10.2 NVODS: DATA NETWORKING FOR AN INTEGRATED OCEAN OBSERVING AND PREDICTION SYSTEM

Steve Hankin,[1,*] Peter Cornillon[2], Joseph Sirott, and Jonathan Callahan[3]
[1]National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory, Seattle, Washington
[2]Graduate School of Oceanography, University of Rhode Island, Naragansette, RI
[3]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington

ABSTRACT

The National Virtual Ocean Data System (NVODS) is an effort to build the data access component of an integrated ocean observing and prediction system - a framework for broad access to ocean information. The key to success of NVODS will be the ability of a user to easily access data for specified locations and times, without special efforts or insights about the data source(s). The NVODS project brings together academic, Federal, and state institutions as well as corporate and non-profit interests.

System components, important to the overall success of NVODS, that will be discussed are:
1) the Distributed Ocean Data System (DODS), the foundation for data communications within NVODS;
2) the Live Access Server (LAS), which provides a uniform minimal ability to visualize, subset, and "fuse" (overlay and compute differences between) NVODS data via standard web browsers; and
3) the Unidata Aggregation Server, designed to provide a single file virtual view from a multi-file collection.

These components make it possible for suppliers of ocean data to offer their data on the Web at minimal effort -- in some cases requiring only the effort of dropping a DODS server binary onto their computer. Users can access the data in a uniform manner through the NVODS.

Further information about NVODS, DODS, and LAS may be found at
DODS - http://www.unidata.ucar.edu/packages/dods/
LAS - http://www.ferret.noaa.gov/Ferret/LAS/
NVODS - http://www.nvods.org/

1. BACKGROUND

In 1999 the National Oceanographic Partnership Program (NOPP) released a Broad Agency Announcement (BAA) requesting proposals for the Planning and implementation of a 'Virtual Ocean Data Hub' (VODHub) activity as a key element of the full community-based system to broaden and improve access to ocean data.

The Distributed Oceanographic Data System (DODS) group at the University of Rhode Island (URI) successfully responded to this request for proposals. The funding is for three years and began in the summer of 2000. The first year of the report was, as requested in the BAA, designed to obtain community consensus on the basics of the system to be implemented as well as to design this system. The second and third years are to be devoted to implementation of the system.

2. THE PLANNING STAGE

The first year has been completed. It involved a suite of four regional workshops one held in the Northeast, one in the Southeast, one on the Gulf Coast and one on the West Coast, each led by a regional coordinator who is also a Co-I on the project. The focus of each workshop was to identify requirements for the data dissemination elements of a distributed data system. In addition to the regional meetings there was also a meeting devoted to issues related to Geographical Information System (GIS) access to data available via the system. Reports for each of the meetings were produced and then these reports were synthesized into one document summarizing the recommendations. The synthesis report served as background material for a National meeting associated with the project that was held in Washington, D.C. in April 2001. The recommendations of this meeting were then used to define the course of the project over the next two years. All of the reports written as part of the planning phase may be accessed from the project web site: http://nvods.org.

3. VISION

The system that emerged from the planning effort, referred to as the National Virtual Ocean Data System (NVODS), is intended to be both a component of the Global Ocean Observing System (GOOS) and to extend beyond the GOOS effort. The vision of NVODS is that of a data system which provides easy access over the Internet to oceanographic data for the world community of users defined quite broadly. The system is envisioned as having the following principal components:

1. A mechanism to provide projects and institutions who

* Steve Hankin, NOAA, Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115; e-mail: hankin@pmel.noaa.gov.

hold data the means to serve those data via the Internet without having to reformat their data;

2. A universal data transport mechanism which is independent both from file formats and from computer architectures;

3. A set of client software applications including basic network browsing tools that allow the users to access any data in the system, and to seamlessly insert those data into the users' application programs without concern for data formats; and

4. One or more data location sites which will allow the user to discover in a simple and logical manner any datasets containing data of interest to the user which are served into the NVODS system.

NVODS will encompass datasets from all sub-disciplines in oceanography as well as data in related fields such as meteorology. It will enable access to comprehensive data for a myriad of specific uses, ranging from providing real-time and near real-time views of coastal currents, to the monitoring of the variability of specific fish species with varying conditions over years, to the analysis and display of cruise and satellite data in the classroom, to the study of global oceanographic and meteorological change. It will provide marine resource managers with access to comprehensive data from many sources where they have had to rely on a few locally collected datasets in the past.

4. THE DODS DATA ACCESS PROTOCOL

NVODS is a data system involving elements from data location to data analysis. Key to the successful operation of this system is the data access protocol used to move data seamlessly from one element in the system to another. The data access protocol to be used by the NVODS project is based on the DODS data access protocol -- i.e., NVODS inherits all of the data that are currently accessible via DODS as well as the system infrastructure associated with the DODS effort. Basically, NVODS replaces the oceanographic part of the DODS effort which, as a result of the generality of the data access protocols, is being used by a number of other disciplines. At the recommendation of the NVODS Executive Committee, a nonprofit corporation has been formed to maintain and evolve the data access protocol for the wide range of communities that are currently making use of it. This corporation is called the Open Source Project for a Network Data Access Protocol (OPeNDAP).

The DODS data access protocol was originally designed to facilitate access to data via the Internet for the oceanographic research community. The design was however sufficiently general that its usefulness is not confined to the research community. DODS performs as an interface between the user's analysis program and the data that s/he would like to use. Regardless of where the data are stored the data look to the user's application as though they reside on the user's computer in a format recognized by the user's application software.

The architecture of DODS grew from two basic assumptions about the way scientists use data. First, scientists rank data obtained from colleagues as very important, so DODS must support scientists as data providers on a par with data centers. Second, to aid in data processing and interpretation, a scientist will generally adopt one of the many available analysis packages (e.g., IDL) or write his or her own programs and this individual does not want to be required to learn how to use a new analysis package to make use of data from a new data system. Therefore DODS must be easy to interface to existing programs (often referred to as "legacy software"). Many analysis packages allow the user to enter data from several of the more common formats such as NetCDF and HDF, but they have not been designed to handle data stored in an unsupported format in a systematic fashion nor are they generally network-enabled in this regard.

At the lowest level DODS may be viewed as replacing the often cumbersome, yet common practice of ftp exchange of data among colleagues. But DODS is much more than this because it provides access to these data under program control and in a consistent format. This means that the user may use scripting languages available within the more sophisticated analysis packages to access subsets of data from remote sites.

In a typical usage scenario the user requests remote data directly from his or her DODS-enabled application. For example, a Matlab user might request a data subset residing on a remote machine via a DODS-Matlab client. The data are subsetted on the remote machine, returned to the user's computer and entered into the Matlab workspace. This operation could as well be performed with a Matlab script allowing the program unattended to query one archive for data, analyze the returned values and, depending on the result of this analysis, launch a new query to the same archive or to a different archive for more data. Other application packages interfaced to DODS, e.g., Ferret, GrADS and IDL, allow for similar functionality.

DODS uses client/server architecture to provide access to data across the Internet. By installing one or more DODS data servers (standard http servers, equipped with a set of DODS Common Gateway Interface (CGI) programs) any computer with on-line data can provide network access to these data. Users employ client programs to access data from the data servers. References to the data resemble typical file paths prefixed by the "http://IP-NAME/DODS-server" that identifies the DODS

Web server.

5. CURRENT STATE OF NVODS

As of 1 September 2001, in excess of 2 terabytes of data contained in over 350 data sets were available from over 20 sites in the US, France and Korea. A list of available NVODS-accessible data sets or pointers to sites with NVODS-accessible data holdings is available under the *Dataset* pull-down menu on the DODS home page, http://unidata.ucar.edu/packages/dods

6. ACCESSING NVODS DATA VIA THE WEB

A subset of data in NVODS may be accessed via your web browser with the Live Access Server (LAS) developed at Pacific Marine Environmental Laboratory (PMEL) (see also sessions J2.2, J2.3, and J2.4). For NVODS to provide access to data served via the OPeNDAP (a.k.a. DODS) data access protocol, LAS must first address the issue of incomplete or inconsistent metadata that is inherent in any system, which, like NVODS, attempts to minimize the effort required on the part of data providers in providing access to their data. To remedy this problem LAS maintains a local cache of metadata stored in the eXtensible Markup Language (XML). The metadata are ingested into XML via OPeNDAP protocols; perl scripts read the OPeNDAP Data Descriptor Structure (DDS) and Data Attribute Structure (DAS) and use OPeNDAP data queries to determine the coordinates of the variables. The ASCII XML files are easily edited to remedy deficiencies in the metadata. The XML also includes configuration information, such as the names of script files, that may be needed to meet specialized visualization requirements.

LAS is a "traffic cop" style of program. It does not, itself, perform the work of generating output products. LAS accepts inputs (using an XML protocol) that identify the data set(s), variable(s), and type of product (GIF file, NetCDF file, etc.) that is desired. LAS packages this request together with auxiliary information, such as script names and graphical defaults (palettes, contour levels,...), and hands the package off to a "back end" application, such as Ferret or IDL to do the job of creating output products.

The XML files form the core of LAS "intelligence". Using the information in these files LAS generates a data-aware user interface for Web browsers; when the user selects a particular NVODS variable the interface updates its Java and JavaScript input widgets to show the latitude/longitude/depth/time range of that variable. When the user requests a product the XML data base is used to screen out erroneous inputs. Finally, the XML data base is the source of the auxiliary information, such as script names, that direct the back end applications.

Advanced features of LAS, such as the ability to make comparisons between data sets (compute differences and overlay graphics) are also made possible by the XML data base. The geometry information contained in the XML identifies those data sets which can be compared (they must overlap on the space and time axes of the requested output).

There are a number of LAS sites around the country providing access to data in the NVODS system. The primary site is at PMEL: http://ferret.wrc.noaa.gov/nopp/

This is PMEL contribution number 2413 and JISAO contribution number 880.

7. ACKNOWLEDGMENTS

DODS was developed with National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA) and National Science Foundation (NSF) funding. The NVODS effort is currently being funding by NOPP. Salary support to P. Cornillon was provide by the State of Rhode Island and Providence Plantations.