AN INTERDISCIPLINARY OCEANIC DATABASE FOR THE EAST ASIAN SEAS

Jiannheigh Wu, Shouyu Chen , Chu-I Lee, Chih-Hsin Chen, Jianheng Wu, Jiayu Kuo, Huilan Hsu, Sen Jan
Institute of Oceanography, National Taiwan University, No.1, Sec. 4, Roosevelt Road, Taipei 106 , Taiwan,
Email: senjan@ntu.edu.tw

ABSTRACT

Founded by Taiwan's National Science Council (NSC) in 1987, the Ocean Data Bank (ODB) is being operated by the Institute of Oceanography, National Taiwan University (IONTU). This is an interdisciplinary oceanic information system based on a Service Oriented Architecture (SOA) for the Western Pacific Ocean. The database has compiled various oceanic data acquired in the surrounding seas of Taiwan since 1986. The service let users search multidisciplinary datasets via web browsers. The user-friendly interface helps users to build up a customized query easily. The built-in data quality filters also help users to export testable data in multiple formats (e.g. Extensible Markup Language, XML, or Comma Separated Values, CSV). This database also integrated with a Geographic Information System (GIS), and generates instant graphs and maps (with 15' grid resolution and monthly timescale) on physical, geophysical, or biological oceanographic information. Special request for customizing graphs, maps, or historical archive data is allowed through a built-in online requesting system, which then goes through a back-end standardized workflow. This web-based SOA also provides various levels of authorization, which is convenient in data management for administrators, managers, technicians as well as scientists and the public. Comparing to the mean data access among all those previous isolating systems over the past 20 years, we anticipate a significant growth of data access within one to two years after this service formally launched online in 2010.

1. INTRODUCTION

In the past 20 years, IT system of ODB has been designed in an isolated structure to focus on silo functions, which also resulted in culture barriers and high maintenance cost that need to be addressed. We have many isolated systems for different goal. They are all based on huge data source of collection from marine cruises of platform Ocean Researcher I,II,III. Before building different goal of system, we need to extract, clean and load the data of physical dataset or marine geophysical dataset or bathymetric dataset or biotech dataset for each system. Most of the isolated systems are not interoperability. Among the heterogeneous systems, a lot of problems are about our research of ocean.

On the software side, there were two choices: port our old “legacy” code to the new host with modernizations or climb on the Commercial Off The Shelf (COTS) software bandwagon. The arguments for porting the old code were familiarity, both on our and our customer’s part, and rational use of our huge investment. Arguments for using COTS programs are legion: Lower cost, better maintainability, more flexibility, the facility to blame someone else for problems,..and the fact that the concept is very much in style. However, we could not find anyone who had used these tools to build large, complex, user-friendly systems.

Before we go on, we need to clarify the subject: We are not talking about turnkey software programs that perform specific tasks. The software under discussion falls between the turnkey systems and the “conventional” software-development systems like FORTRAN and C. They are specialized for specific tasks, such as data acquisition and time-series analysis, and are meant for the development of custom applications. After a few abortive attempts at porting, the COTS option was selected(Aras Innovator). It has taken an innovative approach combining proven technologies, common Microsoft platforms, and open standards to create the enterprise application framework. The license of Aras is free which is saving our budget.

The superiority of SOA(service-oriented architecture) technology is analyzed in details, and a SOA-based integrated model for heterogeneous systems is proposed. This framework is used in an application project, it solves user information integrate between physical system and marine geophysical system and all of the marine systems, shows its advantage in solving related problems. This model is applied to improve the interoperability of heterogeneous systems. Unfortunately, conventional system implement have paid a high price for these achievements, most often in the form of multi-million dollar due to overly rigid systems that are difficult to implement and challenging to adapt to changing functional requirements. Complex coding and schema changes are required which are time consuming, resource intensive, disruptive and expensive. We need flexible enterprise applications that can be quickly tailored to support complex data and easily changed to adapt to feature’s dynamic business environment. In order to meet this challenge a new level of marine application flexibility is required. In addition to the strategic benefits greater software flexibility provides ODB realizes a dramatic reduction in the amount of time and IT resources necessary to deploy and maintain applications translating into significant ownership cost savings. We found an easy way for building a SOA-based application of marine information management.

2. METHOD

SOA can be both an architecture and a programming model, a way of thinking about building software. An SOA enables you to design software systems that provide services to other applications through published and discoverable interfaces, and where the services can be invoked over a network. When you implement an SOA using Web services technologies, you create a new way of building applications within a more powerful, flexible programming model.

Aras Innovator is a Web-based, n-tier, service-oriented architecture composed of web clients, application server(s), database(s) and file server(s) and is based entirely on standard Internet protocols including HTTP / HTTPS, XML and SOAP [Simple Object Access Protocol].

The SOA web services are implemented in the Microsoft .NET CLR languages optimized for flexibility, performance, and security. By utilizing the Microsoft technologies Aras Innovator takes advantage of the common Microsoft platforms such as the Windows Server System, Microsoft SQL Server, and the .NET Framework which make the costs of the associated infrastructure significantly less than J2EE-based systems

while providing excellent performance, scalability, manageability, and security.

3. RESULTS

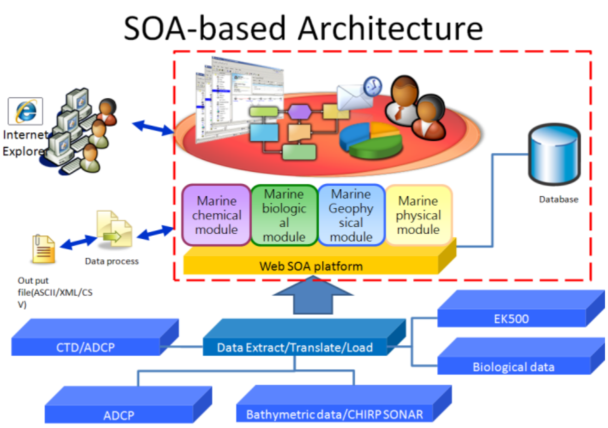

Fig. 1 SOA-based Architecture of ODB

Oceanographic database

The oceanographic database is under the Ocean Data Bank (ODB), which mainly archive and disseminate the Ocean Research Vessels (R/V OR) 1, 2 and 3 data. Currently, there are mainly two different types of data available in the ODB for Ocean physics: Shipboard Acoustic Doppler Current Profiler (Sb-ADCP) data and Conductivity-Temperature-Depth (CTD) data. We use SQL to access and manipulate data in MS SQL Server. In order to do the data transference, recently we have been redesigned in the new data format. To take advantage of the new data format, which can be reconfirmed and deleted the error data in the past. To further reduce the statistical deviation. Until now, it has been cumulative over 10 millions of CTD data in the past 25 years and over forty thousand stations. More than 80 millions ocean current data have been saved and over 3 millions casts (2-min average) and over two decade years. By doing this, we can provide CTD data and Sb-ADCP data for related investigations application in the East Asian Seas Region. Besides, we can also provide the interdisciplinary description on marine phenomena and observations. Not only submit the Data Request Form, but also create value-added services on an interdisciplinary oceanic information system continually. It has been renewing analysis on the spatial and temporal statistics in the past several years. The conclusion of statistics result is different between space and time. On time scale, the annual and seasonal number of observations and distribution of the hydrographic data was using the CTD stations or casts. The annual and seasonal number of observations and distribution of the current data was calculated using the Sb-ADCP in 2-min average data. The hydrographic and current data was based on a one-half degree (1/2°) grid and calculated a new one-quarter degree (1/4°) grid, which greatly improved the spatial resolution of these climatological fields. The annual and seasonal fields are calculated at standard levels from the surface to 5500 meters. To calculate the average of oceanographic data on a 1/4° grid, which can be shown the annual climatological fields and the standard deviation can be shown the different areas. Further, the seasonal climatological fields can show the spatial variations from season to season. These results will be upgraded in the oceanic information system, which can be using the oceanographic data to improve the application of ocean platform.

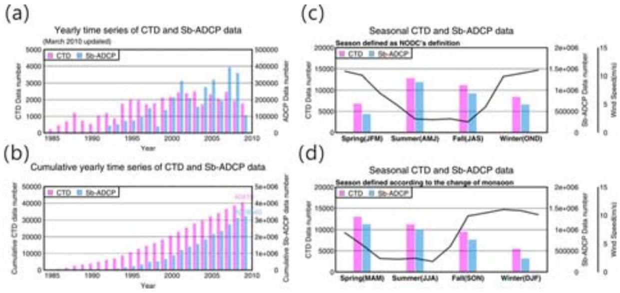

Fig. 2 Time series of the number of CTD and Sb-ADCP casts in the oceanographic database

Fig.2 shows the annual and seasonal statistics of CTD and Sb-ADCP casts. Left panel show yearly time series of CTD and Sb-ADCP data(Fig. 2(a)) and cumulative yearly the series of CTD and Sb-ADCP data(Fig. 2(b)).Right panel show the seasonal statistics: one (Fig. 2(c)) is based on seasons following the Northern Hemisphere conversion .The seasons are defined as: spring (January - March), summer (April - June), autumn (July - September) and winter (October - December). Another (Fig. 2(d)) is divided by the changing of the East Asia monsoon: which means spring is from March to May, summer is from June to August, autumn is from September to November and winter is from December to February. The bold line is monthly statistics of wind velocity on the island of Dongji in the past few years.

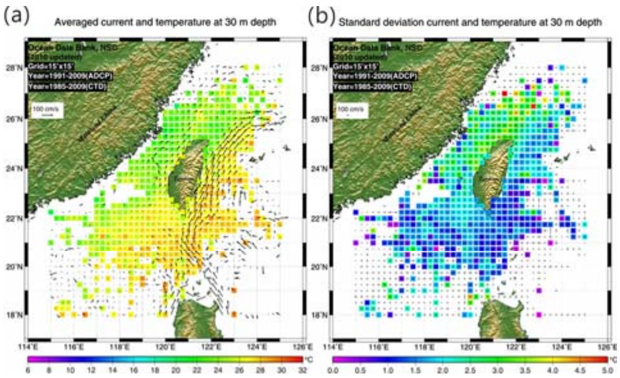

Fig. 3 The average and standard deviation of the annual climatological mean and standard deviation of temperature and current on the 1/4° grid at 30m depth

Fig.3 shows the average and standard deviation of the annual climatological mean and standard deviation of temperature and current on the 1/4° grid at 30m depth. Fig. 3(a) shows the average current and temperature at 30m .Right panel shows the standard deviation current and temperature (Fig. 3(b)). Average of each grid has more than 3 data numbers. The arrow point shows the direction the ocean is flowing to. The corresponding grid means the average temperature (Fig. 3(a)). The cross mark means the double standard deviation of U component and V component. The color on the grid means the standard deviation of temperature (Fig. 3(b)).

Marine Geophysics database

Marine Geophysics database maintains a large volume of bathymetry, seismic navigation data, Sub-Bottom Profiler data and other data types.

Data sources include ocean researcher I、II、III, and collected during marine cruises from 1990 to the present.

Our data sets of these digital databases have been released as a DVD-ROM or VCD-ROM sets , and the Geographic Informatio system (GIS) have been provided with ArcIMS .

Currently only a limited number of lines are available online. Digital seismic reflection data (SEGY files) and Sub-Bottom Profiler data are available on request and can be searched for using WebGIS by geographic area, survey, cruise, or geophysical parameter.

Marine Seismic Reflection Data

1. Equipment： DFS-V、StrataVisor

2. Since 1990, Data Source from Ocean researcher I、II、III

3. 68 cruises、928 lines, Amount 1 TB bytes and more than thousand pieces optic disks (in SEGY files)

4. Established the metadata of seismic dataset.

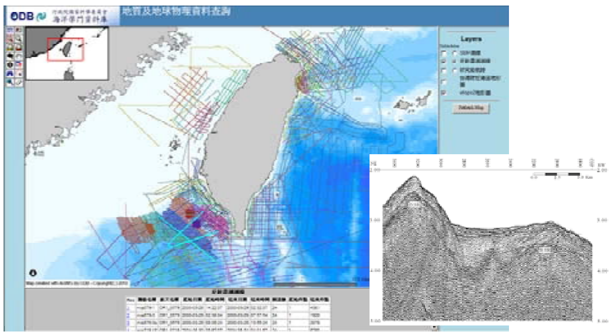

Fig. 4 Marine Seismic Reflection Data

Sub-Bottom Profiler database

1. Equipment
   - ODEC bathy-2000P chirp sonar
   - Edgetech SB-0512i chirp sonar
2. Since 2001, Data Source from Ocean researcher I、II、III
3. 130cruises、4700 lines. Amount 200 GB bytes、more than thousand pieces optic disks.

Established the metadata of seismic dataset.

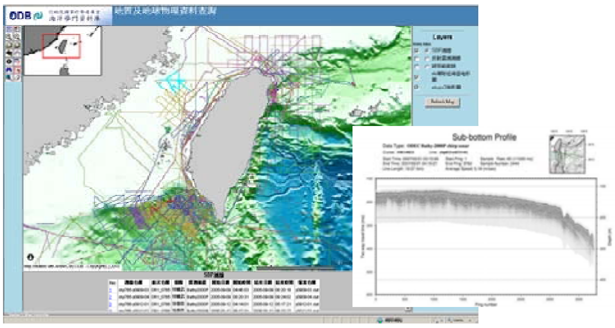

Fig. 5 Sub-Bottom Profiler database

Bathymetric Data around Taiwan

Single-beam echo sounders provide depth information along the ship's trackline. High-density single beam bathymetric data collected by National Oceanographic Research Vessels OR1, OR2, and OR3 since 1989. The amount of data about 1.35 million km, more than 40 million data points. User could inquire through our website.

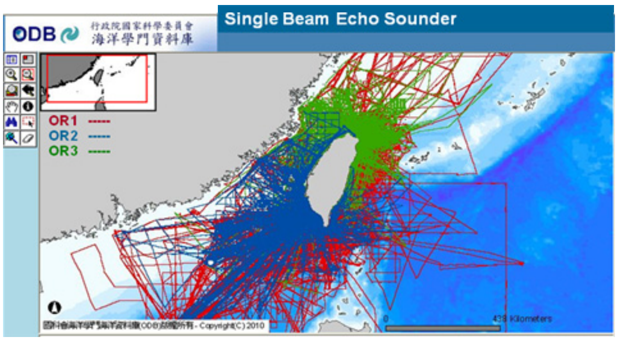

Fig. 6 Shipboard Bathymetric Data Distribution Map

Bathymetric Data around Taiwan

Digital bathymetric data set has been produced by compiling available shipboard bathymetric data supplemented by global bathymetric data sets. It may serve the basis for studying interrelationship among marine geology, physics, chemistry, and biology. We provide two data sets as follows:

500-m gridded data set : Coverage is 18° N to 27° N, and 117° E to 125° E.

1000-m gridded data set : Coverage is 2° N to 35° N, and 105° E to 135° E

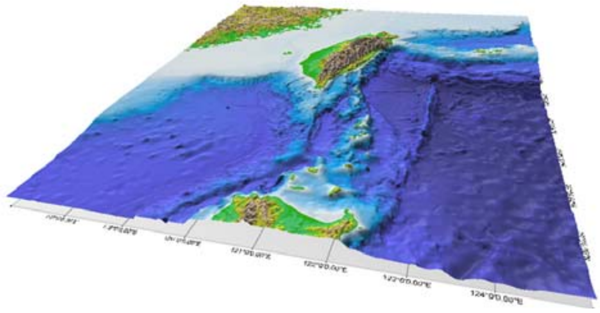

Fig. 7 Taiwan Digital Bathymetry Model, Version6 (TaiDBMv6). Compiled in 1998

EK500 38k/120kHz acoustic image dataset

Scientific echo sounder used to investigate fish distribution, submarine volcanoes and volcanic activity, gas plumes, sedimentary feature, etc. EK500 data Collected by 3 vessels since March 2003, recorded more than 800 cruises.

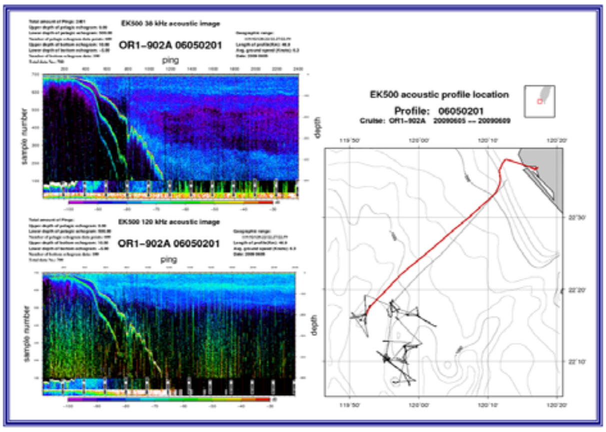

Fig. 8 Example of EK500 data obtained during a gas-hydrates research cruise, 2009.

Bio-database

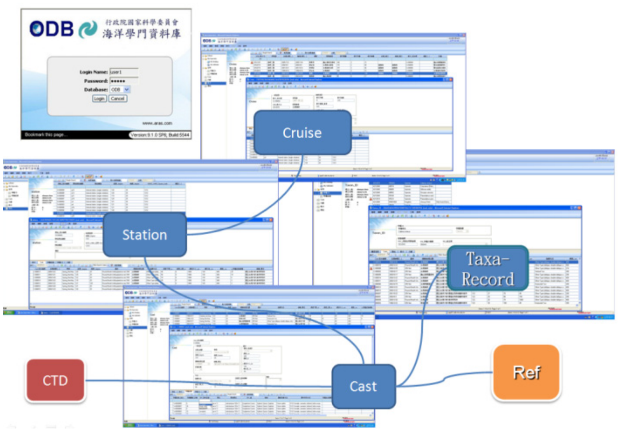

Fig. 9 bio-database

We started to build up a bio-ocean database in late 2009. It is focusing on the zooplankton abundance in the surrounding seas of Taiwan, providing the essential biological parameter to test regional ecological hypotheses. This bio-database is also part of our interdisciplinary ocean integrated database followed with the main structure of World Ocean Database 2009, the world largest ocean database running by National Oceanic and Atmospheric Administration, U.S.A. Users can not only build up a query easily according to their target geographical regions, sampling methods, research ship cruises, stations as well as a particular taxonomic groups or species, but also can cross search the correlated environmental data such as the Conductivity, Temperature, Depth (CTD) records. Our bio-database has been collected over 6000 plankton abundance records over the past six months from historical references. Data after the quality check will be released in the system which is planned launching in fall, 2010.

4. DISCUSSION

The Need for Greater Flexibility

Logical rules need flexible applications that can be quickly tailored to support company-specific competitive processes and easily changed to adapt to today's dynamic business environment. In order to meet this challenge a new level of enterprise application flexibility is required. Improved software adaptability enables companies to easily and cost effectively develop unique applications to support complicated logic processes and practices that deliver:

- Differentiated product and service offerings that excite customers
- Shorter process cycle times that improve responsiveness to customers
- Better profit margins that ultimately increase shareholder value

In addition to the strategic benefits greater software flexibility provides a company realizes a dramatic reduction in the amount of time and IT resources necessary to deploy and maintain applications translating into significant ownership cost savings.

Unique Model-based Approach

The truly unique aspect of the Aras Innovator enterprise application framework is the model-based approach. Aras Innovator takes the premise of defining business processes within enterprise software using graphical models to the next level by running the system directly from the model in real-time. As opposed to "model-driven" systems in which a model is created graphically and traditional source code is generated which is then compiled, linked, tested, debugged, and ultimately deployed, Aras Innovator's model-based approach uses a run-time model to define the business objects. The advantage of the Aras model-based approach over the model-driven approach is considerable. Changes to enterprise applications in Aras Innovator are made in real-time as opposed toweeks or months using conventional compiled code-based systems including model-driven systems. Benefits of the model-based approach include:

- Real-time changes to business rules, forms, workflows, lifecycles, and the data model without complex programming
- Graphical drag & drop solution development with dynamic schema modification for fast and easy application creation and extension
- Separation of business logic and technology simplifies upgrades and enables solution portability
- No proprietary languages to learn

An important point that requires clarification is that Aras Innovator is not running interpreted code. Interpretation of code never occurs, because the models themselves in Aras Innovator are stored in a predefined schema that is utilized by a set of compiled Web services in the SOA. These Web services perform dynamic evaluation of the business object model.

Simplifying Integration

Integration is simplified by open Application Programming Interfaces [APIs] federation services that link deep into the architecture. The APIs send and receive XML / SOAP messages and publish a complete data dictionary providing a standard interface that makes the exchange of data with other systems or integration products, such as Microsoft BizTalk Server, straightforward. In addition, .NET integration capabilities are available for direct integration with ADO, ODBC, flat file, or direct to another system's API.

5. REFERENCES

Model-Based SOA Paper - Aras Corporation(http://www.aras.com/PLM-Software/100063_submitted.ASPX)

Migrating to a service-oriented architecture(ftp://service.boulder.ibm.com/s390/audio/pdfs/G224-7298-00_FinalMigratetoSOA.pdf)

Strether Smith, Lockheed Martin Advanced Technology Center, Palo Alto, CA ,Commercial-Off-The-Shelf (COTS) Software Systems for Data Acquisition and Analysis Love Affairs and Land Mines