Chapter 2
Ocean Big Data Acquiring and Integration Technologies

Xin Zhang, Lei Wang, Xiaoyi Jiang, and Changming Zhu

Abstract Digital Earth is an integrated approach building scientific infrastructure. The ocean data is a typical big data, which can be seen from the data volume, velocity, variety, and value perspectives. The Digital Earth systems provide a three-dimensional visualization and integration platform for ocean big data, which include ocean management data, in situ observation data, remote sensing observation data and model output data. Based on the analysis on the characteristic of ocean big data, this chapter studies the ocean big data acquire and integration technology that is based on the Digital Earth system. Firstly, the construction of the Digital Earth based three-dimensional ocean big data integration environment is discussed. Then, the ocean management data integration technology is presented which is realized by general database access, web service and ActiveX control. Third, the in situ data stored in database tables as records integration is realized with a three-dimensional model of the corresponding observation apparatus display in the Digital Earth system using a same ID code. In the next two parts, the remote

X. Zhang (✉)
State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China
e-mail: zhangxin@radi.ac.cn

L. Wang
Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China

Hainan Province Key Laboratory of Earth Observation, Sanya Institute of Remote Sensing, Sanya 572029, Hainan Province, China
e-mail: wanglei98@radi.ac.cn

X. Jiang
National Marine Data and Information Service, State Oceanic Administration of China, Tianjin 300171, China
e-mail: andyjiangxy@126.com

C. Zhu
Department of Geography and Environment, Jiangsu Normal University, Xuzhou 221116, Jiangsu, China
e-mail: ablezhu@163.com

sensing data and the model output data integration technologies are discussed in detail. The application in the Digital Ocean Prototype System of China shows that the method can effectively improve the efficiency and visualization effect of the data.

Keywords Digital earth • Digital ocean • Ocean big data integration

2.1   Introduction

Since the concept of Digital Earth was put forward by Al Gore in 1998 (2010), many studies have been carried out from the perspective of the Earth. In 2002, Skyline Inc. released the Skyline TerraSuite software. In 2004, NASA launched WorldWind version 1.1, marking the first Digital Earth platform software with a complete scientific research function, which provided scientists with a simulation and display platform for conducting Digital Earth research. In 2005, Google Inc. introduced Google Earth to the world, which raised the Digital Earth application to a new level. Google Earth supports the visualization of DEM data and realizes the virtual representation of the Earth. Similar software includes Leica Visual Explorer (LVE) form Leica Inc., the Visual Earth system from Microsoft Inc., ArcGlobe from ESRI Inc., the digital earth prototype system from the Chinese Academy of Sciences and so on (Guo Huadong 2009). In terms of Digital Earth's potential fields of application, Guo et al. (2009, 2010) studied the digital earth prototype system DEPS/CAS and, as a result of this research, defined digital earth systems as either scientific (such as those of World Wind of the USA, the Digital Earth Prototype system/Chinese Academy of Sciences (DEPS/CAS) of China, Blue Link and Glass Earth of Australia, and the Earth Simulator (ES) of Japan) or commercial (such as Skyline and Google Earth) and proposed that Digital Earth was a comprehensive platform for the integration of future information resources (Guo et al. 2009, 2010).

– With the development of ocean observation technologies, a substantial amount of oceanic data and model products are produced from the three-dimensional ocean observation system, which is composed of diverse monitoring sources such as satellite, airplane, ship, high frequency ground wave radar, buoys (moored and drifting) and land-based stations (Ocean 2010). How to efficiently and effectively integrate the data has become an urgent problem because these heterogeneous and widely distributed data and model products are usually collected in a project-based fashion. So far, the primary ocean observation data integration technologies are GIS (Geographic Information System) and web applications (Yingqi Tang and Wong 2006). In America, portals such as the Oregon coastal atlas, SCCOOS portal (Chongjie Zhang et al. 2007), provide clearinghouses for common decision-support tools, as well as data, maps and ancillary information. In Australia, there are several portals: the Oceans Portal proposed by the Australian National Oceans Office (NOO), Australian Ocean Boundaries Information Systems (AMBIS), and CSIRO's ocean data directory – Marlin (Strain et al. 2006). Global-level ocean portals are being developed as

well, such as the Oceans Biogeographic Information System (OBIS) – a virtual repository of oceanographic and biogeographic information (Malone 2003). Since 2005, the Global Earth Observation System of Systems (GEOSS) has been implemented to achieve comprehensive, coordinated and sustained observations to improve the monitoring of the state of the Earth, to increase understanding of the Earth's processes and to enhance the ability to predict the behavior of the Earth ocean system (GEO 2005). As the oceanographic component of GEOSS, one objective of the Global Ocean Observing System (GOOS) is to foster the development of data management systems that allow users to exploit multiple data sets from many different sources through "one stop shopping" (Thomas 2003). The purpose of the Integrated Ocean Observing System (IOOS) is to make more effective use of existing resources, new knowledge and advances in technology to provide data and information for global and regional scientific study (Ocean.US 2002). The National Science Foundation's contribution to the U.S., IOOS, the Ocean Observatories Initiative (OOI), will construct a networked infrastructure of science-driven sensor systems to measure the physical, chemical, geological and biological variables of the ocean and seafloor. Greater knowledge of these variables is vital for the improved detection and forecasting of environmental changes and their effects on biodiversity, coastal ecosystems and climate (Consortium for Ocean Leadership (COL) 2009). These achievements have provided data and product resources as well as the communications environment for information sharing and scientific research.

– The ocean management data, in situ observing data, remotely sensed data, and model output are produced in a distributed geographical environment. These data have significant scientific value for scientists and decision-makers in government. There are some studies on distributed geographic information processing that focus on geographic information processing from the Geographical Information System (GIS) view (Yang and Raskin 2009; Yue et al. 2009; Friis-Christensen et al. 2009; Zhang Tong and Tsou Ming-Hsiang 2009; Wang Shaowen and Liu Yan 2009). However, given the relatively high temporal frequency and the intrinsic spatial nature of the data, ocean data integration technology based on the Digital Earth system has not been widely implemented. This study is based on ongoing research in China that seeks to construct the China Digital Ocean Prototype System (CDOPS) as part of the China Digital Ocean Information Basic Framework. The experience with the development of the Digital Ocean prototype system in China is relatively rich and thus helpful for addressing this question.

2.2 Acquiring Ocean Big Data

There are many kinds of ocean data produced every day, which operates as an elementary component for the study and application on the oceans. In this chapter, the big data cognition for ocean is studied from four perspectives. Form the data

volume perspective, the ocean data acquire and analysis can produce big volume data. Form the data velocity perspective, the ocean data is collected from the eyes in the sky and objects on-the- ground networks, together with demographic, geologic, and socio-economic data and model estimates. Form the data variety analysis perspective, the database stored data and the unstructured data and model estimates are included. Form the data value mining perspective, there are many kinds of knowledge can be mined. Among the ocean big data, the three-dimensional data play an important role.

The three-dimensional data includes Digital Elevation Model (DEM) data for the seafloor and coast, in situ observational data, remote sensing data and model output data.

The DEM data are acquired from single- or multi-beam ship-borne echo sounders, which are the traditional systems used to map the seafloor topography with high precision results; in addition, data are produced from airborne AIRSAR/POLSAR synthetic aperture data or other methods (Maged et al. 2009). The DEM data from below the sea surface are used to construct the three-dimensional seafloor model as well as the model of the sea surface. The DEM data from the coast are used in land surface three-dimensional modeling, remote sensing data with different spatial resolutions represent as the surface texture.

Station observation data better reflect the environmental condition of the waters within their zones, and the changes in data offer a certain representativeness. Station observation also offers the characteristics of continuity, accuracy, and timeliness and mainly reflects the two aspects of continuous time and continuous space; a reasonable measure of space is continuous in the horizontal direction of the site layout and the vertical direction of the air, surpolygon and subsurpolygon. With regard to ocean profile measurement; continuous time refers to the various processes that can be captured by long-term continuous observation data; its accuracy is derived from the use of the infinite sequence of sample recovery. We care about the continuous process of change. With regard to the station system, on the one hand, the field observation data are a fast, accurate and reliable means of communicating in real time with ocean forecast and other departments for the purpose of controlling the ocean environmental features and evolution process. On the other hand, the field data are a historical resource that can be stored permanently.

The continuous space means in the horizontal direction and vertical direction of the air, surface and subsurface, and ocean profile measurement. The continuous time refers to the various processes in long-term continuous observation.

A self-propelled type of ocean platform is mainly used for underwater, unmanned, wide range, extended underwater environmental monitoring, including the physical parameters, the ocean geology and geophysics, and the ocean chemistry and biology parameters as well as aspects of ocean engineering – all of which can be performed close to the observation area. Its features are the following: low cost; environmentally adaptable; able to surpass the artificial diving limit and enter the field observation area; small size; easy to use; easy to wash; operates according to the acoustic signal remote control or preset program control; built according to

the requirements for related observation projects; independent power and relatively long underwater running time; noise is low; and can be hidden from observation.

The data buoy is anchored or floating on the ocean observation platform; plays an important role in the ocean observation system. Although remote sensing by air and satellite can be performed at great speed and over a wide area, only the surpolygon data are accessible; the unattended buoy submerged in an ocean environment works continuously for a long period of time in combination with other buoys for a comprehensive, profound, all-weather assessment of the ocean environment and the changes that it undergoes.

The main technical instruments for field observation include the following: a specially designed oceanographic vessel and salinity (conductivity)–temperature depth gauge (CTD), acoustic Doppler velocity profiler (ADCP), profiler, side sonar, an underwater vehicle and underwater laboratory, underwater robots, and equipment for seabed deep drilling. Direct observation of the data and mathematical models used to provide reliable reference can also be verified by the results of the experiment and mathematical methods. In fact, the use of advanced research vessels, test equipment and technical facilities for direct observation have indeed promoted the development of ocean science; especially since the 1960s, almost all of the major progress in this field has been closely related to the use of these resources.

Direct observation data can either be used as reliable information for the experimental study and mathematical model or to verify the results of the experiment and mathematical methods. The basic features are direct observation data authenticity and discreteness. Direct observation data are the real basis for understanding complex ocean phenomenon and the calibration of the model test; in addition, the remote sensing data application plays an irreplaceable role. In other words, these are the most basic data, and they serve as a reference for other forms of data and theoretical results. The development of monitoring technology for the ocean environment improved monitoring, prediction and forecast ability and therefore promoted the development of the ocean and coastal economy.

The in situ observational data are being collected from different types of sensors on a range of time scales, such as the data produced by the Argo buoy. The data-obtaining devices include oceanic optical buoys, seafloor moored instrumentation (for dynamic factors), self-locating sub-water tide monitors, sea sound detection buoys, sub-water tide comprehensive admeasuring apparatuses, cruises and others. Real-time and delayed-mode monitoring data collected from the above equipment are transmitted to a local data center. The characteristics of the in situ observational data are that it is related to a fixed x, y, z location in space and has attributes that are attached or related to the location. Most of the data have the required temporal attributes.

– Satellite remote sensing covers most of the ocean environment parameters and information, including sea surpolygon temperatures, ocean currents, sea surpolygon wind fields, concentrations of chlorophyll, suspended mass concentrations, sea levels, gravity anomalies, ocean optical parameters, atmospheric

aerosol ocean rainfall, direction of the wave spectrum, and sea pollution in the surpolygon. Satellite ocean remote sensing involves the electromagnetic wave scope of visible light, infrared waves and microwaves. Visible light remote sensing using the sun as a light source, thermal infrared remote sensing using the sea, and passive microwave remote sensing can be divided into the surpolygon of the microwave radiation source and spaceborne microwave remote sensing active source. Satellite remote sensing instruments presently use radar scatterometers, radar altimeters, synthetic aperture radars, microwave radiometers, visible light/infrared radiometer ocean color scanners, etc.

– The radar scatterometer is an active microwave device; strabismus observations can show the surpolygon wind speed, wind direction and wind stress and sea waves. Using the scattering of the wave field is a reliable basis for sea condition forecasting.
– Spaceborne radar altimeters are also a type of active microwave sensor; they can measure geoids, sea ice, tides, water depths, sea surpolygon wind intensity and effective wave heights, and El Nino phenomena.
– Synthetic Aperture Radar (SAR) is a high azimuth resolution type of coherent imaging radar; it uses phase and amplitude information and is a type of holographic system that can be divided into side, strabismus, Doppler sharpening and bunching of surveying and mapping as well as other applications. According to the difference in SAR image brightness, the sea ice ridge, thickness and distribution, water and ice boundary, tip height and other important information can be extracted. By using SAR images, not only large-area oil pollution can be found in a timely fashion, but sudden pollution incidents can also be detected.
– The microwave radiometer is a passive microwave sensor that conducts remote sensing by measuring the thermal radiation temperature from the sea surpolygon temperature. Sea surpolygon temperature tests one of the most basic parameters, water temperature, which is one of the main measures for determining water mass and is also used to analyze the ocean front and flow. NOAA – 10, one of the last three satellites in the United States, is an advanced, very high resolution radiometer (AVHRR); an image represented by the sensor can be accurately mapped to the surpolygon with a resolution of 1 KM, and 1 °C temperature accuracy is possible. Satellite remote sensing of the sea surpolygon temperature of the global ocean isotherm distribution reveals complex phenomena that conventional methods were unable to discover and has even corrected previous findings.

The visible/near infrared bands of the multispectral scanner and coastal zone color scanner act as passive sensors; these devices measure ocean color, suspended sediment, water quality, etc. Using multi-spectral information and reflectance, suspended load concentration and migration can be extracted. Satellite remote sensing images can display frontal systems, eddy currents, and bodies of water, e.g., mesoscale ocean phenomena, and, combined with other satellite data research, can reveal the phenomena of many dynamic ocean mechanisms and processes.

– Compared with the traditional ship-based monitoring and buoy data, satellite synchronous measurement has major advantages in large areas, higher spatial resolution, long phase measurement, and the combination of multiple platform observations can capture various types of regional phenomena, thus affecting change research and the demand for global change research. In terms of relative sea direct observation, the cost of data is also very low. Satellite remote sensing has obvious disadvantages: observation is only possible on the surpolygon of the sea surpolygon, while visible light remote sensing reaches tens of meters into the sea surpolygon, but this is not sufficient; Various sensors have inherent defects, e.g., infrared remote sensing and spatial synchronization coverage problems in ocean color remote sensing.

Ocean numerical model data is also one of ocean big data sources. With the rapid development of modern numerical computing technology, more and more complex, accurate and efficient ocean numerical models have been developed.

The ocean numerical model is also a data source. Describing the ocean is very complex, and there are numerous forms of mathematical equations, but only in rare cases is it possible to obtain an accurate solution; however, numerical equations and methods can solve these problems to a great extent. With the rapid development of modern numerical technology and computer technology, increasingly complex, accurate and efficient ocean numerical models were created. The Princeton University POM model is one of the most widely used numerical models. MM5 adoption provides much useful data for the research on the atmospheric environment of the ocean. These powerful numerical models can obtain accurate ocean movement scenarios through measured data validation. The ocean model also produces a wide range of data, which can be used for further research.

Ocean large field data commonly uses the form of grid data model. Advantages in the form of a grid data model are as follows.

The main numerical model output products and all kinds of remote sensing observation data are in the form of grid.

The original format of the data, in line with the data acquisition and storage, are studied. The current large data are obtained in the form of grid, the main source of numerical model output data products and all types of remote sensing observation data.

The form of the grid is simple and intuitive and conducive to application. In view of the current data used to evaluate this grid form, it is simple and practical but still maintains a variety of scalable features; it is sufficiently flexible to meet current needs. There are simple topological relations; thus, creating, recreating and all the basic operations are easy.

– The grid is flexible and efficient. Compared with the vector format of geographical information systems, the grid data operation is very easy but not inferior to the vector format with respect to many functions; the grid form can undertake all types of space operations and complete many types of spatial analysis functions, and the straightforward features make it consumer friendly. The model data are usually stored as a multilayer stack of column and row data. The structure can be

defined as regularly or irregularly spaced grid units, with discrete node locations defined in the x, y and z dimensions. The grid nodes or cell values vary over time, and each layer represents a unique slice in space with a given depth.

Heterogeneous data formats increase the difficulty of data sharing and application, prompting the formation of data standards. Currently, ocean field data can be stored in more common data formats, such as the HDF (Hierarchy Data Format) and NetCDF (Network Common Data Form) formats,

Binaries with a small footprint, which offers the advantage of a fast read speed, are more common in the ocean business. As business needs or other application requirements differ, the internal data organization are not the same, and the actual application requirements for the organization of data for analysis is a prerequisite for the processing of the data.

Routine hydrological elements are seawater temperature, salinity, density and sound velocity, all of which contribute to the flow field numerical model for 3-D mesh data. According to the data grid resolution, they can be divided into uniform grid data or variable grid data. Grid data adopt an isometric grid, and the grid resolution will be according to the calculation model; different bands have different resolutions.

Two-dimensional grid data include tidal field water-level data, storm surge water-level data, and surface flow data; this is the basis for the study of ocean phenomena regarding spatial-temporal processes. In terms of the data for the ocean surface, there is no direct concept for the "deep layer", and we enter it as two-dimensional grid field data. However, in the process of time and space analysis of the data, the data and time dimension can also be structured for 3-D data.

Other grid field data file formats are text files, tables, etc.,; the text file is not widely used only because it occupies more space, often as an intermediate data format for quality checking.

Using a numerical model to provide data for ocean research generally involves the integration of fairly measured data; in ocean science, this is often called "data assimilation"; using measured data (including mature remote sensing data), improves upon the drawbacks of the numerical method; the process of calculation can amount to continuously effective intervention. Generally, export data include two categories: data from a variety of data sets that produce export data, which are usually required to pass through the interpolation algorithm that is plugged into the grid-type data.

The ocean big data and information described above are archived or stored in more than one department because they are usually collected or produced based on a project-based approach at different times. At the same time, interdisciplinary research requires different types of data at different times. To effectively and efficiently integrate a substantial amount of ocean big data, one global visualization system should facilitate the sharing, accessing, exploration and visualization of these datasets for scientists, users and decision-makers.

2.3 Characteristics of Ocean Big Data

The ocean is dynamic, and continuous, with fuzzy boundaries of time and space; compared to other industries, ocean data have obvious features. In general, the data characteristics of the ocean are the following.

2.3.1 Acquired from Multiple Sources

Here, data sources are addressed, including data acquisition means and methods. The common marine data types include the in situ observation data, remote sensing observation data and model output data and so on. For example, the ocean environmental elements data can be obtained from different survey platforms, such as CTDs, drifting buoys, sea buoys, ARGOs, and satellite remote sensing, etc.

2.3.2 The Variety of the Data Content

Data reflect the diversity of content including phyletic and various – such as basic hydrological – elements such as temperature, salinity, density, sound velocity, and current; ocean phenomenon data include tide, tidal current, and storm surge; meteorological data involve sea surpolygon wind fields, air pressure, precipitation, relative humidity, etc. as well as biochemical data, geophysical data, etc. Different categories of data on the properties, management and application are not entirely the same. These different uses of the data content as well as the data management, maintenance and solutions to problems, especially in terms of analyzing and accurately obtaining relevant information, are very difficult and require a great deal of manpower and time.

2.3.3 The Heterogeneity Storage Format

Different measuring instruments, calculation methods and tools, data standards have resulted in heterogeneous data storage formats. Different production sectors use custom formats; in addition, semantic heterogeneity, heterogeneous coding, accuracy and data processing heterogeneity and a series of heterogeneous features.

Data from different sources are generally not the same in storage format, even if they are the same type of data, because of the variety of measuring instruments; measuring instruments also have different storage formats, for example, buoys (including mooring buoys and drifters), Nansen stations, different CODAS,

CTDs, ADCPs, and observation ships; remote sensing, satellite and other means of observation have different accuracies, and the data are in different formats.

The people dealing with these methods and tools for ocean data platforms formed a different storage format. In the past, file formats were used to store data; even now, the file format is an important means of data storage. However, the database, with its clear structure, convenient operation, ease of sharing, support for large amounts of data, and other incomparable advantages, has gradually become the most common type of data storage. File data, e.g., the most general text files, binary files, and ranks in the spreadsheet file (such as EXCEL spreadsheets), are suitable for a network with XML files, etc. The database data includes both small- and medium-size databases, such as Access, My SQL, and SQL Server, as well as large databases, such as Oracle, DB2, Informix, and Sybase. For basic geographic data, due to the use of different GIS platforms, the data formats are Shp, Coverage, Map, Tab, Mif, etc.

Heterogeneous data formats result in the difficulty in data sharing, exchange, and applications. There are many general Data Formats such as Hierarchy Data Format (HDF) and Network Common Data Form (NetCDF). The HDF file format is a hypertext document format developed by the University of Illinois. The HDF is suitable for a variety of computer platforms, easy to expand, mainly used to store by different computer platforms to produce various types of scientific data. The NetCDF was created by the University Corporation for Atmospheric Research (UCAR) under the Unidata program according to the characteristics of scientific Data; this format puts forward a kind of oriented array type, suitable for the network sharing. Additional Extensible Markup Language (XML) data are necessary because of its strong data description capabilities, it has in fact gradually become the network data exchange standard. Oceans in such locations as the United States, Australia, Russia and Europe are suitable for the Ocean XML application, and a great deal of development work has successful applied the XML data exchange, data processing and storage, daily management, etc. Because of the different data standards, the data storage format often differs, and different data are used according to the needs of the business; the production department will design their own data format specification. The ocean data format for heterogeneous data is a solution to data sharing.

In the actual data process, we should adopt a unified data structure, to achieve a variety of compatible purposes.

2.3.4   The Large Data Volume

The accumulation of observation involves progress and time; each department has accumulated an increasing amount of historical data; for example, satellite remote sensing for the study of ocean phenomena provides a new set of data, amounting to more than a century of data when combined with ship and buoy data.

After years of ocean observation and special surveys regarding the ocean environment, information has accumulated in areas such as physical oceanography (hydro), ocean surpolygon meteorology, ocean biology, ocean chemistry, ocean environmental quality (pollution), ocean geology, ocean Earth content physics, ocean aviation and satellite remote sensing – amounting to hundreds of gigabytes of data regarding the ocean environment in various disciplines.

2.3.5 The Velocity of Data

Ocean data has the characteristics of spatial and temporal dynamic. The ocean is dynamic and changing, and the dynamic nature of this environment is unlike land; in general, the dynamics of land do not involve the movement of an entire field, for example. Often change is partial and involves only a small area or boundary but lasts for a long period of time; the ocean, however, is changing every moment on a global basis.

Features of time and spatial characteristics and attributes are the three basic characteristics of temporal geographic data, and they are always changing. In time, the state of the space (including the actual space and shape of the space as well as spatial relationships) and its attributes will change. Basic hydrological factors (temperature, salinity, density and sound velocity, current, etc.) of time and space change dynamically, and this is one of the most properties of the ocean. In general, the ocean can be viewed as changing all the time, over a period of a few hours or even a few minutes. A number of attribute values change, which cannot be ignored. The basis of hydrological research is the study of elements of an area of the sea whose common attribute values change. The most obvious ocean data are related to ocean phenomena (storm tides, typhoons, tides) dynamics; the dynamic changes in time and space of ocean phenomena and attribute changes are given equal attention; for example, storm surge processes over time, storm surges in spatial locations (spatial information) and water level values (attribute information) are always changing.

Usually, a series of discrete time data are used to record the changes in hydrological elements or ocean phenomena through the dynamic process of interpolation to simulate changes in time and space. The temporal and spatial characteristics of ocean disasters receive the highest number of requests. The research in the field of geographic information systems has always been a problem that is difficult to overcome and requires constant exploration.

2.3.6 Multiple Temporal and Spatial Scales

From the perspective of oceanography, in terms of the time scale, the ocean environment and climate change characteristics for both information scales (i.e.,

for chronological or dated information) change with the seasonal variation in characteristics; there are synoptic scales that describe the variation in the time scale in days or hours. Thus, time scales range from hours to days, half-days, months, 10-day periods, seasons, years, and other scales. On the spatial scale, environmental information can be divided into large, medium and small categories; this includes phenomena that involve areas of 1000 km or more, such as the Kuroshio, areas measured in kilometers, such as the mesoscale vortex, and areas measured in meters, such as the mixed layer. In addition, spatial data production is the process of determining multiple spatial and temporal scales for ocean spatial data. Time scale performance data are produced in different time sequences. Some stations measure observation data once every hour, and some test once every day or even longer. Numerical model export data are measured according to different purposes. Time scale forecast data are typically measured in hours, and time scale statistical analysis of the data involves years, quarters, or months. Spatial multiscale performance involves precision measurement of regional research and data, e.g., ocean remote sensing data from different satellites. Different spatial resolutions result in multiscale spatial data; examples include the Taiwan Strait and Chinese offshore waters; areas on a global scale, such as the Taiwan Strait, have a complicated flow architecture. In terms of areas prone to coastal storm surges, the spatial scale of the research is generally the Taiwan Strait.

2.3.7   Multi-Level in Depth

According to the ocean study objectives, the ocean data are required in different levels in depth. The ocean surface wind and wave data are mostly used for the mankind applications. While the water temperature, density, currents, chemical composition are different in different depth.

2.3.8   Multiple Levels of Data Users

From the point of view of data applications, the types of users differ because the task-level difference is substantial for the requirements of ocean disaster data. The end user data has three main categories: (1) the public; (2) government management departments; (3) and experts and scholars engaged in scientific research. Clearly, they care about the data product level, which can make a large difference in the corresponding data access levels; experts engaged in scientific research and scholars need primitive-level and feature-level information; government regulators need decision-level information; public policy makers require information at the decision-making level, i.e., conclusive information. In the data production and distribution process, different levels of data content, quality, and additional information make a large difference. Data for production department real-time

performance, quality, and confidentiality, as well as the different needs of users should be introduced into aspects of various levels of data in different production cycles.

2.3.9 The High Potential Value

The data obtained by numerical simulation or observation equipment cannot directly address the complicated relationship between each element; only through deeper analysis, calculation and visual expression can we better understand the data and phenomena and then improve the level of prediction and forecasting. Ocean disaster data are in the form of points, lines, polygons and bodies. In computing visualization, data can be divided into vector data and scalar data. The elements of visual expression are now concentrated in a single static expression, and ocean data are characterized by degeneration; again, there is a mutual influence involved in the interaction between the elements of the relationship; thus, there is an urgent need to study more elements as well as a need for the dynamic visualization of visual expression.

2.4 Primary Study on Ocean Big Data Integration Technology

The Digital Earth System based three-dimensional ocean environment is the basis for the integration of ocean big data. The three-dimensional ocean environment is composed of three parts: the three-dimensional coast, the ocean surface and the three-dimensional ocean floor. The three-dimensional coast is constructed from the DEM data and the land remote sensing data with multi-resolutions. The ocean surface is constructed by pictures with lumpy textures in commercial Digital Earth System platforms such as Skyline. The three-dimensional ocean floor is constructed by the DEM data of the seabed with gradual color rendering. In the three-dimensional ocean environment, a user can zoom all the way from space right down to street level on the coast (with images that in some places are sharp enough to show individual houses or even people) and through the ocean surface to the ocean floor. Some three-dimensional biological or man-made establishment models can also be added between the ocean surface and the three-dimensional ocean floor. Because actual commercial modeling platforms are mainly used for land environments, they have some shortcomings in digital ocean applications, especially in the construction of the three-dimensional ocean environment. Firstly, because the surface is constructed by pictures with a lumpy texture, the region representing the borders of different pictures looks discontinuous. Secondly, the visual resolution can change when users zoom from space right down to sea surface level.

Thirdly, the system running efficiency falls with the accumulation of pictures. Fourthly, it cannot simulate the actual scene to include features such as the ocean current direction. The reason for the shortcomings above is that these commercial systems don't provide programming interfaces for the ocean application. An independent scientific digital earth system can make up for these shortcomings through special arithmetic research and through software development. Because the Component Object Model (COM) has nothing to do with programming languages and can be issued by a binary specification, the scientific digital earth system with the global ocean current presentation function is designed to be developed with COM technology. Consequently, this system can be integrated with a commercial system like Skyline.

– The ocean management data are usually stored in tables of relational database management systems (RDBMS) such as Oracle or SQL Server and the like. And a database record is linked with a geographical feature in a map layer using the same ID number code as the attribute data. As the map layers can be added and visualized on Digital Earth systems, the ocean management data can accessed and viewed through selecting of the corresponding ocean management object such as a fishery region and the like. If the ocean management database is located in the same Intranet network environment with the Digital Earth system application server, the data are accessed with the standard interfaces which include ADO, OLEDB and others. If the ocean management database is located in the distributed network environment with the Digital Earth system application server, the data are accessed through web services. The main style of web services is Simple Object Access Protocol (SOAP). SOAP provides a standard message protocol for communication based on Extensible Markup Language (XML). SOAP web services have two main conventions: any non-binary attachment messages must be carried by SOAP and the service must be described using Web Service Description Language (WSDL). Because the Digital Earth systems have shortcomings in spatial data editing and analysis, it must be used through integration with the two-dimensional GIS system such as ArcGIS and the like to meet the requirement of ocean management. By controlling the display of two web pages which are embedded the ActiveX control of the Digital Earth system and the two-dimensional GIS system, the switch can be realized between three-dimensional Digital Earth display and two-dimensional display. The communication is realized by XML coding to keep the synchronization in spatial location of the two systems. The data synchronization is realized through sharing and access the same database.
– For the in situ observation data, spatial location (x, y, z) and time (t) are critical to the representation and exploration of the data and for further analysis. The location (x, y, z) is used by the Digital Earth systems to display the three-dimensional models of the observation apparatus such as buoys and the like, which are constructed by three-dimensional (3D) modeling software systems such as 3DMAX, Multigen Creator and so on. The measurement data at different time are stored in database tables as records and integrated with

three-dimensional model of the corresponding observation apparatus display in the Digital Earth system by a same ID code. The parameter introduction information can also be included in the web page linked with the apparatus by hot link As the number of the 3D models can influence the system running efficiency, the integration technology mentioned above is suitable for the observation apparatus which are moored at a fixed location. For the drifting observation apparatus, the display on the Digital Earth systems should be used by the two-dimensional symbols and the database tables design should be added the spatial location (x, y, z) information with each record.

– For remote sensing observation data, three integration technologies are used according to different data storage formats. If the remote sensing information products are in format of JPEG or GIF, they can be added to the Digital Earth systems directly as data files. If the remote sensing data are stored in a Spatial Data Engine (SDE) such as ArcSDE and the like, they can be added to the Digital Earth systems through SDE access interface. If the remote sensing data are stored in the Hierarchical Data Format (HDF), they can be added to the Digital Earth systems through Web Map Services (WMS) which is standard web service conformed to the standards set by the Open Geospatial Consortium (OGC).

– According to the spatial area of the model output data, the model output data are divided into two types, i.e. the global data and the regional data. For the global data, it is appropriate to visualize in the globe style. The two-dimensional map is overlaid on the three-dimensional visualizing globe of the Digital Earth systems. For the regional model output data, a special visualization ActiveX control is developed which can be integrated with the Digital Earth systems. Because the commercial Digital Earth systems are not suitable for display multi-layer at different depths, the web control integration technology can make up the shortcoming of it. When a kind of model output data need to be visualized, a region filled with given color is displayed firstly, then the ActiveX control interface is displayed and the model output data can be visualized.

2.5 Applications

The integration technologies discussed in above section are used in the Digital Ocean Prototype System of China. Figure 2.1 shows the system initialization interface and Fig. 2.2 shows the seabed terrain data visualization interface which is a part of the Digital Earth based three-dimensional integration environment.

The ocean management data integration content includes the island management data, economic statistical data, and ocean application region data. Figure 2.1 shows an ocean management data integration example, which is an ocean region application feature, and its attribute information display Fig. 2.3.

The in situ observing data are acquired from different devices which include large-scale environment observing buoys, high frequency ground wave radar, and
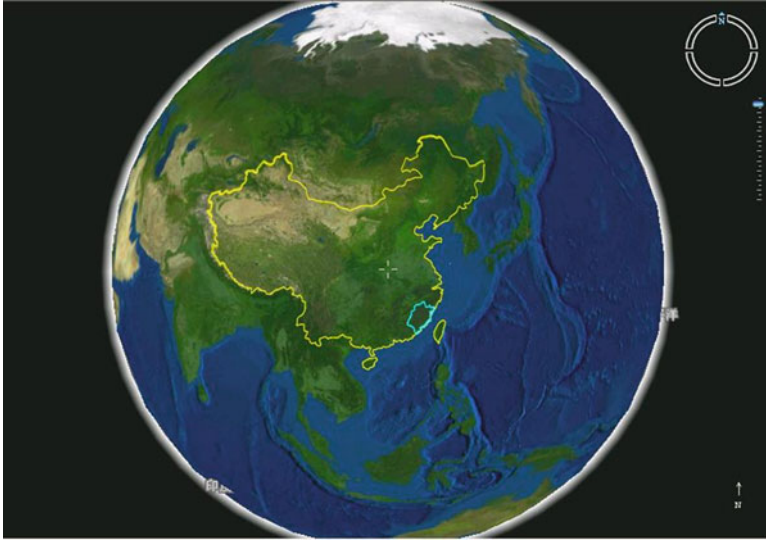
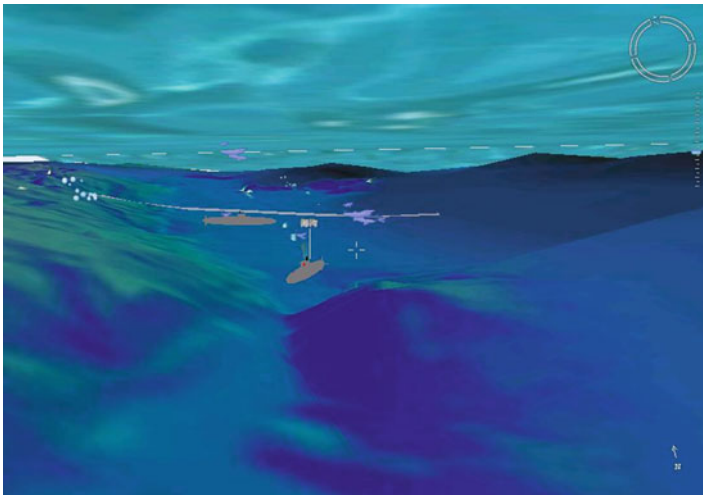**Fig. 2.1** System initialization interpolygon

Fig. 2.2 Digital Earth based three-dimensional integration environment

shipping observing. The observing data are stored in a distributed environment and integrated by a Digital Earth system such as Skyline. The system can help users to access the collected data and devices information. As shown in Fig. 2.4, the key components that make up the data service interface are the map area, the introduction area and data display and the download area. The map area displays the geographical location of the observing device. The introduction area provides the general descriptions of large-scale environment supervision buoys. By selecting the
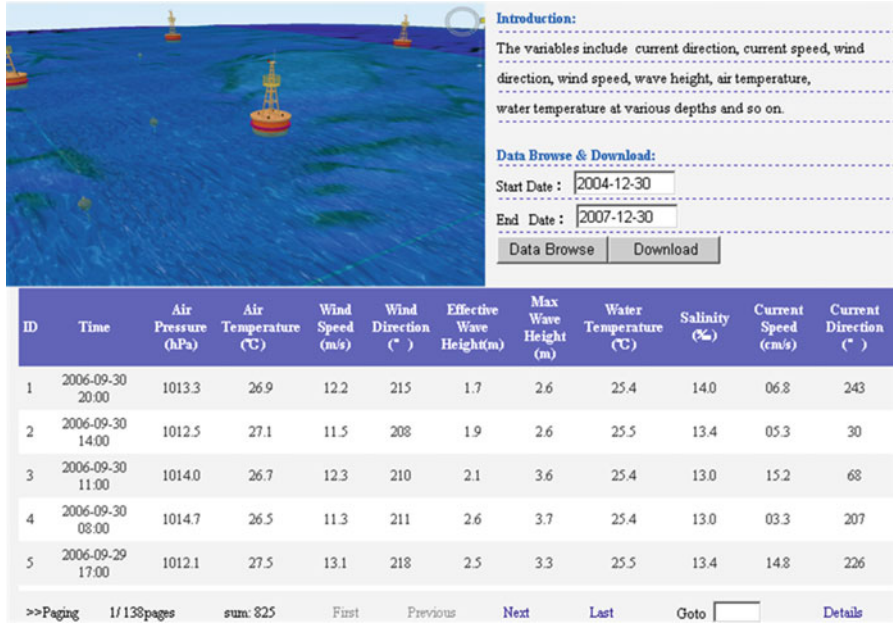
Introduction:

The variables include current direction, current speed, wind direction, wind speed, wave height, air temperature, water temperature at various depths and so on.

Data Browse & Download:

Start Date : 2004-12-30

End Date : 2007-12-30

Data Browse   Download

ID	Time	Air Pressure (hPa)	Air Temperature (℃)	Wind Speed (m/s)	Wind Direction (°)	Effective Wave Height(m)	Max Wave Height (m)	Water Temperature (℃)	Salinity (‰)	Current Speed (cm/s)	Current Direction (°)
1	2006-09-30 20:00	1013.3	26.9	12.2	215	1.7	2.6	25.4	14.0	06.8	243
2	2006-09-30 14:00	1012.5	27.1	11.5	208	1.9	2.6	25.5	13.4	05.3	30
3	2006-09-30 11:00	1014.0	26.7	12.3	210	2.1	3.6	25.4	13.0	15.2	68
4	2006-09-30 08:00	1014.7	26.5	11.3	211	2.6	3.7	25.4	13.0	03.3	207
5	2006-09-29 17:00	1012.1	27.5	13.1	218	2.5	3.3	25.5	13.4	14.8	226

| >>Paging | 1/ 138pages | sum: 825 | First | Previous | Next | Last | Goto | Details |

Fig. 2.3   The large-scale environment observing buoy data integration interpolygon
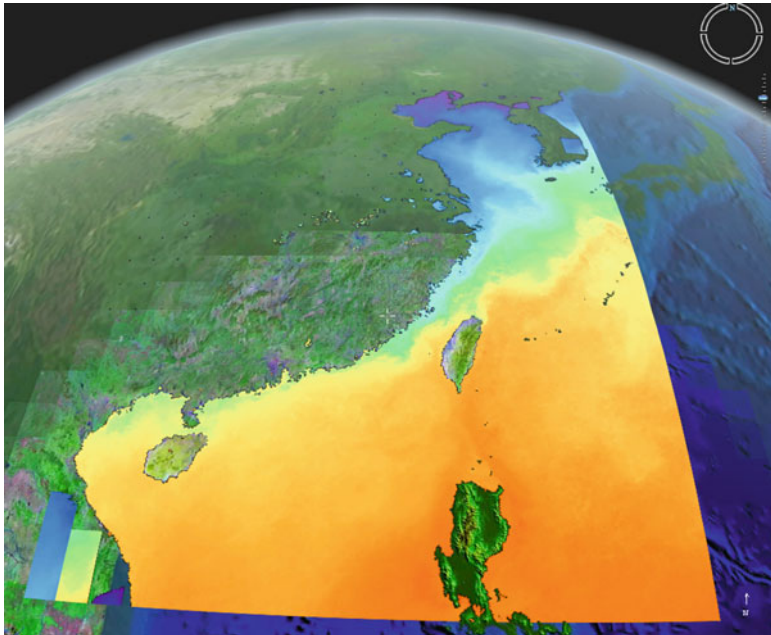
Fig. 2.4   The sea surface temperature (SST) product of MODIS integration
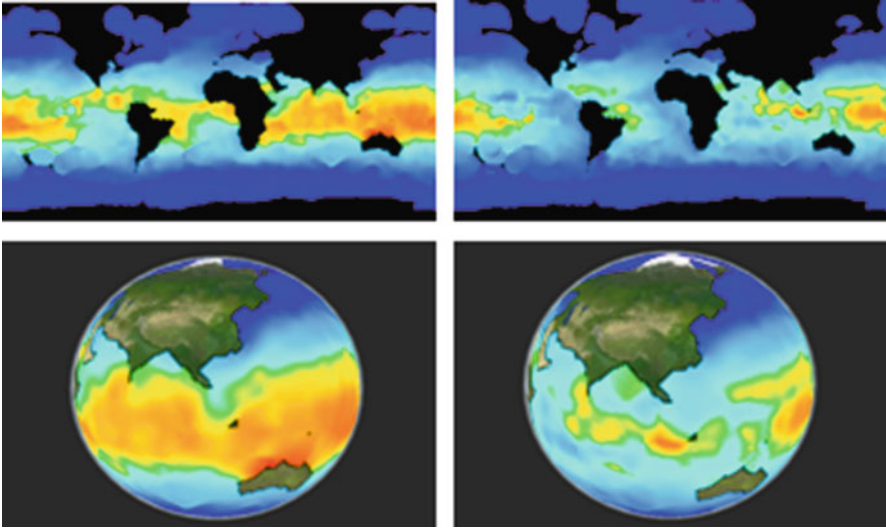
Fig. 2.5 The Integration of the ocean salinity statistical analysis data on different depth dimension

start date and the end date, historical data can be made available to the users in the data display area. The variables include air pressure, air temperature, wind speed, wind direction, effective wave height, maximal wave height, water temperature, salinity, ocean current speed, and ocean current direction. Real time data are updated every 3 h. The data displayed in the table can be downloaded in and XML file. XML was chosen because it is an open standard, and is platform independent and enables easy exchange of information with other organizations.

The satellite observing data integration provides a number of real-time and archived satellite derived products for the Chinese ocean region. The products collected from NASA's MODIS include 250 m RGB, sea surface temperature imagery, chlorophyll imagery and chlorophyll fluorescence. The QuikSCAT wind products and QuikSCAT Level 3 winds products are also available. If the products are available to all users, they can be obtained via direct download from the system. Figure 2.5 shows the sea surface temperature (SST) product of the MODIS integration interface.

Model computing products provide the main ocean data, which include statistical analysis data, ocean models such as the Princeton Ocean Model (POM) computing product data and so on. The data cover the whole or a region of Earth are divided into different levels of vertical dimensions. The original data format is in the grid style.

On the whole, model computing products have some common characteristics.

1. Data includes the longitude, latitude, depth and time, i.e., four dimensions. Longitude, latitude, and depth constitute the element distribution in three-dimensional space; the time dimension determines the time scope of the data set records.

2. The time dimension resolution of data varies because data sources vary, but at equal intervals, it chronicles the elements of the field distribution or elements of the statistical distribution of a certain period of time (mean, median, mode value etc.); and at different times of the nodes on the data, the space characteristics are consistent for those that have the same spatial range and spatial resolution.
3. The depth dimension resolution of the data is determined by access, the latitude and longitude data for different depth ranges and the resolution of two-dimensional spatial characteristics.
4. Data at the same depth layer and the same time point is two-dimensional grid data and applies more in the case of the uniform grid; it is the alternation of longitude and latitude on spacing as a fixed value, but the grid length-width ratio was not necessarily 1.
5. Within a certain area for each grid, it is necessary to record a number, and for the land in the area corresponding to the grid, it is necessary to use a fixed numerical identifier.

According to the above data Ccharacteristics analysis, based on the time dimension, the data can be divided into a series of identical space structure identical three-dimensional data,; using itused in accordance with the depth dimension, they can be further divided into a series of two-dimensional grid data structures, which is consistent with the rules.

From the above analysis, the ocean field element data comprises the three-dimensional space dimensions in addition to four-dimensional time data, but the time dimension can be viewed as the same spatial structure over time, so the time data analysis can be seen as the same operation in different three-dimensional space repeated over time. Its focus lies in the design of the three-dimensional data model.

Due to the need for field element data and general ocean application analysis, the existing three-dimensional data model cannot satisfy their needs, and therefore, it is necessary to study the characteristics and application analysis of the characteristics of ocean scalar fields of feature data based on the design of an appropriate data model.

The data for each depth layer is composed of a regular rectangular grid, and the same is true in the horizontal plane of projection; therefore, the region corresponding to the upper and lower interface can form a rectangular space. Due to the spacing between adjacent layers of uncertainty, rectangular hight can be seen as a variable.

The ocean field data three-dimensional space model can be viewed as the inverse process of data structure analysis. The first step is to build an interface rectangular mesh. The second step is to build a cuboid. The elevation value is only needed because of the single layer of data; to construct the cuboid, it is only necessary to search through a point with latitude and longitude coordinates and compare the elevation values that determine the up and down interface.

At present, the most commonly used approaches in the analysis of the ocean elements includes the following: single point information, profile analysis, plane distribution, and cross-section distribution analysis. The high rectangle model data

structure is one of the foundations of single point information analysis, according to the theory of cuboid model vertex information analysis. Profile analysis can be seen as a vertex information analysis. In addition, the essence of plane and profile analysis is three-dimensional space analysis of the cutting.

A higher cuboid modeling method is used for cutting analysis and will be applied to any area within the scope of the cuboid element analysis. The cutting method can be divided into three types: horizontal cutting, vertical cutting and cutting in an arbitrary direction; In essence, the first two are a special case of the third. Horizontal cutting corresponds to plane analysis, vertical cutting and cross section analysis. Because the analysis of the ocean elements is mostly based on two-dimensional data, any direction of cutting in the analysis of actual business is seldom used.

Because ocean data with a spatial structure can be divided into scalar field elements consisting of the characteristics of a planar grid, the first two special cases of cutting can be easily converted to the simple intersection of plane geometry problem solving and data interpolation.

As a result of current 3-D GIS platforms, terrain scenes and visualization of 3-D model objects and support technology is mature; the visualization technology and 3-D GIS platform can be set up with the aid of this advantage. On this basis, the terrain scene elements of ocean visualization can be more accurate; the image and intuitive data are contained in the message. The general 3-D GIS platform provides the space point, line, and polgon related interface but not special data model analysis and visualization of the interface. Therefore, in the existing 3-D GIS platform, to realize the ocean data visualization of scalar field elements, the key lies in how to make the results of the model analysis and platform point and polgon each other in a relationship that resembles an interface. Based on the above data model, there are two types of analysis.

The two main forms of analysis for the resulting points are numeric and chart. The discovery of some temperature value returns a value; and an analysis of the trend returns is a two-dimensional trend line chart (which may also be other types of diagrams).

For point objects, a three-dimensional GIS platform provides an interface for basic spatial location information but also provides the operational attribute information and type of attribute information for the acceptable data types.

Therefore, in view of the analysis, the analysis results can be used as point object attribute data correlation. Select any point on a sphere, and corresponding point pairs are dynamically created, and this provides the background model analysis results as the attribute data for the corresponding property field; in turn, by its attribute operation interface, visualization is achieved.

In view of the body analysis results for different types of planar space distribution, form a grid (using color for said numerical changes) and vector (in the form of a contour distribution).

For a surface, the 3-D GIS platform provides a very important attribute – the texture attribute.
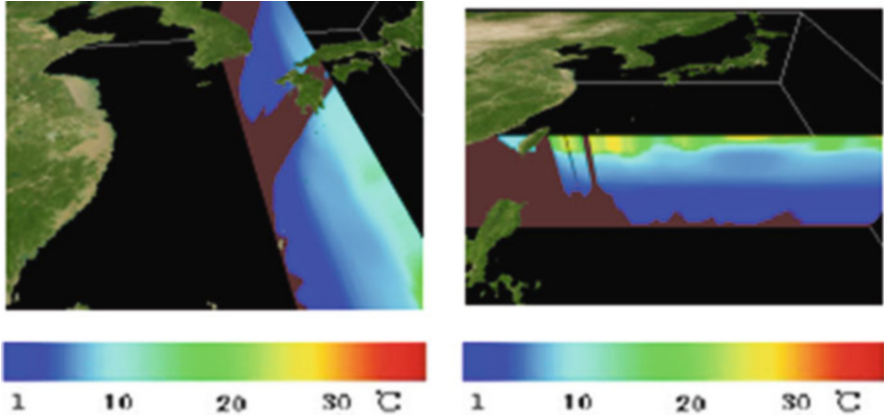
Fig. 2.6   The vertical data visualization integration of a section
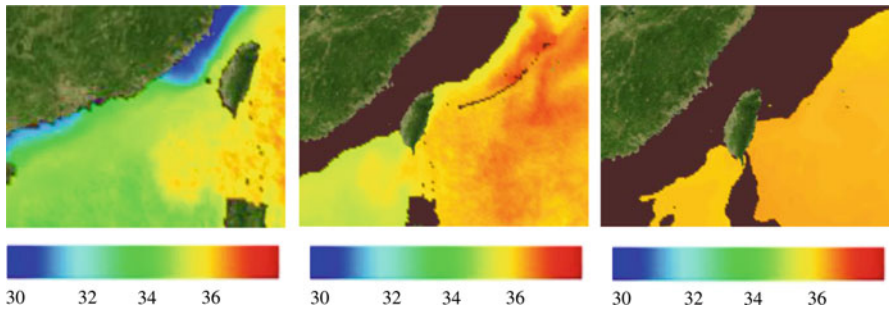

Fig. 2.7   Visualization of model output data of regional seawater salinity at different depth

Thus, for body analysis, in fact, it can be converted into space on a GIS platform that dynamically creates the face image and rendering textures.

Figure 2.6 shows the integration result of the ocean salinity statistical analysis data at the polgon and −75 m. Figure 2.7 shows the vertical data visualization integration result of a section.

2.6   Conclusions and Future Work

The ocean data is a typical big data, which can be seen from the data volume, velocity, variety, and value perspectives. This chapter studies Digital Earth system-based ocean data integration technology, which includes ocean management data, in situ observation data, remote sensing observation data and model output data. The application in the Digital Ocean Prototype System of China shows that the method can effectively improve the efficiency and visualization effect of the data.

In future work, we will study more international standard interpolygons or protocol-based integration technologies to achieve interoperability, which includes OGC-SWE and the IEEE 1451 standard.

References

Al Gore (2010) The digital Earth: understanding our planet in the 21st century [online]. Available from: http://portal.opengeospatial.org/files/?artifact_id=6210. Accessed 14 July 2010

Chongjie Zhang, Chirag Dekate, Gabrielle Allen, Ian Kelley, Jon MacLaren (2007) An application portal for collaborative coastal modeling. Concurr Comput: Pract Exp 19(12):1571–1581

Consortium for Ocean Leadership (COL) (2009) Project execution plan [online]. Available from: http://www.oceanleadership.org/wp-content/uploads/2009/04/. Accessed 14 July 2010

Friis-Christensen A, Lucchi R, Lutz M, Ostländer N (2009) Service chaining architectures for applications implementing distributed geographic information processing. Int J Geogr Inf Sci 23(5):561–580

GEO (2005) The Global Earth Observation System of Systems (GEOSS) 10-year implementation plan [online]. Available from: http://www.earthobservations.org/documents/. Accessed 14 July 2010

Guo Huadong (2009) Digital Earth: ten Years' development and prospect, advances in Earth Science. Int J Digital Earth 24(9):955–962

Guo HD, Fan XT, Wang CLA (2009) Digital Earth prototype system: DEPS/CAS. Int J Digital Earth 2(1):3–15

Guo HD, Liu Z, Zhu LW (2010) Digital Earth: decadal experiences and some thoughts. Int J Digital Earth 3(1):31–46

Malone TC (2003) The coastal module of the Global Ocean Observing System (GOOS): an assessment of current capabilities to detect change. Ocean Policy 27(4):295–302

Ocean.US (2002) Building consensus: toward an Integrated and Sustained Ocean Observing System [online]. Available from: http://www.ocean.us/documents/docs/Core_lores.pdf. Accessed 14 July 2010

Ocean.US (2010) The National Office for Integrated and Sustained Ocean Observations, first annual Integrated Ocean Observing System (IOOS) development plan-a report of the National Ocean Research Leadership Council Prepared By Ocean. US [online]. Available from: http://www.ocean.us. Accessed 14 July 2010

Strain L, Rajabifard A, Williamson I (2006) Ocean administration and spatial data infrastructure. Ocean Policy 30(4):431–441

Thomas CM (2003) The coastal module of the Global Ocean Observing System (GOOS): an assessment of current capabilities to detect change. Ocean Policy 27(3):295–302

Wang Shaowen, Liu Yan (2009) TeraGrid GIScience Gateway: bridging cyberinfrastructure and GIScience. Int J Geogr Inf Sci 23(5):631–656

Yang C, Raskin R (2009) Introduction to distributed geographic information processing research. Int J Geogr Inf Sci 23(5):553–560

Yingqi Tang, Wong DW (2006) Exploring and visualizing sea ice chart data using Java-based GIS tools. Comput Geosci 32(6):846–858

Yue P, Di L, Yang W, Yu G, Zhao P, Gong J (2009) Semantic web services-based process planning for earth science applications. Int J Geogr Inf Sci 23(9):1139–1163

Zhang Tong, Tsou Ming-Hsiang (2009) Developing a grid-enabled spatial web portal for Internet GIServices and geospatial cyberinfrastructure. Int J Geogr Inf Sci 23(5):605–630