Research on Marine Environmental Data Management in China Digital Ocean Prototype System

Jian Liu[1, 2], Xin Zhang[1], Xiaoyi Jiang[3], Bing Jiang[3], Titanhe Chi[1]
1. Institute of Remote Sensing Applications Chinese Academy of Sciences
2. Graduate School of the Chinese Academy of Sciences, Beijing, China
3 National Marine Data & Information Service, Tianjin, China

1. Introduction

Marine environmental data consist of marine environment element data, marine environment phenomenon data, marine biology and chemistry data etc. Marine environmental data provide main basis for decision-making in marine development, environmental protection, disaster prevention and management. Now, database management is commonly used in marine environment data management, which provides convenience and possibility for marine data service and share. However, with the accumulation of data and growing of application demand, database technology hasn't fully played its role in marine basic data management, and the data has not been fully utilized. This was manifested in the following aspects: (1) since the database is running on different hardware and software running platforms, there is inconsistency in the data structure; database system is geared to the needs of project construction and application demand, thereby making unified and sustainable management difficult to achieve; reduplication of some data exists among databases. (2) In data analysis environment, there exist scattered resources, and there is a lack of systematic process management and accumulation for information; and the lack of a comprehensive analysis environment.

With the rapid expansion of marine application demand, people's dependence on the marine environment information becomes increasingly strong. It also puts forward higher requirements for effective management and comprehensive analytic applications of marine environment information. Data warehouse is defined as subject-oriented, integrated, unnewable and time-varying [1]. It is an effective solution for marine environment data storage and integrated application [2, 3].

Based on the characteristics of marine environmental data and application demands, this article studied marine environmental data management based on data warehouse in China Digital Ocean Prototype System (CDOPS). Since the data are stored distributed in the node data center, a mechanism of data registration, publishing and uploading was proposing; through the data discovery, location access services provided by meta-data directory service, distributed data exchange and share of marine environment are realized. The data in the data warehouse are stored at different analytic subject, and ROLAP is used to organize multidimensional data. In fact table, a "hierarchical object model" is used to organize 3D data, and through record valid time, loading time and existing time in time dimension table, the dynamically update data warehouse and identify of data version can be easily realized.

2. Data Information

The source of environmental data demonstrated in CDOPS can be categorized as images derived from remote sensing retrieval, data field derived numerical model, and data products measured from various observation platforms. The elements include seawater temperature, salinity, density, sound velocity, currents, tides, storm surges, dissolved oxygen content, PH value, chlorophyll concentration etc.; the data formats involved include: .tif, .img, .shp, .jpg, .dbf, ASCⅡ files, binary files etc., and the total amount of data is up to

This work is supported by the National High Technology Research and Development Program of China (863 Program) under the grant No. 2009AA12Z208

TB level. Owing to the diversity of marine data sources, formats and storage structures, marine environmental data have the characteristics of multi-source, heterogeneous and multi-state; In addition, the ocean is a dynamic, continuous, boundary-fuzzy time and space information carriers. Thus, multi- temporal scales and temporal-spatial dynamics are the main features making marine environmental data different from other data [4].

3. Solutions of Key Issues

3.1 Mechanism of Data Registration, Publishing and Uploading

The data warehouse can be classified as integrated and distributed data warehouse [1]. The environmental data demonstrated in CDOPS are distributed in the data center of National Marine Data & Information Service, node data centers in costal provinces and Institutes of Marine Research. The marine environmental data warehouse using distributed architecture, in which central data warehouse located in the National Marine Data & Information Service Center, and data markets are established in node data centers. All the data loading to marine environmental data warehouse need to be registered at "Data Registration Center" and the registered information is metadata.

The access of data stored in node data center can be realized by data publishing mechanism: images in the format of .tif or .img derived from remote sensing retrieval can be published through building WMS servers; vector data in .shp format can be published through building WFS servers; data in .jpg, .txt or .dat format can be stored in FTP server and accessed by ftp address; data stored in a database system can be accessed through grant authority and open access interface. Data in the node data centers which need to be uploaded to the central data warehouse and kept in the archives, after the metadata have been registered, should be loaded to the node data exchange directory; then, the automatic updating service automatically retrieve and upload the data through the data bus, and the data uploaded will be automated distribution and loaded to central data warehouse.

3.2 Metadata Catalog Service

Through the data discovery, location access services provided by meta-data directory service, distributed data exchange and share of marine environment are realized. The implementation of meta-data directory service depends on the basic software and hardware, including: metadata editor, data registration center, metadata server and meta-data gateway. The metadata of marine environmental data are organized in xml format. Metadata editor, deployed in each node data center, is used to catalog metadata for various types of product. In addition to data source, format definitions and descriptions, the data publish method, access method and access link are also recorded in metadata. All the data are registered at "Data Registration Center", which is a website providing interfaces for data query and access. The metadata gateway deployed in the National Marine Data & Information Service Center is used to connect data servers at each node data center, and provide proxy service of data query requests.

3.3 Analytic Subjects in Marine Environment Data Warehouse

The analysis subjects in marine environmental data warehouse are established by referring to Object-oriented Spatio-temporal Data Model and Feature-based Spatio-temporal Data Model [5, 6]. The analytic requirements of data and features of data structure are main considerations of division of analysis subjects. For example, although the data structures of typhoon and storm surge are different, and the storm surges are often caused by typhoon, there is a need of comprehensive analysis of typhoon and storm surge. So these two phenomena are regarded as one analytic subject, and data files are stored in one fact table to share

dimension tables. The analytic subjects in marine environment data warehouse include hydrological conventional element subject, marine conventional meteorological element subject, marine phenomena subject, marine disaster subject, marine biology subject and marine chemistry subject.

3.4 Logical Model of Multidimensional data

In this paper, ROLAP (Relational On-line Analysis and Processing) is used to organize multidimensional data. ROLAP is based on relational database, in which data storage and retrieval can be easily realized by SQL sentences [7]. The logic model of each subject is designed based on star structure. The fact table is used to record the measurement of each dimensional intersection. Metadata table and Meta_ID (as foreign key) are recorded in fact table, so that the fact table can be easily related to the metadata table in metadata database. Dimension tables include: the time dimension table, which records the detail level of time, valid time, loading time and existing time; space dimension table, which records the data range and coordinate information; depth dimension table, which records the layer scheme, layer number and value of layer depth; the platform dimension table, which records the observation of data.

In fact table, a "hierarchical object model" is used to organize 3D data filed derived numerical model. Firstly split data by layer on depth dimension, and then store the data in Oracle in BFILE format. The virtue of this method is faster storage, small number of records, faster read and easy for SQL retrieval.

Marine environmental information is characterized with its climatology-scale variation features, seasonal variation rules and weather-scale variation, so the time scales are divided into hour, half-day, day, ten days, month, quarter and year six levels. Time dimension table records the detail level of time, valid time, loading time and existing time, in order to control the data version, and to update data warehouse dynamically. For different data types, the definition of existing time is different. The versions of the data can be got by comparing different loading time.

3.5 Visualization of Data

The analysis operation includes: drilling down, slicing, dicing, and rotation, allowing users to view data from a variety of dimensions and integrated level, and to understand data more deeply. The marine environmental data are stored in the data warehouse in numerical format, so there is a need of data visualization. The general application process of frontal analysis of marine environmental data in CDOPS is: first, the user queries data by certain conditions, and then, select a data analytic operation; the program automatically invokes the data query and visualization interfaces, and the visualization results are displayed at main window; based on the authority level, data download services are provided conditionally. The visualization of the data in CDOPS is shown as follows:

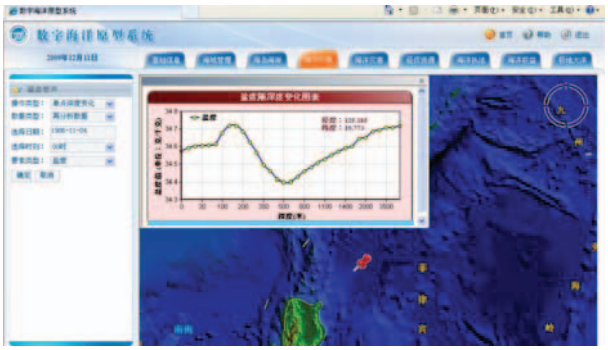

Figure 1 Drilling down of 3D data of seawater salinity at certain position

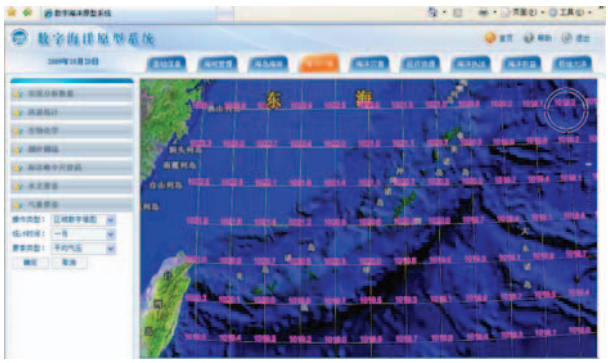

Figure 2 Slicing of pressure data field on depth dimension

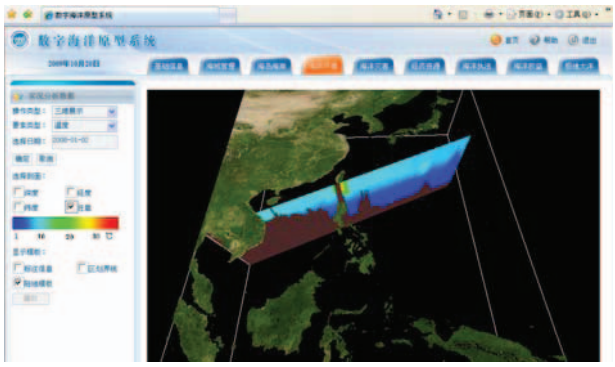

Figure 3 Dicing of 3D data of seawater temperature slice at certain angle on depth dimension

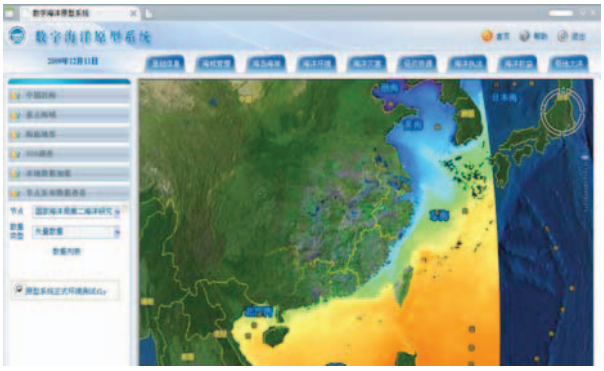

Figure 4 Image of seawater temperature derived and from remote sensing stored in node data center

4. Conclusion

The purpose of marine environment data warehouse construction is to provide comprehensive and multiple marine information services, through integrating marine environment data distributed in different data centers, for policy-makers, marine managers and public users. It also provides data support for Chinese Digital Ocean prototype system. This paper puts forward a proposal for distributed marine environment data warehouse construction, and gives solutions to some of the key issues. Follow-up research may focus on further optimizing the data structure and improving query and retrieval efficiency.

5. References

[1] W.H. Inmon. Building the Data Warehouse, Second Edition.

[2] Hou Wenfeng.Tentative Idea for Development of "Digital Ocean". Marine Science Bulletin, 1999.

[3] Xia Dengwen. Study on Modeling Methods and Related Techniques of Fundamental Data and Business Process in Digital Ocean. Doctoral dissertation, Northeastern University, China, 2006.

[4] Zhang Tianyu. Research on Marine GIS Field Model. Doctoral dissertation, Institute of Geographic and Natural Resources Research, CAS, 2004.

[5] Worboys M F, Hearnshaw H M, Maguire D J. Object-oriented Data Modeling for Spatial Databases. International Journal of Geographical Information Systems, 1994

[6] Wright D.J., Blongewicz M.J., Halpin P.N. and Breman J. Arc Marine: GIS for a Blue Planet, Redlands, CA: ESRI Press. 2007.

[7] Lin Yu. The Principle and Practice of Data Warehouse, Post & Telecom Press, China,2003.

[8] E1eveld MA, Schrimpf WBH., Siegert AG，User requirements and information definition for a virtual coastal and marine data warehouse，Ocean＆Coastal Management [Ocean Coast Manage].2003.