

Development of an Integrated Regional, National and International Data System for Oceanography

Project PI: Peter Cornillon

Graduate School of Oceanography, University of Rhode Island, Narragansett, RI 02882-1197

Phone: (401) 874-6283 FAX: (401) 874-6728 E-Mail: pcornillon@gso.uri.edu

Regional Coordinators:

Mark Abbott – West Coast Coordinator

Oregon State University, 104 Ocean. Admin. Building, Corvallis, OR 97331-5503

Phone: (541) 737-5195 E-Mail: mark@coas.oregonstate.edu

Margaret Davidson – Southeast Coordinator

NOAA Coastal Services Center, Charlestown, SC 29405-2413

Phone: (843) 740-1200 E-Mail: Margaret.Davidson@noaa.gov

Linda Mercer – Northeast Coordinator

Department of Marine Resources/State of Maine, West Boothbay Harbor, ME 04575-0475

Phone: (207) 633-9565 E-Mail: linda.mercer@state.me.us

Worth Nowlin – Gulf Coast Coordinator

Oceanography and Meteorology Building, Texas A&M University, College Station, TX 77843-3146

Phone: (979) 845-3908 E-Mail: wnowlin@tamu.edu

Remaining Co-Is

Brad Butman

United States Geological Survey, 384 Woods Hole Road, Woods Hole, MA 02543-1598

Phone: (508) 548-8700 E-Mail: bbutman@usgs.gov

Richard Chinman

University Corporation for Atmospheric Research/IITA, Boulder, CO 80307-3000

Phone: (303) 497-8696 E-Mail: chinman@ucar.edu

Don Collins

Jet Propulsion Laboratory/NASA, 4800 Oak Grove Drive, Pasadena, CA 91109-8099

Phone: (626) 744-5508 E-Mail: Retired

James Cummings

Naval Research Laboratory, 7 Grace Hopper Ave., CA 93943-5502

Phone: (408) 656-1935 E-Mail: cummings@nrlmry.navy.mil

Glenn Flierl

Massachusetts Institute of Technology, Cambridge, MA 02139-4307

Phone: (617) 253-4692 E-Mail: glenn@lake.mit.edu

Dave Fulker

University Corporation for Atmospheric Research/Unidata, Boulder, CO 80307-3000

Phone: (303) 497-8336 E-Mail: fulker@ucar.edu

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2003		2. REPORT TYPE		3. DATES COVERED 00-00-2003 to 00-00-2003	
4. TITLE AND SUBTITLE Development of an Integrated Regional, National and International Data System for Oceanography				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Rhode Island, Graduate School of Oceanography, Narragansett, RI, 02882-1197				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES A National Oceanographic Partnership Program Award.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 10	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

James Gallagher
University of Rhode Island, Narragansett, RI 02882
Phone: (406) 723-8663 E-Mail: jgallagher@gso.uri.edu

Ted Habermann
National Geophysical Data Center/NOAA, 325 Broadway, Boulder, CO 80305-3328
Phone: (303) 497-6472 E-Mail: Ted.Habermann@noaa.gov

Peter Hamilton
Science Applications International Corporation, 615 Oberlin Road, Raleigh, NC 27605-3129
Phone: (919) 836-7565 E-Mail: Peter.Hamilton@saic.com

Steve Hankin
Pacific Marine Environmental Laboratory /NOAA 7600 Sand Point Way NE, Seattle, WA 98115-0070
Phone: (206) 526-6080 E-Mail: Steven.C.Hankin@noaa.gov

Chris Lynnes
Goddard Space Flight Center/NASA, Greenbelt, MD 20771-0001
Phone: (301) 614-5185 E-Mail: Christopher.S.Lynnes@nasa.gov

Richard Lawrence
Environmental Systems Research Institute, Inc., 380 New York Street, Redlands, CA 92373-8100
Phone: (909) 793-2853 E-Mail: rlawrence@esri.com

Lola Olsen
Goddard Space Flight Center/NASA, Greenbelt, MD 20771-0001
Phone: (301) 614-5361 E-Mail: olsen@lilgcmd.gsfc.nasa.gov

Neville Smith
Bureau of Meteorology Research Centre, PO Box 1289k, Melbourne, Victoria 3001 Australia
Phone: +61-(0)3-9669-4434 E-Mail: N.Smith@bom.gov.au

Ken Tenore
University of Maryland, Chesapeake Biological Laboratory, Solomons, MD 20688-0038
Phone: (410) 326-7241 E-Mail: tenore@cbl.umces.edu

Anders Andren of the University of Wisconsin, Herve Roquet of Meteo-France, France and Ken Turgeon of the Minerals Management Service originally listed as Co-Is effectively dropped out of the project in that they have not participated in the effort in any way since receipt of the award.

Award Number: N000140010889
Related web sites: <http://nvods.org>, <http://unidata.ucar.edu/packages/dods> and <http://opendap.org/>

LONG-TERM GOALS

As part of a phased approach toward an Integrated Ocean Observing and Prediction System, NOPP¹ in 1999 selected a few preliminary steps for development. One of these steps was to plan and implement “a community-based 'system' to broaden and improve access to ocean data”². The project described herein was proposed in response to the 2000 NOPP BAA and has as its long-term goal the development of such a system; i.e., the implementation of a network based system that would provide for the discovery of and seamless access to oceanographic data. The ultimate goal is a system that will provide immediate access to a vast array of real-time and historical oceanographic archives by all segments of the oceanographic community from the one-time user to the modelers and managers requiring regular and automatic ingestion of the highest quality data and data products.

OBJECTIVES

1. The development of a data access protocol capable of handling the exchange of all oceanographic data types – biological, physical, chemical and geological – as well as products derived from these data. This objective is based on the assumption that at the core of any network based distributed data system is middleware that allows for the seamless exchange of data between system elements.
2. The development of a representative set of data servers and data clients based on the protocol. These system elements will be used to evaluate the protocol as well as to provide a catalyst for the use of the system.
3. The linking of in excess of two terabytes of oceanographic data held in over 250 databases at approximately 40 sites via this system – to provide a core set of data in the system to stimulate its use.
4. The establishment of a diverse group from the ocean data user community committed to the success of the system.
5. The building of the basic infrastructure required to couple data providers and users in the science community with data providers and users in the GIS community.

APPROACH AND WORK PLAN

As requested in the BAA, the approach to this project was to be taken in two phases:

1. In the first year community consensus was to be sought for the basic system and the system was to be designed at a high level, and
2. In the second and third years, the system conceived in the first year was to be implemented.

In the first year there were to be five regional workshops run by each of the five regional coordinators. The recommendations of these regional workshops were to provide the starting point for a system design and implementation plan to be undertaken at a national meeting of the project Co-Is and secondary investigators. In the remaining years, a set of pilot projects designed to demonstrate use of the system as well as to test the system were to be undertaken. These pilots were to be undertaken on a regional basis and directed by regional coordinators. The regional coordinators were also to address population of the system in their region. J. Gallagher was to act as the project’s technical lead, coordinating all of the technical efforts and R. Chinman was to act as project coordinator under the direction of the PI P. Cornillon. Chinman resigned in the second year and was replaced with P. Hemenway of the University of Rhode Island.

¹ Acronyms used in this report are defined at: <http://nvods.org> at the bottom of the page.

² Fiscal Year 2000 National Ocean Partnership Program BAA (As published in Commerce Business Daily, September 16, 1999.)

Over the course of the project, M. Davidson was replaced by A. Ball as southeast regional coordinator, A. Andren dropped out of the project as the Great Lakes coordinator (and was not replaced) and S. Weisberg was added as the southern California regional coordinator. T. Habermann was to be responsible for an interface between the system and existing GISs.

WORK COMPLETED

1. In the first year planning stage 4 regional workshops, one topical workshop and a National Meeting involving all of the Partners were completed. The recommendations from these meetings are available at: <http://nvods.org/> (under "Reports and Presentations"). These recommendations formed the basis for the work undertaken in the remainder of the project. The resulting system is referred to as NVODS, the National Virtual Ocean Data System.
2. Development of the data system core infrastructure:
 - a. Three versions (3.2-3.4) of the NVODS/OPeNDAP³ c++ core software have been released with each incorporating major additions and corrections relative to the previous version. A complete history of the software is available at: <http://unidata.ucar.edu/packages/dods>.
 - b. In addition, a JAVA version of the core software has been implemented and released.
 - c. A specification, based on XML and web services, has been written for the protocol – DAP 4.0.
 - d. An “Aggregation Server” has been developed. It allows multi-file array data sets to be viewed as single data objects, thus dramatically simplifying requests for and return of the data.
 - e. Extensive modifications have been made to the Live Access Server (LAS) at PMEL.
 - f. An Ancillary Information Services (AIS) component has been added to the core software. This allows for the addition of sites that provide a consistent view of the metadata associated with data in the system and for these sites to be seamlessly accessed along with the data.
3. Representative data servers and data clients developed and/or upgraded as part of this project:
 - a. Clients – netCDF, HDF4, HDF5, FreeForm, JGOFS, Matlab4, DSP and JDBC.
 - b. Servers – netCDF client library, Matlab and IDL.
4. Data provided by the system: A fairly complete list of data served via the system is available at: <http://www.unidata.ucar.edu/cgi-bin/dods/datasets/datasets.cgi?xmlfilename=datasets.xml>. This list includes data entered as part of the proposed effort as well as data entered by “outside” groups referenced in section 5.a.i below.
5. Establishment of a diverse ocean user community has been addressed from two perspectives:
 - a. Active recruitment of data providers and software developers beyond the group originally identified in the proposal.
 - i. Data providers from “outside” the project include: TPAC-Australia, U. Hawaii, Caro-COOPS, JPL and MIT-ECCO, COLA, CDC, NODC, NCDC and U. Kansas-OBIS Hexacoral data.
 - ii. Groups “outside” the project making significant software contributions to the system:
 1. COLA has developed the GrADS-DODS Server which includes a powerful server-side computational capability as well as the ability to serve GRIB and BUFR data.
 2. NCAR has developed a grid-FTP version of the core software.
 3. A number of groups have developed NVODS clients – ncBrowse, VisAD, IDV, etc.
 4. Several large projects have begun making use of OPeNDAP: GODAE, NOMADS, ESG-II
 - b. Four pilot projects have been undertaken to help entrain users. These projects are focused on the data management community.

³ The data access protocol, at the core of the National Virtual Ocean Data System (NVODS), has been developed and maintained by the Open source Project for a Network Data Access Protocol Inc. (OPeNDAP), a non-profit corporation established explicitly for this purpose. NVODS core software will be referred to as OPeNDAP in this report.

- i. SCCWRP Pilot, Steven Weisberg – A water quality management project in Southern CA.
 - ii. Shrimp Pilot, Linda Mercer and Philip Bogden – A fisheries management project in the Gulf of Maine shrimp fishery.
 - iii. Gulf of Mexico Pilot, Worth Nowlin – A project to make oil spill model predictions available to response team managers in the Gulf of Mexico.
 - iv. CODAR Pilot, Dave Ullman – A project to provide a consistent view of CODAR data collected from multiple, separately managed sites.
6. Several projects have been undertaken to provide GIS access to OPeNDAP-enabled data servers.
- a. With the requisite geo-referencing metadata OPeNDAP data objects may be converted to a GeoTIFF file using NOAA/NGDC software. These files can be read by most GIS packages.
 - b. A GDAL-OPeNDAP driver for array data and an OGR-OPeNDAP driver for point and vector data have been implemented for use in MapServer, a web-based GIS.
 - c. As part of the SCCWRP pilot an interface to EASy, a GIS, has been written for array data.
 - d. An ArcIMS-OPeNDAP connection has been developed for array data.

RESULTS

1. Infrastructure:
 - a. The Ancillary Information Services capability to add, at a remote site, metadata associated with a data set is seen as a major step in dealing with the lack of metadata consistency in a highly distributed data system based to a large part on voluntary contributions of data. The AIS has proven to be a major factor in bringing NVODS data sets into compliance with standards required by standard software packages such as MapServer.
 - b. The Aggregation Server has proven to be a significant first step in simplifying access to large multi-file data sets. Work to address aggregation of multi-file in situ data sets has begun with a project sponsored workshop (January 2004) addressing aggregation of mooring data.
2. Population of the system is growing rapidly with more than 500 data sets now accessible via OPeNDAP servers at more than 40 institutions worldwide. These servers are responding to over 2 million individual data requests per month from approximately 2000 users and result in the transfer of over 200 Gbytes of data per month. The rapid growth in the number of data providers and data users suggests that the system has reached a point at which it can continue to grow without additional seed funding; i.e., the population effort funded as part of this project. Of particular interest here is the length of time required for the adoption of new technologies. As observed in other similar endeavors, it appears that on the order of ten years of development and implementation is required for a significant system component to be widely adopted; six years of funding from other sources – NASA, NOAA and NSF – received for DODS prior to the NOPP funding reported on herein was used to develop the base on which NVODS was built. Figs. 1-3 show usage statistics for two data providers for the past 28 months. The number of data accesses system-wide (not shown) has grown by approximately 115,000 per month over the past 2 years.

IMPACT AND APPLICATIONS

National Security

The potential for use of the system to access environmental data for National Security related issues is expected to grow with the volume and breadth of data available via the system.

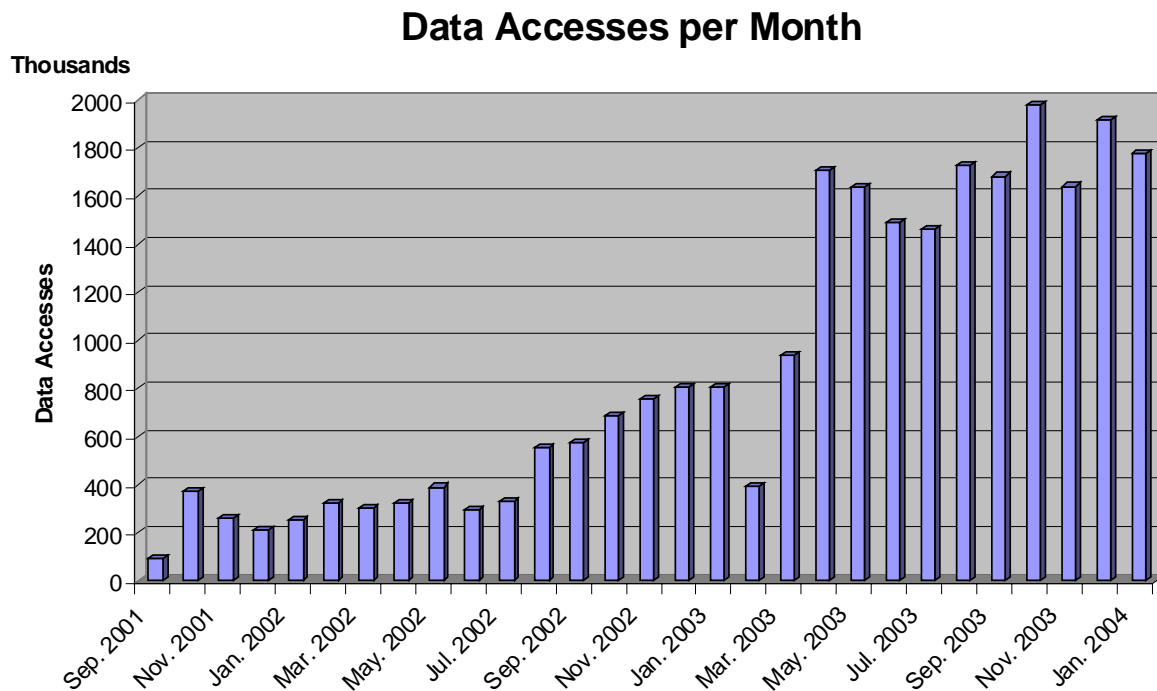


Figure 1. Data accesses per month via OPeNDAP-GDS servers at COLA and UCAR.

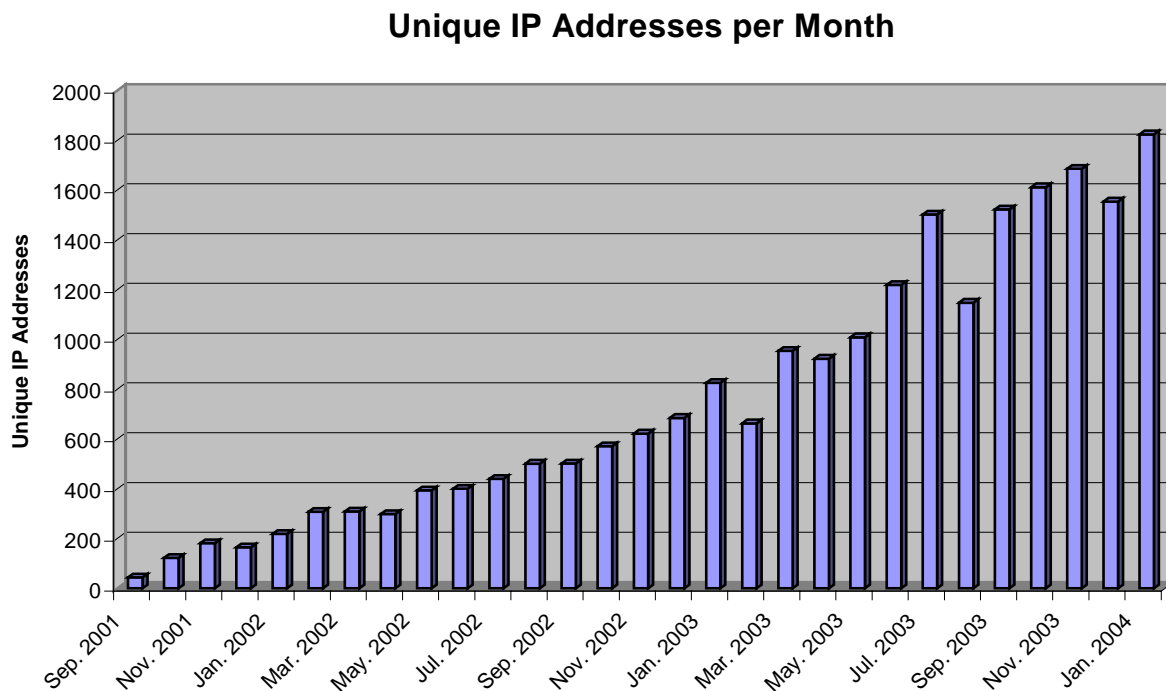


Figure 2. Number of unique IP addresses per month accessing data via OPeNDAP-GDS servers at COLA and UCAR.

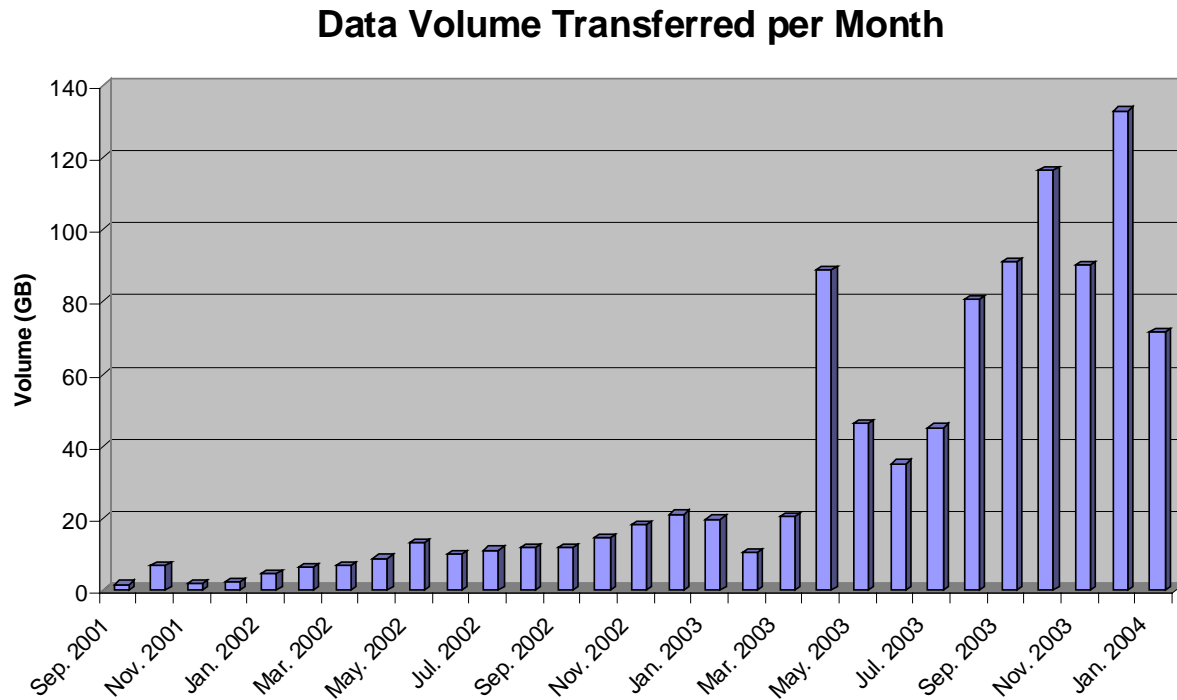


Figure3. Data volume in gigabytes accessed from OPeNDAP-GDS servers at COLA and UCAR. The anomaly in April 2003 resulted from one user downloading a major archive via the system. The system was subsequently modified to prevent such use in the future.

Economic Development

NVODS components are expected to be adopted as elements of future data systems. Adopting system components will result in major time and associated dollar savings in the development of new systems.

Science Education and Communication

Facilitating access to data is at the core of the NVODS effort. NVODS is expected to play a major role in the communication of scientific data both to the research community and to the educational sector.

TRANSITIONS

National Security

Analysis of OPeNDAP access logs suggests that the system is being heavily used by the military to access both meteorological and oceanographic data in an operational mode.

Economic Development

The Data Management and Communications (DMAC) committee of the Integrated Ocean Observing System (IOOS) has adopted OPeNDAP for the data transport component of IOOS. The core software is also being used in a number of other large-scale projects: NOMADS, GODAE, ESG II, PARADIGM, etc. In addition use of the system to help address resource management problems in fisheries, oil spill mitigation and water quality monitoring is being explored via pilot projects.

Science Education and Communication (Delete this section if there are none)

NVODS is heavily used in the scientific community to access oceanographic and meteorological data.

RELATED PROJECTS

The nature of this project has been to form collaborations with a large number of other groups. A number of these collaborations have resulted in a continuing close working relationship that are expected to continue beyond this grant. These collaborations fall into three areas:

1. Many groups are providing data via OPeNDAP servers, but in most cases these groups are doing so with little contact with the project office. Those working closely with the project office and with which we anticipate a close continuing relationship are: W. Nowlin, Texas A&M; B. Butman, USGS; R. Raskin, JPL PO-DAAC; P. Sharfstein, FNMOC; B. Domenico, Unidata; T. Habermann, NGDC; P. Hamilton, SAIC; S. Hankin, PMEL; K. Casey, NODC, B. Blumenthal, LDEO; J. O'Brien, FSU; J. Kinter, COLA; P. Bogden, GoMOOS; R. Schweitzer, CDC; E. Chassignet, U. Miami; F. Muller-Karger, USF; J. Lever, NAVOCEANO; and, S. Kempler, Goddard DAAC.
2. Similarly there are now many groups beginning to develop OPeNDAP-enabled software, clients, servers and specialized functions. Those with which we anticipate a close continuing relationship are: B. Domenico, Unidata; J. Kinter, COLA; S. Hankin, PMEL; D. Denbo, PMEL; N. Potter, OSU; L. Olsen, GCMD; T. Burk, MapServer and, P. Fox, HAO.
3. In addition there are several large projects that are making use of OPeNDAP (a.k.a. DODS) as a component of their data transport layer: NOMADS <http://nomads.ncdc.noaa.gov/>, GODAE http://www.bom.gov.au/bmrc/ocean/GODAE/Meetings/7thIGST/data_sharing_paper.doc, DMAC/IOOS <http://dmac.ocean.us:8101/>, ESG II <https://www.earthsystemgrid.org/about/overview.jsp>; and, COLA <http://grads.iges.org/grads/gds/>.

PUBLICATIONS

Cornillon, P., J. Gallagher and T. Sgouros, "OPeNDAP: Accessing Data in a Distributed, Heterogeneous Environment", Data Science Journal, 2, November 2003, pp 164-174.

CONSIDERATION FOR EXCELLENCE IN PARTNERING AWARD (COMPLETE THIS SECTION ONLY IF YOUR PROJECT FUNDING FINISHED THIS YEAR OR IN THE LAST 12 MONTHS, OTHERWISE PLEASE DELETE THIS SECTION.

1. **Ocean Sector Diversity:** Project partners come from all of the oceanographic subdisciplines – physics, biology, chemistry and geology – as well as from academia, state government, the federal government, industry, and the international community.
2. **Partner Involvement:**
 - a. Mark Abbott – Hosted the West Coast Regional Workshop in year 1. Served OSU data via the Java based JDBC server. Moderate-to-low participation.
 - b. Margaret Davidson – Replaced by Anne Ball who ran the Southeast regional workshop. Moderate-to-low participation.
 - c. Linda Mercer – Hosted the Northeast Regional Workshop. Hosted several other regional workshops to identify a regional project. Organized and ran the Shrimp Pilot. Served on the NVODS Advisory Committee. Moderate-to-high participation.
 - d. Worth Nowlin – Hosted the Gulf Coast Regional Workshop. Worked very actively with regional data providers to help/encourage them to serve their data via the system. Very active

- member on the NVOADS Advisory Committee. Developed the Gulf of Mexico Oil Spill Response Pilot. Organized the Gulf Coast Regional group into the base for a Gulf Coast Regional Observatory. Very high participation.
- e. Brad Butman – Served a significant volume of data via the system. Moderate participation.
 - f. Richard Chinman – Coordinated the overall effort in the first year. Was replaced by Paul Hemenway in the second year. Very high participation.
 - g. Don Collins – Served data from the PO-DAAC into the system. Moderate participation.
 - h. James Cummings – Participated in the first National Meeting of the project. Low participation.
 - i. Glenn Flierl – Very active member of the NVOADS Advisory Committee. Also provided input periodically with regard to system performance and evolution. Moderate-to-high participation.
 - j. Dave Fulker – Replaced by Ben Domenico – Provided user services for the system. Served Unidata's significant near-real time data holdings via the system. Developed the Aggregation Server. Included the OPeNDAP core in the Java version of netCDF. High participation.
 - k. James Gallagher – Project technical lead. Very high participation.
 - l. Ted Habermann – Lead the effort to interface system accessible data to ESRI GIS products. Also developed the GeoTiff output engine. Served NGDC data via the system. Moderate-to-high participation.
 - m. Peter Hamilton – Installed servers for SAIC's extensive mooring archive. Analyzed the organization of OPeNDAP-accessible mooring data and hosted a workshop to develop a common data model for these data. Moderate-to-high participation.
 - n. Steve Hankin – Very active member of the NVOADS Advisory Committee. Actively promoted the system – many of the non-funded data providers were entrained into the system by him. Extended LAS. Provided access to a significant fraction of the NVOADS-accessible data sets via LAS. Installed OPeNDAP servers for PMEL data sets. Very high participation.
 - o. Chris Lynnes – Installed servers for the very extensive Goddard DAAC data holdings. Moderate-to-high participation.
 - p. Richard Lawrence – Worked with Habermann on the ESRI-OPeNDAP data connector. Low participation.
 - q. Lola Olsen – Developed the OPeNDAP Data Portal to the Global Change Master Directory. Worked actively with the project to describe NVOADS-accessible data sets in the GCMD. High participation.
 - r. Neville Smith – Participated very actively in the NVOADS National Meetings. Installed a server for some of their data holdings. Moderate-to-low participation.
 - s. Ken Tenore – Participated in one of the annual meetings. Very low participation.
3. **Partner Long-Term Commitment:** Co-Is with which we anticipate a close continuing relationship in the future are: W. Nowlin, Texas A&M; B. Butman, USGS; R. Raskin for Don Collins, JPL DAAC; P. Sharfstein for Jim Cummings, FNMOC; B. Domenico, Unidata; T. Habermann, NGDC; P. Hamilton, SAIC; S. Hankin, PMEL; L. Olsen, GCMD and, S. Kempler for Chris Lynnes, Goddard DAAC. These colleagues continue to serve their data, add new data to their sites or develop software for the system as well as to participate in on-going planning efforts and in the submission of joint proposals. In addition to these project Co-Is, there is an equal number of groups/individuals who were not Co-Is, but who are participating at a similar level.
 4. **Success in Project Objectives:** All of the objectives identified in the proposal have been met. Specifically, a community-consensus infrastructure has been developed for the consistent syntactic delivery of data via the network directly into the user's analysis package; the system has been heavily populated with oceanographic and meteorological data sets; the system is being

heavily used; and, a vigorous community of data providers and software developers has been established. In addition to the tasks initially identified, a number of other tasks related to the system have been undertaken such as the pilot programs, the development of an Aggregation Server for array data, the development of an Ancillary Information Services capability and interfaces to several different GIS web-based systems. Our success in meeting these objectives is underscored by the fact that several large projects are making use of OPeNDAP (a.k.a. DODS) as a component of their data transport layer:

- a. NOMADS <http://nomads.ncdc.noaa.gov/>,
- b. DMAC/IOOS <http://dmac.ocean.us:8101/>,
- c. ESG II <https://www.earthsystemgrid.org/about/overview.jsp>;
- d. COLA <http://grads.iges.org/grads/gds/>; and,
- e. GODAE http://www.bom.gov.au/bmrc/ocean/GODAE/Meetings/7thIGST/data_sharing_paper.doc.