure 7 for the case of the Haro Strait experiment sound speed and current inversion [17]. The solid line corresponds to the actual inversion that was implemented. Had the addition of a local circulation model been possible (dashed line), a fully multidisciplinary inversion combining the large sampling coverage of the acoustic data set with the physics of a circulation model would have been achieved.

Poseidon architecture involves a paradigm shift in

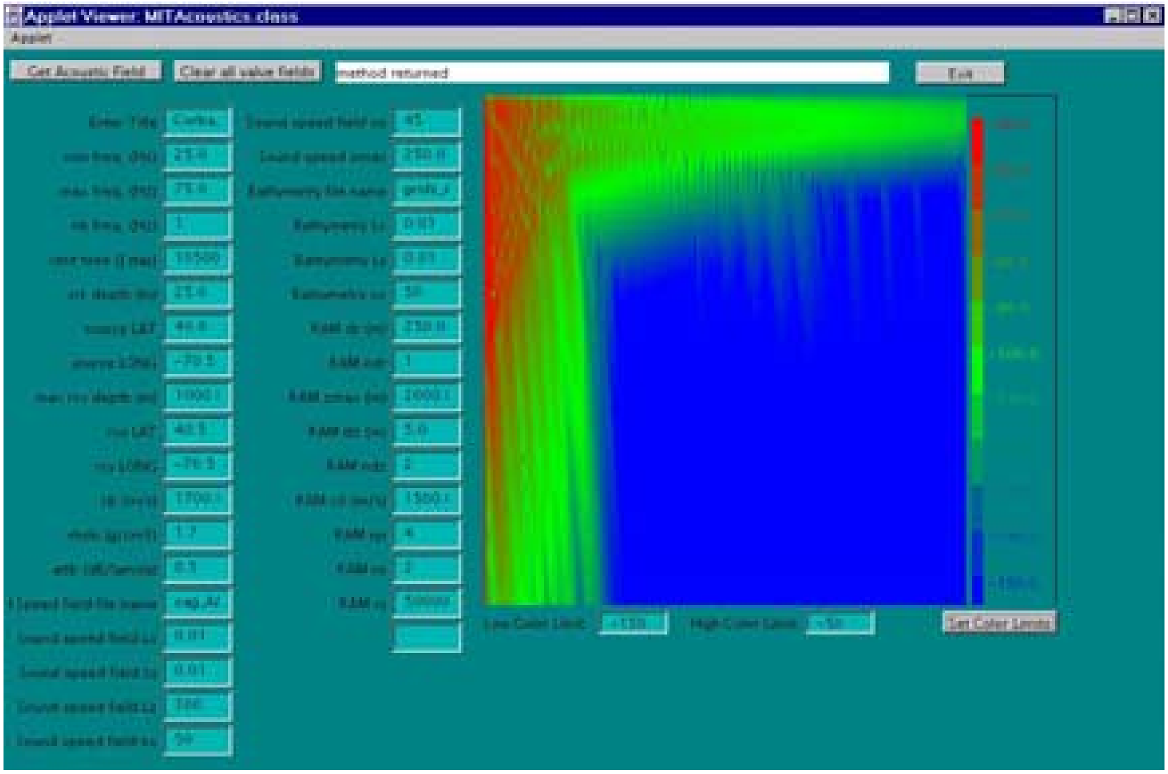


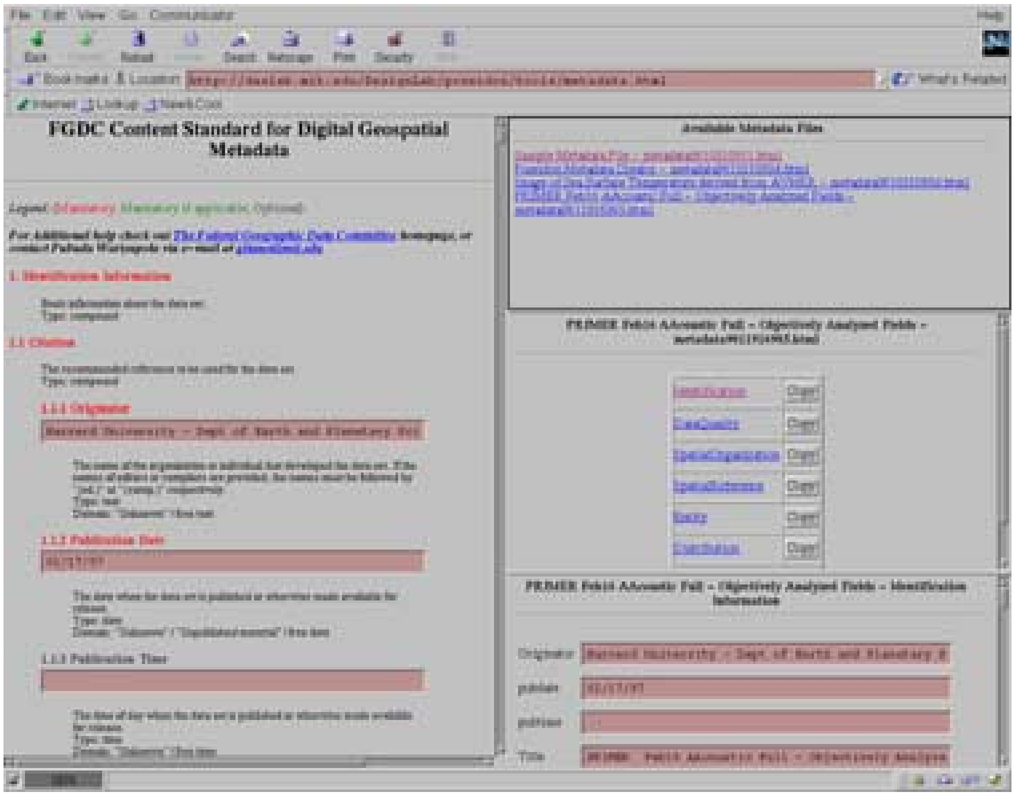
Figure 6: POSEIDON's applet with received acoustic field

large ocean simulations. Currently, an expert scientist downloads legacy software and data to his/her own computer and largely *manually* navigates through the workflow of program executions, based on the expert knowledge and assistance from the authors of legacy software and owners of measured data. POSEIDON, on the other hand, will allow for efficient and largely automated workflow creation for certain problem domains based on the concept of metadata for scientific software. While the example workflow we have created in Figure 6 is completely static (i.e., the sequence of operations and the data exchange mechanisms are hard-coded, and is transparent to the user who sees only one interface to a black box), we have experimented with dynamic, near run-time, workflow creation, to allow the user to combine a series of software and data objects into a workflow and execute it through a GUI. The components of the workflow are represented graphically as objects with input and output parameters that can be combined with other objects on the canvas and executed as a combined workflow. The user sees the graphical objects, but not the transparent back-end execution of the workflow. Ultimately, we plan to develop a system that will be able to automatically create workflows to generate user-specified output parameters from combinations of existing data and software objects, using metadata, and workflow graphs as well as heuristic and AI methods.

In addition, POSEIDON will create a *service-oriented prototype* for ocean simulations without the current downloading of software to a single environment. The World Wide Web (`http://www.w3.org/`) and CORBA provide the basic underlying technologies for service-style oceanographic simulations from hosts able to run simulation resources and models. This service-oriented architecture could have provisions for network security, user authentication, and accounting.

6 Ontology and Metadata Creation

The POSEIDON project involves the development of sharable ontologies (or common vocabulary) and metadata (or data about data) for actual measured/simulated *data* [18, 19] and for modeling *software*. The use of metadata to describe the properties and characteristics of data is beneficial for the construction of automated scientific information retrieval systems. However, current metadata systems would require duplication of large portions of the metadata, which makes the construction of detailed metadata


File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location What's Related
Internet Lookup New&Cool

http://[illegible]/poseidon/tools/metadata.html

FGDC Content Standard for Digital Geospatial Metadata

Legend: Mandatory Mandatory if applicable Optional

For Additional help check out The Federal Geographic Data Committee homepage, or contact Pubudu Wariyapola via e-mail at [illegible]

1. Identification Information

Basic information about the data set.
Type: compound

1.1 Citation

The recommended reference to be used for the data set.
Type: compound

1.1.1 Originator

Harvard University - Dept of Earth and Planetary Sci

The name of the organization or individual that developed the data set. If the names of editors or compilers are provided, the names must be followed by "(ed.)" or "(comp.)" respectively.
Type: text
Domain: "Unknown" / free text

1.1.2 Publication Date

02/17/97

The date when the data set is published or otherwise made available for release.
Type: date
Domain: "Unknown" / "Unpublished material" / free date

1.1.3 Publication Time

The time of day when the data set is published or otherwise made available for release.
Type: time
Domain: "Unknown" / free time

Available Metadata Files

[illegible]

PRIMER Febit AAcoustic Fall - Objectively Analyzed Fields - metadata[illegible].html

Identification	Open
DataQuality	Open
SpatialOrganization	Open
SpatialReference	Open
Entity	Open
Distribution	Open

PRIMER Febit AAcoustic Fall - Objectively Analyzed Fields - Identification Information

Originator Harvard University - Dept of Earth and Planetary S

pubdate 02/17/97

pubtime

Title PRIMER: Febit AAcoustic Fall - Objectively Analyze

Figure 8: POSEIDON Metadata Creator user interface

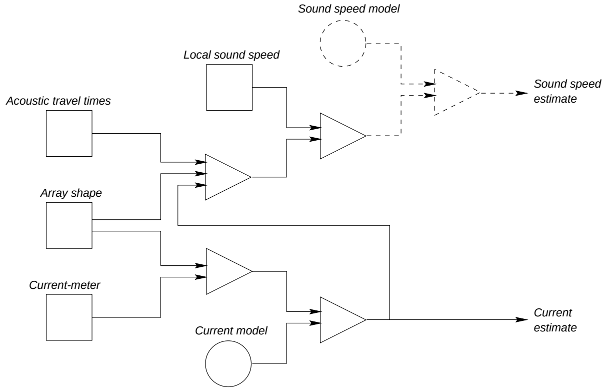
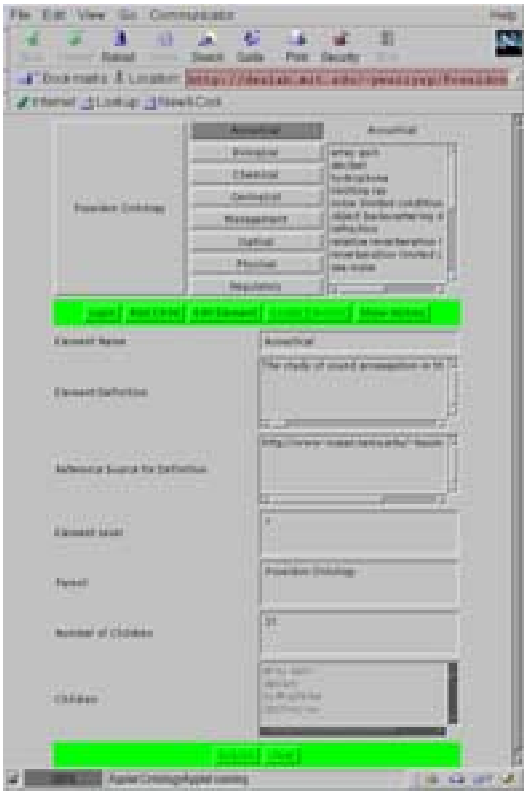


Figure 7: Dataflow dependency graph corresponding to the Haro Strait oceanographic experiment (1996) inversion [17]

cumbersome and time consuming. Our solution to this problem can come from borrowing concepts from object-oriented programming systems, which provide hierarchical taxonomic structure for scientific data and software.

We have implemented a method that incorporates two existing metadata standards (Dublin Core for documents, FGDC standard for geospatial metadata) in an expandable object-oriented structure known as the Warwick Framework. Furthermore, since metadata is expensive and tedious to produce, a Web-based software tool has been developed that simplifies the process and reduces the storage of redundant information - see Figure 8. While we can search the metadata for the information we need, we still need an ontology to ensure that we can identify data unambiguously. Since no existing vocabulary encapsulates all aspects of the ocean sciences and ocean systems management, we have facilitated the production of such a resource by creating a Web-based tool (Figure 9) that will allow specialists with the requisite domain knowledge to populate the ontology independently. The tool handles all the logistical issues of storage, maintenance, and distribution.

To date, we have focused our efforts on metadata for data. Construction of complex oceanographic work-



Poseidon Ontology

Element Name: Acoustical

Element Definition: The study of sound propagation in th

Reference Source for Definition

Element level: 1

Parent: Poseidon Ontology

Number of Children

Children

Figure 9: Poseidon Ontology Creator user interface

flows will require resource discovery of both data *and software.* Metadata for software is an important research issue [20] that would need to involve formal characterization of the input and output structure, range of modeling validity for input parameters, and parametrization in terms of cost, accuracy and methodology of solution. A promising approach involves semantic networks and rule-based inference engine [21].

7 Conclusions

We have presented ongoing work on a modern distributed computing and networking infrastructure for multidisciplinary ocean research. The advent of Littoral Ocean Observing and Predicting System (LOOPS) enables realistic field estimates, including real-time nowcasts and forecasts as well as simulations, in which a set of coupled interdisciplinary models are linked to an observational network, consisting of various sensors mounted on a variety of platforms, via data assimilation schemes. Poseidon is a knowledge network which will allow seamless search, exchange, analysis, and visualization of resources, and enable widely distributed computing such as efficient forecasting and adaptive sampling of the ocean. The Model Management System (MMS) of the Poseidon architecture is responsible for: (1) creation of a Graphical User Interface (GUI) that allows a user to build an information workflow; (2) validation of the workflow; and (3) management of the execution of the workflow. The Poseidon project also involves the development of sharable ontologies and metadata for actual measured/simulated data and for modeling software.

Acknowledgments: Department of Commerce (NOAA, Sea Grant) under grant NA86RG0074, National Ocean Partnership Program via ONR grant N00014-97-1-1018. The authors also thank Prof. S. Lalis for his comments on this work.

References

[1] A. S. Grimshaw and W. A. Wulf. Legion: The next logical step toward the world-wide virtual computer. *Communications of ACM*, 40(1), January 1997.

[2] M. Sato, H. Nakada, S. Sekiguchi, S. Matsuoka, U. Nagashima, and H. Takagi. Ninf: A network-based information library for global world-wide computing infrastructure. In *HPCN '97*, pages 491–502, 1997.

[3] I. Foster and C. Kesselman. Globus: A metacomputing infrastructure toolkit. In *IPPS/SPDP '98, Heterogeneous Computing Workshop*, 1998.

[4] S. Vinoski. CORBA: Integrating diverse applications within distributed heterogeneous environments. *IEEE Communications Magazine*, 14(2), February 1997.

[5] H. Casanova and J. Dongarra. NetSolve: A network server for solving computational science problems. *Journal of Supercomputer Applications and High Performance Computing*, 111(3):212–223, Fall 1997.

[6] NASA: Earth observing system data and information system. (http://spsosun.gsfc.nasa.gov/New_EOSDIS.html)

[7] R. C. Dengel and J. T. Young. The Unidata recovery system. In *Proceedings, Ninth International Conference on Interactive Information and Processing Systems for Meteorology, Hydrology, and Oceanography*, Anaheim, CA, January 1993.

[8] J. Gallagher and G. Milkowski. Data transport within the distributed oceanographic data

system. In *Fourth International World Wide Web Conference*, Boston, MA, December 1995. (http://www.w3.org/Conferences/WWW4/ Overview.html)

[9] E. Mesrobian, R. Muntz, and E. Shek. OASIS: An EOSDIS science computing facility. In *International Symposium on Optical Science, Engineering, and Instrumentation, Conference on Earth Observing System*, August 1996.

[10] Master Environmental Library (MEL): (http://mel.dmso.mil/)

[11] C. Lagoze and J. Davis. Dienst: an architecture for distributed document libraries. *Communications of the ACM*, 38(4):45, April 1995.

[12] C. Houstis, C. Nikolaou, M. Marazakis, N. M. Patrikalakis, J. Sairamesh, and A. Thomasic. THETIS: Design of a data management and data visualization system for coastal zone management of the Mediterranean sea. *D-lib Magazine*, November 1997. (http://www.dlib.org/)

[13] A. R. Robinson *et al.* Realtime Forecasting of the Multiscale, Interdisciplinary Coastal Ocean with the Littoral Ocean Observing and Predicting System (LOOPS). *Abstract of Lecture at AGU Ocean Sciences meeting, January 2000.* (http://czms.mit.edu/poseidon/publication/)

[14] T. Curtin, J. G. Bellingham, J. Catipovic, and D. Webb. Autonomous Ocean Sampling Networks. *Oceanography*, 6(3):86–94, 1993.

[15] A. R. Robinson, P. F. J. Lermusiaux and N. Q. Sloan, III. Data Assimilation. in *The Sea: The Global Coastal Ocean I*, Processes and Methods (K. H. Brink and A. R. Robinson, editors), Volume 10, John Wiley and Sons, New York, NY, 1998.

[16] K. P. Mihanetzis. *Towards a Distributed Information System for Coastal Zone Management.* MIT, Joint Ocean Engineer and M.S. in Ocean Systems Management Thesis, May 1999. (http://czms.mit.edu/poseidon/publication/)

[17] P. Elisseeff, H. Schmidt, M. Johnson, D. Herold, N. R. Chapman, and M. M. McDonald. Acoustic tomography of a coastal front in Haro Strait, British Columbia. *Journal of the Acoustical Society of America*, 1999, in press.

[18] P. C. H. Wariyapola, N. M. Patrikalakis, S. L. Abrams, P. Elisseeff, A. R. Robinson, H. Schmidt, and K. Streitlien. Ontology and Metadata Creation for the Poseidon Distributed Coastal Zone Management System. *Proceedings of IEEE Forum on Research and Technology Advances in Digital Libraries, IEEE ADL'99,* pp. 180–189. May 1999, Los Alamitos, CA:IEEE, 1999.

[19] P. C. H. Wariyapola. *Towards an Ontology and Metadata Structure for a Distributed Information System for Coastal Zone Management.* MIT, M.S. in Ocean Engineering Thesis. September 1999. (http://czms.mit.edu/poseidon/publication/)

[20] N. M. Patrikalakis, P. J. Fortier, Y. Ioannidis, C. N. Nikolaou, A. R. Robinson, J. R. Rossignac, A. Vinacua, and S. L. Abrams. Distributed Information, Computation, and Process Management for Scientific and Engineering Environments. *D-lib Magazine*, April 1999. (http://www.dlib.org/)

[21] V. Christophides, C. Houstis, S. Lalis, and H. Tsalapata. Ontology-Driven Integration of Scientific Repositories. *NGITS '99, New Generation Information Technologies, Lecture Notes in Computer Science,* Elsevier, Hobart, Israel, July 1999.