The Distributed Oceanographic Data System: A Framework for Access to Scientific Data in the EOS Federation

by

Peter Cornillon — University of Rhode Island
Mark Abbott — Oregon State University
Elaine Dobinson — Jet Propulsion Laboratory
Robert Evans — University of Miami
Glenn Flierl — Massachusetts Institute of Technology
Michael Folk — University of Illinois
Peter Fox — National Center for Atmospheric Research
David Fulker — University Corporation for Atmospheric Research
Steven Hankin — Pacific Marine Environmental Laboratory of NOAA
Judy Holoviak — American Geophysical Union
Lola Olsen — Goddard Space Flight Center of NASA

Submitted to

Creative and Innovative Working Prototype Earth Science Information
partnerships in Support of Earth System Science
CAN-97-MTPE-01
National Aeronautics and Space Administration
400 Virginia Ave. SW
Suite 700 Washington D.C. 20024

October 22, 1998

Contents

Abstract

In response to concern in the scientific community about the potential shortcomings of the EOSDIS data distribution system, a "distributed" data system using Internet servers is proposed. Under this model, a scientist is as likely to be contributing data as using it. Not only could such a system reduce the costs of the EOS data distribution, but it would allow scientists to retrieve data from each other in much the same manner as data would be retrieved from the central EOS archives. For a user whose software requires a specific data format, the system would automatically translate data formats from the archive to suit those requirements. In addition to making the EOS data itself more easily accessible, the combination of the translation facility with the distributed nature of the data system will facilitate the sharing of data among earth scientists. The basic framework of this system is already implemented as the Distributed Oceanographic Data System.

1 Introduction

MTPE-97-CAN-01 and -02 are part of NASA's[1] response to concern within the scientific community about EOSDIS. This concern is based on two quite different issues: (1) the lack of a sense of "ownership" within the scientific community of the production of EOS data sets and of the whole data distribution portion of EOSDIS, and (2) what most scientists perceive as the high cost of EOSDIS—especially given its apparent functionality.

Our collaboration (Section 6) of 11 non-profit institutions proposes to address these issues by integrating the Distributed Oceanographic Data System (DODS) into the EOS data distribution framework. This will provide for significant enhancements to data delivery over what is currently planned for EOSDIS. In addition, DODS will accommodate the seamless addition of earth science data sets not currently in the EOSDIS repertoire.

The proposed effort will explore how the availability and distribution of data affect each stage of the research process. The specific tasks proposed may be divided into three groups: (1) the addition of data sets to the proposed data system, (2) improving and enhancing the software of the system itself, and (3) developing community interactions. These tasks are discussed in detail in Section 2. Although the focus of this work will be on the oceanographic and meteorological communities, we also intend to explore DODS as a mechanism to achieve the same ends for other earth science disciplines.

DODS is a broad-based effort, now three and a half years old, to develop a distributed data system for earth science data. One of the fundamental design criteria of this system is that each research scientist be viewed as a potential data provider as well as a data user. Partly because of this central tenet, and partly because this development effort has been initiated and carried out within the oceanographic community itself, DODS by its very nature, addresses the first of the issues raised above.

In addition, we expect that moving to a distributed data system such as DODS will result in a long term cost savings for EOSDIS because it makes use of distributed resources within the scientific community both to produce and distribute data, reducing the need for large personnel, networking and CPU resources at the relatively small number of centralized sites in EOSDIS.

[1] See Section 11 for a list of acronyms.

1.1 The Problem

Although national archives are important sources of research data, they are not the only sources: a large fraction of data used in research is obtained by scientists directly from their colleagues. These data might be reprocessed versions of data held in federal archives, such as a climatology generated from data archived at the National Climate Data Center, or field samples measured on an oceanographic cruise. Both of these data sets might eventually make their way into a federal archive, but the lag time can be long and a significant number of data sets are never archived. This is especially true of data sets produced from data that already reside in a federal archive, such as the climatology mentioned above. However, it is often these data that are of the most interest to others in the scientific community, there being significant value added to the original data set in such cases.

Since the initiation of the EOS effort, it has become clear that there will be a number of data products generated by independent researchers that will not be part of the official EOS suite of products, but will be of interest to a broader audience than the researchers who create them. Unfortunately, EOSDIS will not provide access to these "orphaned" data sets.

Concern about the cost of EOSDIS has arisen in part because all that most scientists see is the data delivery portion of the system, the part that will deliver data to them. In fact, this is less than a third of the project's total cost. The rest is associated with the acquisition and preliminary processing of the primary—and admittedly very substantial—data stream. The difficulty is that the data delivery portion of the system has been intimately coupled with the data acquisition, making it conceptually difficult to separate the two. Part of the purpose of MTPE-97-CAN-01 is in fact to achieve this separation.[2]

1.2 The Genesis and Current Status of DODS

Three and a half years ago, a workshop[3] consisting of oceanographers, computer scientists and oceanographic data archivists met to discuss and design a data system that would facilitate the exchange of data over the Internet between research scientists, federal archives, and private organizations. This group in effect was addressing many of the issues that MTPE-97-CAN-01 has been created to study, specifically the data distribution issue raised in the previous section. The workshop and the resulting development effort produced an architectural framework for a distributed data system designed with two primary criteria:

- The scientist is a data provider as well as a data producer, and
- The scientist should be able to use the software package with which he or she is the most familiar to analyze data available via the system.

The first of these criteria requires that serving data must be easy. That is, installation and maintenance must be straightforward, and the overall impact on the scientist must be low. The second point requires both that the system be easily interfaced to *existing* application software, either commercially produced, free-ware or the researcher's own programs, and that data be provided in the form required by that analysis software.

DODS is the system that was designed to satisfy these criteria. Implementation of a basic working version of this system has been completed with Version 2.13 of the DODS software now available (ftp://dods.gso.uri.edu/pub/DODS/) to the community. The system

[2]Of course, it is important that the link between the two not be completely severed or moving data from the front end of the system to its data distribution portion will become difficult or impossible.

[3]http://dods.gso.uri.edu/DODS/reports/workshop1/workshop1.html

is described in detail in a number of documents on the DODS Web site,[4] and is summarized in the following section.

The Application Program Interfaces (APIs) for the NETwork Common Data Format (NetCDF) and the Joint Global Ocean Flux Experiment (JGOFS) are fully supported in DODS on both the client and server side of the system and data servers are supported for MATLAB, DSP, HDF and FreeForm[5].

We have just been notified by NSF that our proposal[6]: (1) to populate DODS with data from oceanographic archives at Scripps (World Ocean Circulation Experiment (WOCE) hydrographic data), Oregon State University (OSU) (the west coast time series) and U. Hawaii (Acoustic Doppler Current Profiler (ADCP) data); (2) for a modest level of user support, and; (3) for some feature augmentation, will be funded.

A proposal for additional enhancements to DODS is pending with National Oceanic and Atmospheric Administration (NOAA)[7]. This proposal focuses on: (1) locating coastal data sets along the eastern seaboard of the US; (2) writing DODS servers for these data sets, and; (3) porting the DODS core software to Windows NT.

It is important at the outset to note that although DODS has been developed with oceanographic applications and data in mind, there is nothing about the design and implementation of the system that constrains its use to oceanography alone.

1.3 DODS Overview

DODS provides a way for researchers to access scientific data anywhere on the Internet using the analysis software of their choice. This is accomplished by re-linking the software packages with DODS-provided versions of the packages' data access subroutine libraries. (In many cases, to re-link an analysis package is a matter of a few minutes and requires no modification to the researcher's application source code.) The relinked program is then able to read data from any DODS-compliant server on the Internet in addition to being able to read files from a local disk. Data are read into these programs by issuing "load" or "read" commands similar to those issued in the original program but with the data set name replaced by an Internet-standard Uniform Resource Locator (URL), such as **http://dods.gso.uri.edu/cgi-bin/nph-nc/data/fnoc1.nc**. The result is that the researcher has effectively converted his or her analysis package into a powerful network client. The analysis programs can be commercially available analysis tools such as Interactive Display Language (IDL), MATLAB or PV-Wave, free-ware such as Ferret[8] or the Generic Mapping Tool (GMT), or an analysis program developed by the user.

The process of analyzing data may be split into two phases: acquisition and visualization. A scientist first retrieves some data set and then uses whatever programs are needed to put the data into a form that will illustrate the question at issue. The visualization phase is often a highly idiosyncratic process—as it should be. DODS only addresses the data acquisition phase of the process, leaving scientists free to use the analysis programs and methods that best fit their needs. The steps (which can often be quite tedious) required to acquire the data and prepare it for the analysis program are avoided; the DODS core software does this work automatically.

---

[4] http://dods.gso.uri.edu/DODS/

[5] The FreeForm Data Access System, developed at NGDC, implements a unique and powerful method for translating data from one format to another. It is based on simple data descriptions.

[6] See http://dods.gso.uri.edu/DODS/proposals/NSF/NSF.html for the text of the NSF proposal.

[7] See http://dods.gso.uri.edu/proposals/NOAA/NOAA-prop.html for the text of the NOAA proposal

[8] Ferret is a visualization and analysis application designed by oceanographers at PMEL to meet the needs of oceanographers and climate researchers analyzing large and complex gridded data sets.

In addition to providing remote access, DODS offers other important features such as subsetting, format translation and data location:

Subsetting — DODS possesses a sophisticated subsetting mechanism that reduces network traffic as well as the load on a user's system. It accomplishes this with a rich "constraint expression" syntax that allows a user to select virtually any subset from a remote (or local) data set. A user can even specify a subset with a conditional data access expression dependent on an entirely distinct data set, stored at another site.

Translation — DODS incorporates a powerful data translation facility, so that data may be stored in a format defined by the provider, but may be accessed by the user in a manner identical to the access of local data files on his or her own system. Furthermore, the translation facility can be extended to data types. If, for example, a user program cannot accommodate a certain data type (a JGOFS program, set up to display sequences, may not be able to read directly an array from a NetCDF archive), it still may be possible to read and display those data by converting ("translating") them to a data type more familiar to the program.

Data Location — To help users find data on the network, a data locator service is currently being implemented for DODS compliant data sets. Data providers may register their data sets with this service, allowing a remote user to query the data locator for the URLs of any data set satisfying given criteria. With a set of returned URLs, the user can then refine the search by querying the given DODS servers directly.

DODS also contains provisions to increase the amount of data available across the network, since the DODS architecture is designed to make it easy for scientists to become their own data providers. Data need not be sent to a central authority nor converted to a standard format in order to make it available to others. Any machine connected to the Internet can host a DODS server, which can make data available to any DODS client.

A small part of the functionality provided by DODS is already available on the Internet. Tools that exist now, such as Netscape and Java, are being used to implement data browsers and catalogs. Netscape, however, is not a data system, and Java is no more than a portable programming language. These tools do not inherently prevent two different developers from creating two completely incompatible data services. The DODS system, on the other hand, is built on a sophisticated and versatile protocol—and the functions to implement it—for data transmission and representation. This ensures that data from any DODS server can be correctly read by any DODS client.

DODS also provides an open-ended environment to scientific data users. Without DODS, for example, a data provider could include a sophisticated Java applet on their site, allowing a user to examine data. Prototypical sites like this already exist. However, the scientist looking at those data with such tools is constrained to look at the data in ways anticipated by the applet designer. If he or she wants to see the data in ways not accommodated by the site designer, with a contour map, for example, or wants to use some sophisticated curve-fitting algorithm to extrapolate data, there is no recourse. By providing *only* the data access tools and protocol, DODS allows scientists to use whatever tools they feel are appropriate to the situation, instead of being forced to adapt to the tools provided by others.

Another very significant difference between DODS and all of the other existing systems serving earth science data with which we are aware is that network accessible data may be acquired under program control. This means that a researcher interested in a given phenomena can write a program or a script in his or her analysis package that will acquire data from a DODS compliant data server, operate on these data and, depending on the result of the operation, request more data from the same site or from another DODS compliant site.

To conjure an elaborate—but not at all extraordinary—example, a script could be written that first determines how many in situ measurements of sea surface temperature were made in a given space-time window by querying the National Ocean Data Center's DODS Expendable BathyThermograph (XBT) server. If more than 10 are available, the script could query the Jet Propulsion Laboratory (JPL) Physical Oceanography – Distributed Active Archive Center (PO-DAAC) DODS server to determine if the wind speed exceeded 10 m/s in the vicinity of any of these measurements. If so it could then acquire the corresponding sea surface temperature image(s) from the University of Rhode Island (URI)'s DODS server. The resulting data triplet could be stored, the space-time window incremented and the process repeated until a space-time range initially specified by the user has been exhausted. In this example, many requests to the three different archives may be made, but only the data of interest are acquired, and the whole process can be initiated with only one command by the user.

2 Approach

The tasks presented here are aimed at developing a distributed data system that will adequately serve the needs of EOS scientists. Motivation for undertaking these tasks can be partitioned into four groups: basic system enhancements; general support; experimental efforts, and; research efforts. Note that some tasks come under more than one category.

1. *Basic System Enhancements*. These tasks are based on current technology and are extensions of tasks already accomplished in the current DODS architecture. They are doable and the rationale for undertaking them should be clear. Additional rationale is presented in the description of the task where deemed appropriate. The tasks, all of which relate to DODS are: development of a catalog server (2.2.1); providing Java clients and servers (2.2.2); enhancing format, structure, and unit translation (2.2.3); adding support for new APIs (2.2.5); establishing virtual data sets (2.2.6); providing efficient access to very large data sets (2.2.8); the incorporation of Distribute Active Archive Center (DAAC) data into DODS (2.1.2); development of an Artificial Intelligence (AI) metadata generator for DODS data sets (2.2.4), and; development of a push-pull interface (2.2.10).

2. *General Support*. These tasks are designed to facilitate the use of the system (2.3.1) and to increase awareness within the earth science community of its existence and utility (2.3.3). Again the rationale for the tasks should be clear.

3. *Experimental Efforts*. These are tasks designed to explore either new uses that might be made of DODS or how well DODS transfers to other earth science research communities. We propose three experiments all of which involve integrating DODS with other systems. In the first we will investigate how to scale DODS access in a high volume data system. This will be done by combining DODS with the Unidata "broadcast" system (2.2.10) and new "virtual" servers (2.2.6) to provide access to experimental MODIS data sets (2.1.1). In the second experiment we will investigate data fusion through the combination of DODS access with Ferret visualization/manipulation (2.2.7). This experiment will also rely on unit translation (2.2.3). In the third experiment we will investigate how well DODS works in a very different earth science community, those interested in the upper atmosphere (2.1.3). The rationale for these experiments is presented in the discussion of the primary task associated with each.

4. *Research Efforts.* These tasks address the speculative, but possibly very significant, contributions that may be made with distributed systems such as DODS. They are tasks that we would not have thought of prior to our development of DODS because of the difficulty in accessing information needed to carry them out. They include: the use of Artificial Intelligence to characterize users' needs based on their past usage of data in the system and then, using this characterization, to make recommendations about other data sets that may meet their needs (2.2.11); providing mechanisms for user feedback related to data quality (2.3.2) with the objective of increasing the overall quality of data in the system, and; evaluating the potential use of distributed data systems in the peer-reviewed publication process (2.3.4).

Because the categories presented above overlap, we have decided to organize our presentation of the details of each task differently in the following. They are divided into three subsections: (1) new data sets for DODS, (2) system enhancements, and (3) community interaction. The tasks are also summarized by performing group in Appendix B.1.

2.1 New earth science data sets to be added to DODS

Over 300 oceanographic data sets are already accessible to DODS[9]. Additional oceanographic data sets will be added with recently obtained National Science Foundation (NSF) funding. In this section we propose adding to this suite of DODS-compliant data sets: (a) MODIS ocean data products to be developed at the University of Miami (UMiami), (b) standard EOSDIS ocean data products to be archived at the JPL and the Goddard DAACs, and (c) solar/terrestrial upper atmosphere data sets held by the High Altitude Observatory (HAO) of the National Center for Atmospheric Research (NCAR).

The UMiami MODIS research data sets were chosen because they are expected to be of importance to the oceanographic community and because they represent Principal Investigator (PI) produced EOS data sets. These UMiami data sets also provide test sets with which we can address the push-pull distribution mechanism proposed in Section 2.2.10. Data sets from the two DAACs were chosen again because of their general importance to the oceanographic community and because they will allow us to experiment with interoperability issues associated with serving large EOS data sets. The upper atmosphere data were chosen for three reasons. (1) These data are of prime importance to the global change problem and there are no plans to integrate them into the EOSDIS framework. (2) They represent a very different data type from those included in the DODS repertoire, hence they offer a good opportunity to learn about extending the DODS model to other earth science data types. (3) University Corporation for Atmospheric Research (UCAR)/NCAR is currently putting together a strategic plan for the handling of its vast data archives in the future. The DODS-HAO collaboration, in addition to addressing a critical data delivery problem in HAO, would allow UCAR/NCAR to evaluate the performance of a truly distributed data system first hand.

2.1.1 MODIS-derived ocean parameters produced at UMiami

The MODIS Oceans Team is developing the "community consensus" algorithm for the MODIS ocean data products. This algorithm will be implemented at the Goddard DAAC which will serve the EOSDIS sanctioned MODIS data. The Oceans Team will continue to tune the algorithm after launch, processing the MODIS data stream in a research mode as the data are received. We propose making a subset of the research data stream available

[9]http://dods.gso.uri.edu/DODS/home/dods-list.html

to the community shortly after it is produced. The research community will then be free to choose the EOS sanctioned product or the data processed with the most recent innovations: research products without the level of scrutiny of the officially sanctioned version, but possibly of a higher quality.

These data sets will be "broadcast"[10] to OSU and a subset will be "broadcast" to URI. DODS servers will be installed at all three institutions for these data and a "virtual" DODS server pointing to the holdings of the three sites will be installed at URI. User requests for MODIS research ocean data will be made to the virtual server. This will decrease the load on any one site and allow us to experiment with the "push-pull" technology discussed in Section 2.2.10 with conditions under our control.

2.1.2 JPL PO-DAAC and Goddard Space Flight Center (GSFC) DAAC Ocean Data Sets

JPL PO-DAAC and GSFC Global Change Master Directory (GCMD) staff will work with the DODS core group[11] to install DODS HDF-EOS[12] servers at the JPL and GSFC DAACs for all EOS sanctioned ocean data sets[13] they plan to keep on line[14]. Currently there are in excess of 3 TB of level 2 and higher data on line at the PO-DAAC and this will increase by approximately 2 TB per year between 1998 and 2000. Additionally, standard DODS HDF servers will be maintained at JPL for the Pathfinder and early EOS data sets that are not in HDF-EOS. Also, JPL PO-DAAC and GSFC DAAC staff and the DODS core group will investigate enhancements in the efficiency of serving the large image data sets via DODS in the context of the data compression work being done with National Center for Supercomputing Applications (NCSA) (section 2.2.8).

2.1.3 Upper atmosphere and solar data set

The High Altitude Observatory division of NCAR investigates the sun and the earth's space environment, focusing on the physical processes that govern the sun, the interplanetary environment, and the earth's upper atmosphere. As part of its mission, HAO is a focal point for two important programs: (1) the Coupling, Energetics and Dynamics of Atmospheric Regions (CEDAR) program designed to enhance the capability of ground-based instruments to measure the upper atmosphere and to coordinate instrument and model data for the benefit of the scientific community [15] with the goal of better understanding various forcings of the earth's climate, and (2) the Radiative Inputs from Sun to Earth (RISE) program designed to address causes of variations in the Sun's radiation as a star as well as the source of radiant energy impinging on Earth. The overall aim is for an improved physical understanding of present, past, and future radiative output variations in the extreme ultraviolet, ultraviolet, visible and infrared.

[10] We use the term broadcast here to mean automatically sent to a subscriber list as they are produced.

[11] The DODS core group refers to the URI/Massachusetts Institute of Technology (MIT) collaboration that designed and implemented the DODS core software.

[12] HDF-EOS servers will be developed by JPL as part of this effort — Section 2.2.5.

[13] JPL is the designated archive for all physical oceanographic EOS data while the GSFC is the designated archive for ocean color data

[14] Over the past year, the JPL PO-DAAC has, in an experimental mode, been serving NASA SCATterometer (NSCAT) data via the DODS HDF server (http://dods.gso.uri.edu/DODS/). As part of this experiment we have found that once a DODS server is constructed installing and maintaining it requires very little additional effort.

[15] http://www.hao.ucar.edu/public/research/tiso/cedar/cedar.html

Additionally, HAO's observatory monitors the sun with a number of different instruments (the Digital Prominence Monitor, the Coronal Helium Imaging Photometer, the White Light Coronameter, the PSPT, and the Low-L helioseismology experiment). All data obtained from these instruments are returned to HAO and made available to the community for use in understanding solar and space weather, in support of space based observatories (Yohkoh, UARS and SOHO) as well as in support of basic solar-terrestrial research efforts.

The volume of the CEDAR data holdings[16], are significant and are expected to grow rapidly over the next 10 years (to $\sim$ 100GB or more). Unfortunately, however, these data are difficult to access and expensive to maintain. HAO has therefore begun (with expected completion later this year) the process of migrating the data out of the current Data Base Management System (DBMS) with the hope of providing access to them via a distributed system.

HAO believes that DODS meets most of its data distribution requirements and intends to make its CEDAR holding accessible to the community via DODS. Specifically, HAO, with DODS core group help, will build (where necessary) and install DODS servers for all of their CEDAR data holdings. Additionally, they will build a MATLAB Graphical User Interface (GUI), similar to the one already constructed by the DODS core group, and participate in testing of the IDL GUI being developed, to facilitate location and access of these data. (These data will also be registered with the GCMD (Section 2.2.4).) This will provide high functionality access to the 800+ member CEDAR community and to the broader community which we expect will find a number of the data sets (for example, measurements from Incoherent Scatter Radar, Lidar, and All Sky cameras, and synthetic data from Large Model output) to be of significant interest.

2.2 Enhancements to the DODS core system

Although DODS is a functioning system, we have learned a great deal about issues associated with networked data systems over the past year and we feel that there are significant enhancements that can be made to the system which will improve its usefulness by increasing: its functionality; the data classes that it handles; the user community that has access to it, and; the efficiency of its operation.

This section details these enhancements.

2.2.1 The Catalog Server

Current DODS data servers provide access to data at a fairly low level. In essence these servers are a generalization of the file-level organization of data sets. At present, access to DODS data sets is performed at this level, hence the user requires knowledge of the file structure associated with the data set to efficiently extract data from it. We propose the addition of catalog servers to the DODS system[17]. These servers will provide a higher level of access such as that needed for data discovery. By clustering multiple files into a single homogeneous data set, adding extra information to the data set and supporting a search mechanism tied to standardized parameters, the catalog server will provide a second, more sophisticated, access point for a given data set. DODS catalog servers will be built using existing lower level servers.

[16] http://www.hao.ucar.edu/public/research/tiso/cedar/cedar.html#CDB.

[17] A prototype catalog server for DODS have been developed at JPL (Dobinson, E. and R.G. Raskin, Earth Science Remote Access Tool, Proceedings of the Ninth International Conference on Scientific and Statistical Database Management (SSDBM), August, 1997).

Once the basic server has been built, we will extend the server design so that the separation between the higher level catalog server access and data access is eliminated. Thus, a user may choose to access data using either the high-level catalog server or the lower-level server.

The addition of the catalog server will not sacrifice the basic tenets of DODS; that providers not be required to reformat their data for inclusion in the system and that they be free to make data available without being forced to participate in every aspect of the system. Thus, the old style DODS servers will be completely supported and the new intelligent servers will function by merging ancillary information with the information contained in the data set.

2.2.2 Java class clients and servers

Porting the C++ "native" DODS core library to Java is expected to be a straight-forward task. Both are object-oriented environments. Using the new Java DODS core library we will create both client applications and data servers which users will be able to run on a wide variety of hardware platforms.

DODS will provide a Java class library for development of DODS clients in Java. This library will enable the development of DODS clients which run in World Wide Web (WWW) browsers as well as stand-alone clients. Because Java byte code software is not tied to a particular hardware platform, the number and type of DODS users will be greatly expanded.

Also, the Java library will provide a platform on which to experiment with other modes of access within DODS, including Java-CORBA and Java-RMI. Thus, the Java library will pave the way for integration with other systems (e.g., those using Common Object Request Broker Architecture (CORBA)).

2.2.3 Translation

The success of any system which hopes to achieve interoperability among many data sets rests on the ability to translate between many different data representations. DODS implements data interoperability by breaking the translation problem into three parts: format translation, structure translation, and unit translation. While the DODS 2.x design completely realizes our goals for format translation, we will continue to pursue both structure and unit translation as part of the EOS Federation.

Data Structure Translation — The DODS data model supports a group of basic data types for representing scientific information, in order to meet the expectations of users and their analysis software. Data types differ in the relative ease with which different operations can be performed, and often differ considerably in the kind of information they represent. To achieve data structure translation the software must be able to translate information stored in one data type into information stored in another with minimal information loss or corruption.

We will generalize and unify type translation in DODS by adding it as a module to the DODS core software. The translation module will use metadata available from each DODS server and the DODS Common Data Representation (CDR) object. The CDR is an overarching schema serving as an intermediate representation for each of the supported data types. The CDR, when compared to a simpler pair-wise approach, reduces the number of translation modules required for n data types from $n(n-1)$ to n. In addition, the CDR simplifies long-term maintenance of the translation software by reducing the effort needed to add support for new data types.

Data Unit Translation — Because DODS servers support uniform access paths to many data sets, it is an ideal base on which to build data fusion applications (Section 2.2.7). But, data fusion requires that users be able to translate not only data types but also the units in which data values are expressed, for example, all ocean currents might be expressed in meters per second.

DODS currently does not support unit information. We will add units to the data transfer software. Thus, data types and their units will be translated. Unidata has addressed this complex (and probably unsolvable in a general sense) problem with considerable success in its UDUnits (http://www.unidata.UCAR.edu/packages/udunits/index.html) software package. We propose using UDUnits as the basis for the DODS units translation software.

2.2.4 Mining DODS compliant data sets for directory information

We propose to document all DODS-compliant data sets registered at the GCMD to the Directory Interchange Format (DIF) standard[18]. This effort will consist of two parts: manual documentation and the development of an automated population mechanism. The first objective will be met with a full time "DIF coordinator" who will work at the GCMD. Documentation of DODS-compliant data sets will begin with registration by the data provider via a Web-based registration form. The registration process will obtain some information from the provider and some from the data set. The "DIF coordinator" will then supply missing information either by querying the data set via DODS or, if need be, asking the data provider.

After the development of a serviceable directory at the GCMD we will investigate techniques for using the metadata available with each data set in the Data Descriptor Structure (DDS) or the Data set Attribute Structure (DAS) to automatically generate the DIF.

Particularly difficult in most metadata integration strategies is attribute/entity matching. To accomplish this we will develop an AI system that will operate as part of the registration process. Programs at the GCMD will automatically interact with the data provider during this process to validate the results of the AI system and will request additional information of the data provider if the AI system needs help.

When this system has been successfully installed we will extend the process to a technique that does not assume particular access patterns, thus permitting later data mining through unexpected paths and/or views of the data. For example, it might be determined that spatial and temporal data density are important search parameters not originally included in the metadata population effort. The extension of the AI system would provide for a means of obtaining this information for all, or a subset of, DODS data sets registered at the GCMD.

2.2.5 Addition of New API's and Continued Support for Existing API's

One of the design tenants of DODS has been to approach data formats from the APIs used to access them. With this approach, all data sets using a particular format become DODS-compliant by simply writing the DODS servers for the associated API. At present, we support: MATLAB-4, NetCDF, HDF, FreeForm, DSP and JGOFS[19]. To this suite we propose adding GRid In Binary (GRIB) and Binary Universal Form for the Representation of meteorological data (BUFR). This will provide DODS-compliant capability for a very large number of meteorological data sets. In addition we propose upgrading our MATLAB, NetCDF, and HDF to include enhanced versions of these APIs.

[18] A DIF to FGDC translator has been written by GCMD staff, hence DIF-compliant data will also be rendered FGDC-compliant by the GCMD.

[19] FreeForm and JGOFS are ideally suited as mechanisms to serve simple PI-held flat file data sets via DODS.

DODS currently provides a high level of support for MATLAB-4; both a data server and a graphical interface client. MATLAB-5 provides a much improved base for processing and storing earth-science data. Among other enhancements, MATLAB-5 provides for multi-dimensional arrays; critical to meteorological and oceanographic research. We therefore propose extending our MATLAB-4 support to MATLAB-5.

NetCDF and HDF are also very important software systems for the earth-sciences. These systems are used to store self-describing data, which greatly facilitates serving data via DODS. We will add support for NetCDF 3.x and NetCDF 4.x to the existing NetCDF client and server libraries. We also propose to add support for HDF-EOS and the emerging HDF-5 system to the existing HDF server. The DODS-HDF server currently supports translation of HDF's SDS, Vdata, and raster data types. We will add server translation for the point, swath, and grid data types, using the HDF-EOS API. An HDF-EOS client-library will also be built.

2.2.6 Virtual data sets

DODS makes use of the hypertext transport protocol (http) to move data from the provider to the user. Data are requested from a DODS server via a URL which is used to characterize the subset of data desired. In DODS, a synthesized data set could be a collection of URLs, the actual collection of data spanning a number of data archives. We anticipate that such virtual data sets will assume an increasingly important role as data are gathered from open ended data sets such as those associated with EOS. DODS access to MODIS research data products (Section 2.1) is one example of a virtual data set. We propose adding to DODS the capability to construct "virtual" data servers. This will initially be done in the context of MODIS ocean data research products (Section 2.1.1).

2.2.7 Data fusion of DODS compliant data sets with Ferret

DODS is uniquely well suited to data fusion – the joining of data from distributed locations into merged visualization and analysis (V&A) products. With DODS an existing desktop V&A application can become a data fusion environment simply by relinking the application with the DODS software.

The WWW and DODS acting in concert have the potential to break through the barriers that have traditionally been associated with distributed data location and diverse file formats. A Web site that can provide "live" V&A can be extended to provide live comparisons (over-plotting and differencing) between distributed data sets. We propose developing such a Web server with a focus on the needs of gridded data sets (especially for modelers and climatologists).

The system we propose uses the Ferret[20]-based Live Access Server (LAS). LAS provides the user with on-the-fly visualizations, subsetting, and delivery of data files in a user-selectable choice of file formats. Ferret is tailored for data fusion providing for over-plotting, rescaling, and the types of regridding needed to take differences of data sets with differing spatial and temporal resolutions. The LAS Web server, accessible at http://www.ferret.noaa.gov, is widely used within NOAA and elsewhere, hence clearly provides a strong foundation upon which to build the proposed Web data fusion server.

The DODS-WWW-Ferret system will tie a community of closely affiliated data suppliers into a single virtual data base with a Web server interface offering visualization, subsetting, downloading and fusing. This system will be ideally suited to collaborative efforts in which the members are both suppliers and users of data, making it an ideal tool for scientific

[20]Ferret has been successfully relinked and tested with DODS

research. Additionally, it provides us with another platform on which to experiment with DODS extensibility.

2.2.8 Storing and Accessing Long Time Series of Large Data Arrays

DODS allows access down to the single datum. This provides the user with enormous flexibility in requesting data, but for some of the larger data sets to be served, efficient access can become a significant problem because of the number and size of the data files. We discuss some of the issues involved as well as the approach that we plan to take to address these issues in the context of a problem that we currently face.

With NASA funds we (URI) are applying the community consensus NOAA/NASA Pathfinder Sea Surface Temperature (SST) Algorithm to all of the full resolution Advanced Very High Resolution Radiometer (AVHRR)/2 data in the URI/UMiami archive (June 1981 to present). The resulting data set will consist of more than 15,000 SST fields, one per satellite pass, with each field comprising approximately 8 million pixels. The data are processed and saved in satellite coordinates. However, most users either want images in some standard earth projection or they want a long time series of sea surface temperature values at a given location.

The problem that arises is that the processed data set is too large for us to maintain on-line uncompressed, ≈ 256 GB. However, once cloud covered pixels have been set to zero, the data set will compress to ≈ 60 GB and we can provide access to this data set on-line via DODS. But this means that for each request of data from a particular satellite pass we must decompress a file. If we have stored the data as compressed images, it is straightforward to meet such a request. For those requesting a time series, this quickly becomes impractical. For example, to provide a 15 year time series of SST values at one location would require decompressing 60 GB to 256 GB to deliver 30 KB of data.

To address this problem we propose to examine the use of the HDF format for storing the data. HDF already supports compressed data, and the University of Illinois' HDF project has recently added support in HDF for array "chunking", whereby large arrays are split into smaller chunks, which can then be individually compressed and stored in HDF files. A request to retrieve data results in some subset of the smaller fields being decompressed and reassembled into a sub-image containing the area (or point) of interest.

Additionally, we will examine the use of HDF-EOS for storing the data. HDF-EOS is a data model that has been developed for EOSDIS to support storage of earth-science data types in HDF. One such data type is a "grid" structure. There is an HDF-EOS API and associated library that supports compression and chunking. The combination of projection information and compressed chunks might provide a very effective data storage scheme for DODS.

Testing of the efficiency achieved by "chunking" will be performed on storage and retrieval of data from the very large image archives at URI, UMiami and OSU. Projections related to demand will be made and then the CPU time and disk storage requirements will be estimated as well as the resulting savings that might be obtained by adopting this method across the EOSDIS data product line.

2.2.9 Data Redundancy and Overlap

When combining data sets, the DODS catalog servers may encounter the problem of data redundancy. The problem is not that data redundancy is bad, but that when two or more servers return overlapping values the catalog server must arbitrate between them. This problem is a complex one; it will not be limited to simple cases where values are all identical, but rather will include situations where values for the same "range" are slightly different.

We believe that a solution to this problem goes beyond the current effort, hence all that we propose here is to quantify the problem in the context of the "virtual" servers introduced in sections 2.1.1 and 2.2.7.

2.2.10 Coupling DODS Access to a Broadcast Based Data Delivery System

When the demand for data sets is great, the servers and networks that access these data, are likely to become congested. Such congestion will be greatly exacerbated if users' requests tend to cluster, such as immediately following the first availability of the data set. Server-side congestion is further increased with the enhanced functionality of DODS compared with the simpler "order" based mode of EOSDIS. This is likely to be a significant problem for some of the expected high use EOS data sets such as those derived from MODIS.

The meteorological community, through Unidata, has already dealt with this problem as it relates to accessing near-real-time weather data, where the clustering of requests is particularly severe, via a broadcast system known as Internet Data Distribution (IDD). IDD employs a subscription mechanism by which users express their interest in receiving classes of products as soon as they are available. A few users, following Unidata criteria for "top-level relays", subscribe at the source, other users register their subscriptions with these relays, and so forth, until all subscriptions are satisfied.

The Unidata IDD represents true distributed computing, in which a large number of data deliveries occur without imposing an extraordinary load on any single server or link in the system. IDD applied to EOS data will foster a federation-oriented esprit de corp, because data users and providers cooperate as a community to achieve prompt and reliable delivery, as has been shown in Unidata.

We propose linking IDD (push) with DODS (pull) to substantially increase the capability of the system to satisfy user requests for data. Our demonstration will entail the definition of one or more regularly generated "data streams" that are of wide interest[21]. Following the Unidata model, we will engage a set of users (who desire immediate data access) in a push-based system of data relays. Each participant will embed the pushed data into a DODS-compliant server. These "mirror" servers are updated automatically whenever new products become available via IDD. They will be augmented to notify a DODS "virtual" server for the data set. The virtual server will keep track of data set status at each of the "mirror" sites. Users' data requests are made to the "virtual" site. The request triggers an analysis to determine which sites hold the data requested and the current burden (both communication and CPU) of each these sites. The request is then passed on to the least burdened site which returns the requested data to the user. Thus, we will be combining two operational systems to efficiently meet the needs of the primary users of either system. As an aside we note that this system also provides for some redundancy in the overall system.

We will evaluate the performance of this system in the context of general access to EOS data. Specifically, what cost savings might be possible to EOSDIS or to the WP-Federation by adopting this approach with a large fraction of the standard EOSDIS products? This addresses the question of the scalability of the system.

2.2.11 Data Location and Usage Helpers

Finding and using data in a large distributed system is complicated because users have a difficult time obtaining adequate results from conventional search tools; important data often

[21] We anticipate beginning with the research ocean product data sets to be developed at the UMiami from the MODIS data stream, with OSU acting as the mirror site, but do not want to constrain ourselves to these data sets only.

remain undiscovered and superfluous data are often returned. We will build an AI system to address this problem.

Each data request made of a DODS server contains information about data usage. It identifies the target data set, the desired subset of this data set and the data requester. We propose capturing this information and making use of it to design the AI based system which will inform users of other data sets that might meet their needs. This determination will be made based on the requests that they are currently making. To capture the requests, DODS servers will be augmented with software that will echo the request to a computer at URI. Initially this data stream will be used to design the AI system. Once built, the same data stream will be used to determine a given user's interests, to extrapolate them to other data sets and ultimately to inform the user of these other options. The AI system that we propose will be based on data warehousing/mining techniques[22]

Additionally, the request data stream will be used in evaluating system performance (Section 4).

2.3 Community interaction

In this section we address issues related to the interaction between DODS and the scientific community.

2.3.1 User Support

Essential to the success of DODS is adequate support for the installation of clients and servers and for use of the system. User support will be addressed as part of the NSF funding recently received by the DODS core group (section 1.2). However, the support services funded by NSF will cover neither the additional functionality proposed for DODS here nor the new user communities to be added.

Additional funding is requested in this proposal to cover the integration of DODS into the WP-Federation framework. The support services requested will be provided by Unidata using a user support model that they have developed for the meteorological community over the past decade. In this model one member of the support team acts as the principal point of contact for a given software package or system. This individual acts as a filter to the user community, answering the obvious questions immediately or passing the question to the individual in the group with the most familiarity with the associated software. Rapid feedback (same-day turnaround) to users is a cornerstone of this user support model.

To maximize their familiarity with the software employed in DODS, the user support staff will work with DODS core group programmers to enhance or correct the software. To minimize their need to write new responses to every question received, the user support staff will also create a searchable record of their transactions with users. Most of this record will be externally accessible through the Web, augmenting DODS documentation.

DODS user support functions will also encompass user feedback mechanisms. A DODS Users Committee will be appointed by the DODS-MTPE Steering Committee (section 5) and will meet twice per year. Finally, a list server will be maintained to support conversation among DODS users and, as appropriate, subgroups thereof having special interests.

2.3.2 User Feedback as Relates to Data Quality

The increased use associated with facilitated access to scientific data results in de facto quality control of the data. Each user of the data will, in their own way, evaluate its quality

[22]Data Mining, Communications of the ACM, 39 (11), 1996.

(this is the scientific way) and these user evaluations are often the most probing that the data sets will see. This is one of the primary arguments offered by the Committee on Geophysical and Environmental Data of the National Research Council against the commercialization of data. They argued that commercialization of data would reduce access them, hence their use and as a result their quality. To render this level of quality control effective however requires mechanisms to record the findings of users and to make this information available to future users. We propose to investigate such mechanisms as part of the user services development by Unidata. Specifically, a Web-based input mechanism will be developed, maintained and moderated by DODS user services allowing DODS data users to comment on the quality or other characteristics of the data they use. These comments will become part of the "metadata" associated with data sets and they will be linked to the data sets via the associated DODS server.

2.3.3 Stimulating community use of DODS

The primary objective of scientific organizations such as the American Geophysical Union (AGU) is to facilitate communication between its members. Such communication includes publications, meetings and the one-on-one exchanges of data and text. Communication is being improved by better access to the internet, although primarily as improved exchange of text. However, improvements in network based *access to data* can play an even more significant role in scientific communication. The AGU is therefore extremely interested in investigating the implications of network based data exchange in geophysical research as well as facilitating its use. To this end the AGU proposes to work as part of this collaboration. Specifically, the AGU can reach a very broad scientific community to discuss the merits of DODS[23], it can work with the DODS core group to obtain feedback on design and use of DODS, and it can help to educate users. The AGU will contribute toward these goals through its scientific meetings and its publications. The AGU will provide one free standard booth at the Spring, Fall, and Ocean Sciences meetings for the 3 years of collaboration. These booths will provide DODS with direct contact with scientists at the meetings. The AGU is also willing to provide news coverage in EOS, the AGU Transactions, subject to their standard editorial policies. (See the AGU letter in Appendix A.)

2.3.4 DODS and the Publication Process

We anticipate that network-based distributed data systems will play a role in the actual publication process in the future. The AGU's Publications Committee has already developed some general polices concerning the publication of data and the necessity of long-term availability. The AGU currently provides an electronic data depository for small data sets related to published articles and it anticipates an expanding interest in augmenting the formal literature with citable data in multiple formats. Simply put, the AGU would like to explore how data accessibility of this kind fits into the evolution of electronic publications. To address these issues we propose the following three tasks:

1. *Exploring DODS access to the AGU electronic data depository* — In this task we will investigate the installation of a DODS server at the depository. Not only would this put these "published" data sets into the DODS repertoire but it would also allow the AGU to gauge interest in such on-line data sets.

2. *Investigating the use of DODS for format conversion* — One of the issues to which the AGU must be sensitive is long term maintenance of its data depository. The AGU

[23] The AGU membership exceeds 35,000 earth scientists in 115+ countries, with 10,000+ ocean affiliates.

would like to explore the use of a system like DODS to perform the appropriate format translations required when migrating data to new storage systems.

3. *Linking DODS delivery of data to on-line delivery of electronic journals* — At present, AGU's on-line journals provide figures as gif images. The AGU is considering an alternative to provide the data in place of the figure. A Java-applet that comes with the on-line manuscript could then generate the plot when requested by the user. With additional software in the applet the reader could actually manipulate the data while reading the article; DODS Java clients could be enhanced to work with the AGU on-line manuscripts achieving the functionality introduced above.

Although, the AGU is interested in these issues for all of its members, in this effort the focus will be its ocean affiliates.

3 The WP-Federation

The DODS collaboration believes that the WP-Federation is an appropriate way to address what we perceive to be the scientific community's concerns with regard to EOSDIS. Furthermore, we believe that the underlying notion of the federation will benefit the DODS constituency. As should be apparent by now, the fundamental design criteria of DODS are based on a view of the scientific community as a federation: from the DODS perspective each research scientist is a potential provider of data as well as a potential user. Through our collaboration with other WP-Federation members we will substantially increase the data available via DODS. By providing bridges to other systems we will substantially increase the number of earth scientists who have access to DODS data. We see the earth sciences community benefiting in both cases.

We believe that a good starting point for the development of the federation will be to carefully examine large scientific collaborations that work. To this end we believe that we have a lot to offer:

1. Run by UCAR, the Unidata effort now helps 120+ universities to acquire and analyze atmospheric data. Unidata is managed by the Unidata Program Center which obtains "its direction from the common needs and interests of its users via a users committee, and occasional surveys of the community".

2. The AGU, a 75+ year old scientific society with more than 35,000 members in over 115 countries, lists among its objectives: to promote cooperation among scientific organizations involved in geophysics and related disciplines, and to advance the various geophysical disciplines through scientific discussion, publication, and dissemination of information; objectives that in many ways are synonymous with those that we envision for the WP-Federation.

3. HAO/NCAR with its 800+ member CEDAR community interested in earth's upper atmosphere and ~100 member RISE community interested in solar variability are examples of demanding users with a broad range of computer expertise and patience with accessing remote datasets. These users continue to help shape the direction of data needs through steering committees and working groups dedicated to the data products.

3.1 WP-Federation Interoperability

We define 'interoperability' to be the capability of systems to exchange information without human intervention. We address the issue of interoperability between DODS and the systems listed in the CAN in the context of this definition.

3.1.1 ESIP Interoperability

The DODS project addressed interoperability at its outset by specifying a flexible transaction-oriented client-server system based on an extensible set of data objects. The DODS Data Access Protocol (DAP) defines a simple system that is used to move information from DODS servers to DODS client programs. The DAP uses http to move the information on the network, and so builds on the extant Web. DODS can achieve interoperability with other ESIPs through its use of http.

However, planning for interoperability using Z39.50 and CORBA is also important given the current state of software. DODS can use commercially available gateways to achieve interoperability with both CORBA- and Z39.50-based systems. Alternatively, DODS servers can support other protocols (e.g., CORBA) in parallel with the DAP. Both of these solutions are greatly facilitated by the modular, transaction- and object-oriented design of DODS.

Because the DODS data transfer protocol is very simple, and does not rely on shared state between the client and server, it is straightforward to build proxies and/or gateways between DODS servers and other systems. This is true regardless of whether the target systems also use a stateless or stateful transaction protocol.

3.1.2 EOSDIS and CEOS/CIP Interoperability

We do not propose full interoperability between DODS and EOSDIS; we will not provide access through EOSDIS for DODS users because EOSDIS does not provide the full data retrieval and subsetting capabilities of DODS. However, we will work with EOSDIS to provide access to DODS data via EOSDIS. There are two requirements that must be met to achieve this: the data must be found using EOSDIS (JEST) and it must be delivered in a form compliant with EOSDIS. The data location requirement is readily met using the DODS DIF-based metadata (Section 2.2.4). If the EOSDIS client requires full HDF-EOS we may not be able to meet the data delivery requirement in all cases since it is unlikely that we will be able to provide the full suite of metadata implied by HDF-EOS for all DODS-compliant data sets. We will however meet this requirement in some cases and we can meet a requirement for HDF in all cases with the HDF-EOS DODS client-library (Section 2.2.5).

DODS will be interoperable with the Committee on Earth Observation Satellites (CEOS) Catalog Interoperability Protocol (CIP) as viewed from the CEOS side to the extent that DODS is interoperable with EOSDIS from the EOSDIS side, since CEOS CIP and EOSDIS will to be interoperable.

3.1.3 FGDC Compliance

The DODS DAP contains provisions (currently implemented, but not necessarily used, by all DODS servers) for augmenting datasets accessible by DODS servers with metadata without actually modifying the data themselves. This mechanism is powerful enough to maintain several parallel sets of metadata so that DODS datasets can be mapped to more than one metadata standard. We will, using our data set registration system currently under development with the GCMD, use the DODS ancillary DAS object to add Federal Geographic Data Committee (FGDC)-compliant metadata to all registered DODS datasets. (Section 2.2.4).

4 Metrics

The first order metrics that measure the DODS contribution to the WP-Federation must address the degree to which DODS is used by the community for which it was designed and the degree to which it is used for the tasks for which it was designed. Therefore, these metrics must address the fact: (1) that DODS has *not* been designed to acquire very large volume, gigabyte+, data sets over the network, but rather to obtain subsets of these and smaller data sets, and; (2) that DODS has been/will be populated with oceanographic and high altitude atmospheric data sets; i.e., the target user communities are the oceanographic researcher and the high altitude atmospheric scientist. Clearly others will access data via DODS, but for the work proposed herein, the primary effort to advertise the existence and utility of the system will target the oceanographic and, to a lesser extent, the HAO communities.

To address these two issues we propose the following first order metrics:

Metric 1: *The fraction of the oceanographic community desiring subsets of ocean data sets at the PO- and Goddard DAACs that make use of DODS to obtain these subsets.*

Metric 2: *The fraction of the HAO community acquiring CEDAR data that make use of DODS to do so.*

These are easily obtained statistics and measure in some ways the perceived value of DODS. We are however concerned that easily determined quantities such as these or others of a similar nature (volume of data delivered, degree of documentation or number of directory entries in a high level directory) may not necessarily be the appropriate parameters with which to measure the effectiveness of a system. By defining such measures we may in fact be contributing to inefficiencies in the system. It might help to present a simple example.

> Recently, the PL, pursuing a research effort related to wind forcing of first mode baroclinic Rossby waves, required 3 years of numerically generated wind data of the North Atlantic. He quickly located these data on the Web and initiated an ftp session to acquire them. In the middle of the first file he ran out of disk space. (Yes, he should have probably checked the size before ftping the files, but he didn't and this is the sort of frustration faced by users all the time.) This particular data set is stored in 100+ MB files each containing one year of global data. The story continues... the files were not in a format with which he was familiar... In the end he spent a couple of days to obtain the 18 MB of data that he needed, from the three 100 MB files available from the data archive.

In terms of volume delivered this was a success. In terms of expenditure of time, use of CPU, disk and network resources this was a failure. Far from being atypical, such scenarios are common. Metrics devised to evaluate system performance must address these issues.

The above is an example of problems that we encounter in defining metrics in the system as it is today. Our work with DODS over the past several years suggests that the difficulty of defining metrics to evaluate system performance will be further exacerbated as we increase our use of highly functional distributed systems. For example, a user might make a data request to the DODS server at the Goddard DAAC for MODIS data covering the Gulf of Mexico on 1 July 1999. One of the constraints on this request might be that the wind speed, available via a DODS server at NDBC, at any given mooring in the Gulf exceeded 10 m/s. If this constraint is not met, no data are returned. In this example the DODS server at the Goddard DAAC would use information, embedded in the request, to access the NDBC. The data server at NDBC evaluates the request and returns a yes or no answer regarding wind speed at the moorings. *No data were delivered to the user from NDBC, and yet data were used.*; i.e., the system performed very efficiently.

The conclusion that we reach from these examples is that to properly evaluate system performance will require a significant amount of personnel time. The metrics will have to be based on quantities that can, for the most part, only be determined by careful interaction with the user communities. We base the following set of metrics on the assumption that: optimal system performance from the user's perspective requires that the system delivers *only the data needed* in a *timely fashion*, in the *format required* by the user's analysis package, all with a *minimal expenditure of the user's time.* We further constrain these metrics to apply only to the research scientist's perspective.

In the following sections (4.1, 4.2, 4.3) we include as a quote the description of the metrics presented in the CAN before each set of proposed measures. In Section 4.4 we introduce a new category of metrics.

4.1 WP-Federation Metrics

> Identify and enumerate existing obstacles to making data easier to find, access and use and develop metrics to measure removal of obstacles. *Goal: to measure success of the Federation in improving data access and use.*
>
> Identify user community types; identify services required by community, enumerate those services and develop metrics to measure total services provided to the community. *Goal: to measure success of Federation in providing services required by a large diverse user community.*

The community types associated with DODS are: research user; data provider and; ESIP-3 users. Metrics 3–7 focus on the needs of the research scientist and metric 8 measures the system's usefulness from the perspective of the data provider. To gather information for metrics 3 and 5 to 8 our "DIF-coordinator" working at the GCMD will survey a group of randomly selected users. To gather the data needed for metric 4, we will use tracking software to provide a quantitative measure of access times and delays (see Section 2.2.4). Additionally we will couple these requests to data discovery statistics at the GCMD.

Each of these metrics can be used to measure the performance of DODS as well as other WP-Federation contributors and EOSDIS. We will determine them for DODS.

We do not feel comfortable describing the needs of ESIP-3 users; therefore we do not define metrics for measuring them. However, we will work with other WP-ESIPs to discover these needs and implement ways to measure the system's impact on those requirements.

Metric 3: *The fraction of users who obtained data they sought.* (Goal: To determine if the users are finding the data they want, and once found obtaining it.)

Metric 4: *Elapsed times from the initial search the formal data request to receipt of data.* (Goal: To determine if the user obtained data in a timely fashion.)

Metric 5: *The fraction of data delivered that is used.* (Goal: To determine if the user is able to determine what s/he wants from a data set at the time that it is requested and, if so, can s/he request the desired subset?)

Metric 6: *The fraction of the data acquired that were used in the format in which they were delivered.* (Goal: To determine how efficient the data system is at format translation.)

Metric 7: *The amount of time required to determine how to use the data beyond that associated with format conversion.* (Goal: To determine if sufficient metadata accompanied the data for them to be used "out of the box".)

Metric 8: *Does the system adequately support data provider's needs; different versions of data sets, updates to metadata, etc.* (Goal: To determine how easy it is to install and maintain DODS servers.)

4.2 Earth System Science

> Develop metrics for usefulness of new products to specific user community, and speed of responsiveness to user community needs for introduction or improvement to new products. Develop metrics for usability of new products, including documentation and metadata. *Goal: To offer improved products and services to Earth Science User community.*

As part of this proposal (Section 2.3.1) we propose a tracking system for user feedback related to the products as well as the use of the system. This feedback will be analyzed by the user support group to determine which data sets are of value and to whom. Also, as part of the research to enhance data location in the system (Section2.2.4), we will track every access to the system. This data base will provide additional information with regard to use.

Metric 9 *Quantify usage statistics to determine usefulness of data provided via DODS.*

4.3 Technology Development Metrics

> Develop metrics to measure use of *existing* systems by end-users (i.e., science user, programmer, operations staff). Same metrics will be used to compare with new implementation approach of service and existing EOSDIS services. *Goal: To prove equal or better functionality.*

This has been addressed above.

> Define metrics showing scalability of WP-ESIP's solution to the size and scale of EOSDIS in terms of function and cost. *Goal: To assess WP-ESIP technology prototype approach relevance and applicability to a full scale Environmental Information Economy.*

Metric 10: *User services full time equivalents (FTEs) per data request to the system as a function of the number of requests.* (Goal: To determine the scalability of user services.)

Metric 11: *Access time (request to delivery) as a function of the number of requests made. These statistics will be stratified by server, data type, degree of subsetting, volume of data returned, etc. This metric will also be addressed in the context of the push-pull experiment (Section 2.2.10) since it may be possible to accommodate an increase in load via the IDD-DODS coupling.* (Goal: To determine how well the data delivery portion of the system scales.)

4.4 Data Standards

We propose an additional category of metrics:

> Define metrics for analysis of data standards. *Goal: To determine how effective data standards are in the context of the WP-Federation.*

Metric 12: *Rate data standards supported by the federated system for their overall effectiveness in achieving interoperability and for their cost of implementation and maintenance.*
We will make these estimate for DODS. The question of effectiveness will be determined subjectively; the questions of cost will be determined objectively.

5 Management Approach

This proposal involves a collaboration between 11 non-profit institutions lead by the University of Rhode Island. The management structure at URI will consist of Peter Cornillon, the Project Leader, providing the overall direction on a day-to-day basis. Working with Cornillon will be a Project Manager, hired to maintain communication between the various elements of this project as well as with other DODS funded work, and James Gallagher, the primary architect of the system, who will continue in his role as Technical Lead for DODS. Cornillon, with Glenn Flierl of MIT will also provide day-to-day scientific input to the development effort.

One or more systems or scientific programmers from each participating institution will form the DODS Development Group. This group will meet a minimum of 3 times per year and will be lead by Gallagher. The purpose of these meetings will be to synchronize new releases of DODS as well as the general development effort. Such meetings have been found to work well in the past. Gallagher will also continue to maintain the DODS Developer's Bulletin Board.

Management and scientific input from a broader perspective will be obtained from the DODS-MTPE Steering Committee which will consist of the proposal's Project Members (PMs) and will be chaired by Cornillon. This group will meet annually at a minimum, but may also be called to a special meetings prior to and/or following the establishment of the WP-Federation if deemed appropriate.

Project Member (PM)'s associated with this effort will be encouraged to attend the convention in which the WP-Federation is formed. The effort will be represented on the WP-Federation by Cornillon, if limited to one representative. If more are permitted, additional members will be chosen from the PM's by a vote of the DODS-MTPE Steering Committee.

Articles of Collaboration will be drafted prior to award if we are selected. These articles will be based on the management plan outlined above and the milestone schedule presented in Section 9.

6 Project Leader's and Project Members' Vitae

Peter Cornillon — University of Rhode Island

Education: Cornell University, Ph.D., 1973, Experimental High Energy Physics
Cornell University, B.S., 1969, Engineering Physics
Employment: 1990–Present: Professor of Oceanography, University of Rhode Island.
Inter/National Committees: TOPEX Science Team; NSCAT Science Definition Team; GCMD Advisory Panel; GCOS Data Management Panel
Professional Community Service: Editor Journal Geophysical Research - Oceans 1994-present
Publications: Dr. Cornillon has 50+ refereed publications. Recent ones include:

Polito, P. and P. Cornillon. "Long Baroclinic Rossby Waves Detected by the Topex/Poseidon Altimeter." J. Geophys. Res., 102, 3215-3235, 1997.

Everson, R., P. Cornillon, et al. "An Empirical Eigenfunction Analysis of Sea Surface Temperatures in the Western North Atlantic." J. Phys. Oceangr., 27, No. 3, 468-479, 1997.

Cayula, J.F. and P. Cornillon. "Cloud Detection from a Sequence of SST Images." Remote Sens. Environ., 55, 80-88, 1996.

Lee, T. and P. Cornillon. "Propagation and Growth of Gulf Stream Meanders between 75W and 45W." J. Phys. Oceanogr., 26, 225-241, 1996.

Lee, T. and P. Cornillon. "Temporal Variation of Meandering Intensity and Domain-Wide Lateral Oscillations of the Gulf Stream." J. Geophys. Res., 100, 13,603-13,613, 1995.

Cayula, J.F. and P. Cornillon. "Multi-Image Edge Detection for SST Images." J. of Atmos. and Oceanic Tech., 12, 821-829, 1995.

Mark R. Abbott — Oregon State University

Education: PhD, 1978, University of California Davis, Ecology
B.S. 1974, University of California Berkeley, Conservation of Natural Resources
Employment: 1993-present Professor of Oceanography, Oregon State University
Inter/National Committees: National Academy of Sciences Committee on Earth Studies, Chair; EOS Payload Panel, Chair; EOS/MODIS Science team; EOS Science Executive Committee
Publications: Dr. Abbott has 40+ refereed publications. Recent ones include:

Abbott, M.R., and B. Barksdale, Wind forcing and phytoplankton pigment patterns off central California. J. Geophys. Res. 96: 14,649-14,667, 1991.

Denman, K.L., and M.R. Abbott, Time scales of pattern evolution from cross-spectrum analysis of advanced very high resolution radiometer and coastal zone color scanner imagery. J. Geophys. Res., 99, 7433-7442, 1994.

C.P. Summerhayes et al., eds., Variability in upwelling systems as observed by satellite remote sensing, Abbott, M. R., and B. Barksdale, In: Dahlem Workshop on Upwelling in the Ocean: Modern Processes and Ancient Records, John Wiley, New York. pp. 221-238, 1995.

Letelier, R.M., and M.R. Abbott, An analysis of chlorophyll fluorescence algorithms for the Moderate Resolution Imaging Spectrometer (MODIS). Remote Sens. Environ., 58, 215-223, 1996.

Letelier, R.M., and M.R. Abbott, Chlorophyll natural fluorescence response to upwelling events in the Southern Ocean. Geophys. Res. Letters, 24, 409-412, 1997.

Elaine Dobinson — Jet Propulsion Laboratory

Education: University of Southern California, M.A., Computer Science
University of Chicago, B.A., MAT Mathematics
Employment: Technical Group Supervisor and Deputy Task Manager for the JPL PO-DAAC.
PI on an ESDIS prototype Earth Science Remote Access Tool using DODS.
IMS Technical Lead for the Version 0 EOSDIS/IMS at JPL.
Data Dictionary co-lead at the IMS system level.
Planetary Data System Lead Database Designer.

Dobinson, E. and R. G. Raskin, Earth Science Remote Access Tool, Proceedings of the Ninth International Conference on Scientific and Statistical Database Management (SSDBM), Olympia, WA, August, 1997.

Robert H. Evans — University of Miami

Education: North Carolina State University, Ph.D. 1973, Electrical Engineering
North Carolina State University, M.S., 1970, Electrical Engineering
Employment: 1987–Present: Research Professor of Oceanography, University of Miami.
Professional Activities: JPL DAAC User Working Group; SeaWIFS Science Working Group; NOAA/NASA Ocean Color Science Working Group; NOAA/NASA Calibration Science

Working Group; EOS V0 Steering Group; NASA Earth Science and Applications Data Systems Committee; JOI Committee for Satellite Ocean Data System Science Working Group; AMS/NCAR Committee for Standards in Satellite Data Systems

Recent Publications:

Evans, Robert H. and Howard R. Gordon, 1994. CZCS "System Calibration": A retrospective examination. J. Geophys. Res. 99(C4), 1994: 7293-7307.

Brown, J., O.Brown and R. Evans. Calibration of Advanced Very High Resolution Radiometer Infrared Channels: A New Approach to Nonlinear Correction. J. Geophys. Res. Oceans 98(C10), 1993, 18,157-18,269.

Gordon, H.R. and R.H. Evans. Comment on "Aerosol and Rayleigh radiance contributions to coastal zone color scanner images" by Eckstein and Simpson. International Journal of Remote Sensing, 14, 1993, 537-540.

Weatherly, G.L., R.H. Evans, O.B. Brown. A comparison of GEOSAT altimeter inferred currents and measured flow at 5400m depth in the Argentine Basin. Deep Sea Res. II, 40(4/5), 1993: 989-999.

Glenn R. Flierl — Massachusetts Institute of Technology

Education: Harvard University, Ph.D., 1975, Physics
Oberlin College, B.A. with highest honors, 1970, Physics

Employment: 1988–Present: Professor, Massachusetts Institute of Technology

Publications: Dr. Flierl has 60+ refereed publications. Recent ones include:

Yano, J–I. and G.R. Flierl (1994) Jupiter's Great Red Spot: compactness conditions and stability. Ann. Geophysicae 12, 1–18.

Flierl, G.R. (1994) Rings: Semicoherent oceanic features. Chaos 4, 355–67.

Flierl, G. R., and K. Haines. The decay of modons due to Rossby wave radiation. Phys. Fluids 6(10), 3487–97.

Yasuda, I. and G.R. Flierl (1995) Two dimensional asymmetric vortex merger: contour dynamics experiment. J. Oceanogr. 51, 145–70.

Michael J. Folk — University of Illinois

Education: Syracuse University, Ph.D. Systems and Information Science
University of Chicago, MAT Mathematics
University of North Carolina at Chapel Hill, BS Mathematics

Employment: 1988-Present: Technical Project Manager of HDF, Software Development Division, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.

1979 - 1988: Professor, Oklahoma State University. Professor of mathematics and computer science. Primary research area: computer file structures.

Inter/National Committees: National Academy of Sciences Mapping Science Committee

Recent Publications:

Folk, M, "HDF: the Next Generation," Proceedings of Scientific Information Management and Data Compression Workshop, Oct. 96.

Folk, M, "Data Models for Scientific Data," Proceedings for Visualization '93, Oct. 93.

Peter A. Fox — National Center for Atmospheric Research

Education: Monash University, Ph.D., 1985, Mathmatics
Monash University, B.Sc. (Hons), 1980, Mathmatics

Employment: 1995–Present: Chief Computing Scientist, High Altitude Observatory at

NCAR
1991–1995: Staff Scientist, High Altitude Observatory at the National Center for Atmospheric Research
Inter/National Committees: SOLERS22 Working Group 1 U.S. Solar Terrestrial Energy Program; NSF RISE Steering Committee
Professional Community Service: Associate Editor of Fundamentals of Cosmic Physics
Publications: Dr. Fox has 40+ refereed publications. Recent ones include:

Fox, P., P. R. Wilson and P. McIntosh, "Coronal Holes and the Polar Field Reversals", *Solar Physics*, (in press), 1997.

Fox, P. and P. A. Gilman, "Joint Instability of Latitudinal Differential Rotation and Toroidal Magnetic Fields below the Solar Convection Zone", *Ap. J.*, (in press), 1997.

Fox, P., J. Fontenla, O. R. White, E. H. Avrett, and R. L. Kurucz, "Calculation of Solar Irradiances I: Synthesis of the Solar Spectrum", submitted to *Solar Physics*, 1997.

Fox, P., J. Fontenla, O. R. White, K. L. Harvey and E. H. Avrett, "Calculation of Solar Irradiances II: Synthesis of Solar Images", submitted to *Solar Physics*, 1997.

Dr. Fox has over 40 publications in peer refereed journals.

Dave Fulker — University Corporation for Atmospheric Research

Education: University of Colorado, MA 1970, Mathmetics
University of Colorado, BA 1966, Mathmatics
Employment: 1984-Present: Unidata Program Director, UCAR
Recent Publications:

Fulker, D, S Bates, and C Jacobs: "Unidata: A Virtual Community Sharing Resources via Technological Infrastructure", BAMS Vol 78 #3, March 1997, Pp 457-468.

Ramamurthy, M K, C Murphy, J Moore, M Wetzel, D Knight, P Ruscher, S Mullen, R DeSouze, D Hawk, and D Fulker: "Teaching Mesoscale Meteorology in the Age of the Modernized Weather Service: A Report on the Unidata/COMET Workshop", BAMS, Vol 76 #12, December 1995, Pp 2463-2473.

Domenico, B, S Bates, and D Fulker: "Unidata Internet Data Distribution (IDD)", Preprints, Tenth International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology, in Nashville, Tennessee, published by the AMS, January 1994, Pp J15-J20.

Sherretz, L A, and D Fulker: "Unidata: Enabling Universities to Acquire and Analyze Scientific Data", BAMS, Vol 69 #4, April 1988, Pp 373-376.

Steven C. Hankin — Pacific Marine Environmental Laboratory

Education: University of Washington, Seattle, M.S., 1984 Applied Mathematics
Reed College, Portland, 1975 B.A., Physics
Professional Experience: 1984-Present: Computer Scientist, NOAA Pacific Marine Environmental Laboratory
Professional Awards: NOAA Performance Awards (7 awards: 1986-97)
Inter/National Committees: NOAA High Performance Computing and Communications computing committee; American National Standards Institute, Computer Graphics Committee, X3H3; Association for Computing Machinery, SIGGRAPH
Recent Publications:

Callahan, J. and S.C. Hankin, "Improving Web access to gridded data: Java tools for climate data servers", 13th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, 1997.

Hankin, S.C., "A portable Web server for gridded data: creating a culture of data interchange", 13th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, 1997.

Hankin, S., D.E. Harrison, J. Osborne, J. Davison, and K. O'Brien, "A strategy and a tool, Ferret, for closely integrated visualization and analysis", Journal of Visualization and Computer Animation, 7 , pp149-157, 1996.

Hankin, S., D.E. Harrison, J. Osborne, J. Davison, and K. O'Brien, "Remote and local visualization and analysis with the Ferret program", 12th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Jan. 28-Feb. 2, 1996, Atlanta, GA, AGU, 139-144.

Judy C. Holoviak — American Geophysical Union

Education: Westminster College, B.A. Magna Cum Laude, 1963, English with Chemistry minor.
Employment: 1980–Present, Director of Publications; Deputy for Executive Director; Director for International Activities, American Geophysical Union.
Professional Activities: Association of Earth Science Editors: Vice-President (1984), President (1985)
Society for Scholarly Publishing: Secretary-Treasurer (1980-1986); Vice-President (1988-1990); President (1990-1992)
American Institute of Physics: Member of Governing Board (1997-2000)
Recent Publications:

Seitter, K.L., and J.C. Holoviak, Earth Interactions: An Electronic Journal Serving the Earth System Science Community, Bulletin of the American Meteorological Society, pp. 2095-2100, September 1996.

Holoviak, J.C., "The mixed blessings of society publishing", LOGOS, vol. 7, no. 1, pp. 106-112, 1996.

Holoviak, J.C., "Pricing–A Scientific Society's Perspective", GSA Today, p. 93, April 1993.

Holoviak, J.C., "Scholarly Communications: New Realities: The Risks and Rewards", Library Hi Tech, vol.10, no. 3, pp.73-75, 1992.

Holoviak, J.C., "Working to Improve the Entire System", The Journal of Academic Librarianship, vol. 15 no. 2, pp. 77-78, May 1989.

Lola M. Olsen — Goddard Space Flight Center

Education: M.A., University of North Carolina, Geography/Earth Science
B.S., Michigan State University, Biological Science
Professional Experience: 1994-present: Project Manager, Global Change Master Directory, NASA, Goddard Space Flight Center
Professional Awards: Five NASA/GSFC Certificates for Outstanding Achievement Award since 1991; NASA/GSFC Special Act Group Award, 1995; NASA Merit Increase Award, 1991, 1992; GSFC Group Achievement Award, 1992
Inter/National Committees: Committee on Earth Observing Satellites, Working Group on Information Systems and Services Access Subgroup, Task Team Leader
Recent Publications:

Olsen, Lola, Global Change Master Directory Enhances Search for Earth Science Data, Eos, Transactions, American Geophysical Union, Vol. 77, Number 18, April 30, 1996.

Scialdone, J. and L. Olsen, The Global Change Master Directory, Library Hi Tech, Volume 13, Number 1-2, pp 44-46, 1995.

Olsen, Lola, and NCDS staff, Greenhouse Effect Detection Experiment (GEDEX) CD-ROMS, 1992.

Olsen, Lola, and NCDS staff, International Satellite Cloud Climatology Project (ISCCP) CD-ROMS, 1992.
Olsen, Lola M. and Charles McClain, Cooperative Efforts in Support of Ocean Research Through NASA's Climate Data System, Eighth International Conference on Interactive Information and Processing Systems, American Meteorological Society, January 1992.

7 Other Participants

In addition to the Project Members listed in the previous section systems programmers and faculty involved on this effort are:

James Gallagher - URI - Technical lead for DODS - M.S. Computer Science, URI, 1990.
Dan Holloway - URI - Systems programmer/manager - M.S. Computer Science, URI, 1993.
James Kowalski - URI - Associate Professor of Philosophy and Computer Science - Ph.D. Purdue University, 1975
Reza Nekovei - URI - Systems programmer - Ph.D. Electrical Engineering, URI, 1995.
Joan Peckman - URI - Associate Professor of Computer Science - Ph.D. Computer Science, University of Connecticut 1990.
Rob Raskin - JPL - Systems programmer - Ph.D. in Atmospheric Science 1992 University of Michigan.

8 Proposed Costs

The following pages present the unified budget for the proposed effort. Aside from computer programmers we are requesting salary support for a Project Manager and for a technical writer. The duties of the project manager are described in Section 5 and URI will match part of this individual's salary. The technical writer, Tom Sgorous, will continue to work part time as a consultant to document all aspects of the system. Sgorous wrote the first version of the User's Guide. One staff member will work at the GCMD. This person will be responsible for DIF generation and for maintaining the DODS server at the Goddard DAAC.

Travel will be used by computer staff to attend the quarterly developers meetings and the annual Steering Committee meetings as well as WP-Federation meetings.

Two computers will be purchased to absorb the Central Processor Unit (CPU) load associated DODS servers at the Goddard and JPL DAACs. There will be very little need for additional disc space at the DAACs to accommodate the DODS servers. Three smaller computers will be purchased with URI match. These will be used by staff to develop software and to manage the overall effort.

An un-named subcontract has been included in the budget. This subcontract is for straightforward programming tasks such as the development of Java clients and servers. The level of effort for these tasks is estimated to be $1\frac{1}{2}$ full time programmers. We could hire these programmers at the university but would prefer not to have the long term commitment that full time employees entail. If however, through the bid process we can not find a vendor to provide these services at the estimated level we will hire programmers and perform them in-house.

All of the other subcontracts are for the Project Members. We have included them in the budget as subcontractors, but if we are selected, NASA may elect to fund them directly thus saving the overhead on these subcontracts. This overhead comes to $47,500.

9 Cooperative Agreement Payment Schedule

Below we summarize, by group, payment milestones. A detailed breakdown of the URI milestones is presented in Appendix B.2.

Year 1

Qtr 1

URI: URI 1 Qtr milestones. *Total: $166,469*
NCAR/HAO: Install DODS servers for CEDAR data holdings *Total: $42,235*
UNIDATA/PUSH: Define a suitable datastream through IDD *Total: $33,632*
UNIDATA/US: Set up a list server and Web pages *Total: $33,632*
JPL/HDF-EOS: HDF-EOS to DODS GRID type conversion *Total: $53,800*

Qtr 2

URI: URI 2 Qtr milestones. *Total: $166,469*
NCAR/HAO: Build a Matlab GUI for access to the CEDAR data *Total: $30,360*
UNIDATA/PUSH: Implement a single-sender/single-receiver *Total: $27,694*
UNIDATA/US: Begin fielding questions *Total: $27,694*
PMEL/DF: Demonstrate a virtual data base with LAS *Total: $45,207*
JPL/HDF-EOS: HDF-EOS to DODS point type conversion *Total: $41,925*
NCSA/HDF–EOS: Report on study HDF-based storage formats *Total: $44,228*

Qtr 3

URI: URI 3 Qtr milestones. *Total: $166,469*
NCAR/HAO: Add user login, password verification, etc. *Total: $30,360*
UNIDATA/PUSH: Design a DODS server configuration *Total: $27,694*
UNIDATA/US: Create Web pages for searching support archive *Total: $27,694*
PMEL/DF: Install an operational LAS at PMEL *Total: $33,332*
JPL/HDF-EOS: HDF-EOS to DODS Metadata conversion; HDF Server Maint. *Total: $41,925*

Qtr 4

URI: URI 4 Qtr milestones. *Total: $166,469*
NCAR/HAO: Develop netCDF or HDF form of MLSO data *Total: $30,360*
UNIDATA/PUSH: Implement the DODS server at IDD rec. node *Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
PMEL/DF: Demonstrate operational LAS systems with DODS *Total: $33,332*
JPL/HDF-EOS: Conversion of SDS *Total: $41,925*
URI/CS/DMM: Development of Directory Mining Tool *Total: $147,648*
NCSA/HDF-EOS: Report on mapping between HDF–EOS and DODS data types *Total: $56,103*

Year 2

Qtr 1

URI: URI 1 Qtr milestones. *Total: $166,469*
UNIDATA/PUSH: Test IDD/DODS system *Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
JPL/HDF-EOS: Conversion of VData *Total: $41,925*

Qtr 2

URI: URI 2 Qtr milestones. *Total: $166,469*
NCAR/HAO: Access to all HAO data with WWW base browser *Total: $60,720*
UNIDATA/PUSH: Test IDD/DODS system *Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
PMEL/DF: Install all JAVA interface at 3 sites *Total: $33,332*

JPL/HDF-EOS: Conversion of raster *Total: $41,925*

Qtr 3

URI: URI 3 Qtr milestones. *Total: $166,469*
UNIDATA/PUSH: Replicate the IDD receiver/DODS server *Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
PMEL/DF: Install shopping basket functionality *Total: $33,332*
JPL/HDF-EOS: Conversion of Metadata; HDF Server Maint. *Total: $41,925*

Qtr 4

URI: URI 4 Qtr milestones. *Total: $166,469*
NCAR/HAO: Implement registration of datasets *Total: $60,720*
UNIDATA/PUSH: Extend the system by replicating *Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
PMEL/DF: Demonstrate a data-fusing virtual data base *Total: $33,332*
JPL/HDF-EOS: Conversion of GRID *Total: $41,925*
URI/CS/DMM: Specification of Expert System for DMM *Total: $147,649*
NCSA/HDF 5: Implementation of HDF 5 API with DODS protocol *Total: $117,936*

Year 3

Qtr 1

URI: URI 1 Qtr milestones. *Total: $166,469*
NCAR/HAO: Install DODS servers at remote CEDAR sites *Total: $30,360*
UNIDATA/PUSH: D&I IDD-¿GCMD update *Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
JPL/HDF-EOS: Conversion of point *Total: $41,925*

Qtr 2

URI: URI 2 Qtr milestones. *Total: $166,469*
UNIDATA/PUSH: D&I a prototype "virtual DODS server" *Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
PMEL/DF: Demonstrate operational servers *Total: $33,332*
JPL/HDF-EOS: Conversion of swath *Total: $41,925*

Qtr 3

URI: URI 3 Qtr milestones. *Total: $166,469*
NCAR/HAO: Search and Query procedure for all HAO data *Total: $60,720*
UNIDATA/PUSH: Test and refine the above two prototypes *Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
PMEL/DF: Demonstrate fusing gridded data with Sat data *Total: $33,332*
JPL/HDF-EOS: HDF Server maintenance *Total: $41,925*

Qtr 4

URI: URI 4 Qtr milestones. *Total: $166,469*
NCAR/HAO: Measures total quantitatively usage and satisfaction *Total: $30,360*
UNIDATA/PUSH: Create a design document*Total: $27,694*
UNIDATA/US: Operational User Support *Total: $27,694*
PMEL/DF: Demonstrate an server with extensions *Total: $33,332*
JPL/HDF-EOS: Conversion to HDF-EOS Metadata *Total: $41,925*
NCSA/HDF 5: Report on DODS/HDF 5 *Total: $58,968*

10 Current and Pending

This proposal is clearly pending with all PMs listed below. It will be presented only for the Project Leader (PL). Hey, why killl another tree? Also "current and pending" are only presented for those requesting funds from this proposal.

Peter Cornillon — Project Lead — URI

Current Support

Note: Location of all Research is the Graduate School of Oceanography
University of Rhode Island

"Biological and Physical Responses of the Upper Ocean to Wind Stress" (with J. Yoder) 17% *JPL/NASA 957627* $941,942 10/1/88–12/31/97

"The Oleander Project: Continued Monitoring of Annual and Subannual Mass and Heat Fluxes in the Northwest Atlantic Across the Sargasso Sea, the Gulf Stream, and the Shelf Between Bermuda and New Jersey" 8% (with T. Rossby and E. Gottlieb) *NOAA NA56GP0220* $733,642 4/1/95–3/31/98

"Production of Western North Atlantic Full Resolution SST Fields from the AVHRR Sensor" 8% *NASA NAGW4764* $175,319 7/1/95–6/30/97

"Support for CEOS Working Group on Data Catalog Subgroup Activities as a NASA Representative" 0% *NASA NCC5123* $60,000 9/15/95–9/14/97 (3rd Yr $30,000)

"Wind Forcing of Physical and Biological Processes in the Upper Ocean" 8% *NS033A06 Oregon State University* $159,954 9/15/95–12/31/00

"An Analysis of Persistent Ocean Fronts on the Continental Shelf Between Cape Hatteras and Cape Cod 8% (with S. Nixon) *NOAA NA66RG0303* $187,101 10/01/96-2/28/98

"Support for the Oceanographic Remote Sensing Lab" 2% *NOAA NA57FE0542* $49,451 9/1/96–8/31/97

"US Globec: Spatial and Temporal Variability in the Occurence Distribution and Structure of SST Fronts in the Georges Bank Region" 8% *NOAA NA57FE0542* $130,450 10/1/96–9/30/98

"Development of DODS Compliant Client and Server Libraries to Support Internet Access to NODC Oceanographic Datasets" 4% *DOC NA57FE0448* 7/1/95–6/30/97 (3rd Yr $30,000)

"Survey of Ocean Fronts with SST, Altimeter and In Situ Data" 8% (with I. Belkin) *NASA NAG53736* $126,000 3/1/97–2/28/98

"The Impact of First Mode Planetary Waves on Western Boundary Currents" 8% *Fed Sub/NASA 960913* $84,083 3/1/97–12/31/97 (2nd Yr $79,899, 3rd Yr $83,556)

"Population and Support of the Distributed Oceanographic Data Systems" 8% *NSF OCE9617804* $186,000 6/1/97–5/31/98 (2nd Yr $168,817, 3rd Yr $142,548)

Pending Support

This proposal, "The Distributed Oceanographic Data System (DODS): A Framework for Access to Scientific Data in the EOS Federation", 8% NASA/MTPE, $4,000,000 1/1/98–12/31/00

"Production of Western North Atlantic Full Resolution SST Fields from the AVHRR Sensor" 4% NASA $171,489 10/1/97–9/30/99

"Preparing Students, Practitioners and Researchers for Ocean Remote Sensing in the Twenty-First Century" 6% NASA $284,571 10/1/97-09/30/98

Elaine Dobinson — Project Member — JPL

Current Support

"NASA/JPL PODAAC Deputy Task Manager", 100%, Location of Research: JPL, NASA UPN 428-33, $165,000 1/1/97–12/3098, (2nd Year $170,000)

No Pending Support

Mike Folk — Project Member — U. Illinois

Current Support

Note: All Research conducted at NCSA/University of Illinois

"HDF Support for the ASCI Program", 0%, *ASCI/Livermore Labs*, $140,447 8/1/97 – 7/31/98
"HDF Support for ASCI w/Special Tasks", 0%,*ASCI/Los Alamos*, $57,500 6/1/97–5/31/98
"DAAC-in-the-box", 0%, *NASA 505/Hughes STX* $120,000 1/1/97–12/31/97
"HDF Configuration Records & QA" 50% *NASA Hughes ITS* $134,342 $537,366 3/11/96–12/31/99
Support for EOSDIS", 0%, *ESDIS Coop. Agreement* $992,246 8/1/95–7/31/01
"Support for Large Datasets in EOSDIS", 25%, *MTPE/NRA* $538,332 12/1/94 - 12/31/97
"Project Horizon", 25%, *NASA* $2,146,342 9/30/94–10/1/97

Pending Support

"The Distributed Oceanographic Data System: A Framework for Access to Scientific Data in the EOS Federation" 0%, *NASA/MTPE* $203,928 1/1/98–12/31/00
"Earth Imagery Coalition " 0%, *NASA/MTPE* $295,150 1/1/98–12/31/00
"BOREAS Data and its Access" 0%, *NASA/MTPE* $353,048 1/1/98–12/31/00

Peter Fox — Project Member — NCAR

Current Support

Note: Location of all Research is at NCAR

"Convective Turbulence and Mixing in Astrophysics", 5%, *NASA/HPCC*, $325,000 (FY93).

"SunRISE - Application of Spectral Synthesis to the Study and Improvement of Surrogates for Solar Radiation in the Visible, UV, and EUV.", 20%, *NSF-Special*, $298,972, 1993/12/01-1997/10/31.
"SunRISE - Solar Variability in the Minimum and Rise of Cycle 23.", 5%, *NSF-Special*, $139,353, 1997/07/01-2000/06/30.
"SunRISE operational budget", 15%, *NSF-Special*, $212,225, 1997/07/01-1998/06/30 (with Knölker, HAO/NCAR)

Pending Support

"SunRISE - Application of Spectral Synthesis to the Study of Solar Visible, UV, EUV and Infrared Variability", 20%, *NSF-Special*, $350,000, 1997/11/01-2000/11/01

Dave Fulker — Project Member — UCAR

Current Support

Note: Location of all Research is at UNIDATA
University Corporation for Atmosheric Research

"Unidata: 1993-1998", *NSF ATM–9218790*, $10,799,404 5/93–4/98
"NIDS Data Products", *NOAA 57GP0576*, $174,375 1994–1999
"SOO and University Access to COMET Case-Study Data", *NOAA NA67WD0097*, $170,498 5/96–5/99 (FY 1999 $100,000)

Pending Support

"Using Instructional Technologies and Satellite Data for College-level Education in the Atmospheric and Earth Sciences", *NSF Lower Atmosphere Research Section*, $45,411, FY97
"Unidata 2003", *NSF Lower Atmosphere Research Section*, $18,124,644 May 1998 thru April 30, 2003

Steve Hankin — Project Member — PMEL

Current Support

Note: All Research conducted at NOAA's Pacific Marine Environmental Laboratory

"From the Web to the Desktop: Unified Access to Gridded Data with Java and the Ferret Program", a Java-enhanced interface to the Ferret-based "Live Access to Climate Data" server – *NOAA/ESDIM*, $100K
"Data Fusion of Point and Gridded Data from Remote and Local Sources", a collaborative effort with NOAA/NCDC to merge point data from NCDC with gridded data from PMEL – *NOAA/ESDIM*, $80K

"Server-in-a-box", Packaging the "Live Access to Climate Data" server for easy configurability and installation. – *NOAA/HPCC*, $70K

Pending Support

"Web Access to Archived and Near-Real Time Oceanographic Data at Pacific Fisheries Environmental Laboratory", a joint National Marine Fisheries Service/PMEL proposal – *NOAA/OGP*, $20K (PMEL portion)

"Seasonal-to-interannual variability of the global ocean circulation based on altimetry, hydrography, and the Princeton Ocean Model", a joint NationalMarine Fisheries Service/PMEL proposal – *NOAA/OGP*, $25K (PMEL portion)

""A Collaborative Environment for OACES Model Sharing (Using the World Wide Web)", building a Web-based system to assist with collaborative modeling of the global carbon cycle – *NOAA/OGPP*, $40K

"A Web Bathymetry Visualizer", a joint National Geophysical Data Center/PMEL proposal – *NOAA/ESDIMP*, $40K (PMEL portion)

"A Research-Quality Regional Weather Perspectives Browser", a joint National Weather Service/PMEL proposal – *NOAA/ESDIMP*, $30K (PMEL portion)

"Live Access to Global Atmospheric/Oceanic/Biological Indicators Derived from the Climatic and Synoptic Observed Data", a joint National Marine Fisheries Service/PMEL proposal – *NOAA/ESDIMP*, $20K (PMEL portion)

"Advanced Web Server for Gridded Data" – *NOAA/ESDIM*, $50K

11 Acronyms

ADCP Acoustic Doppler Current Profiler.

AGU American Geophysical Union.

AI Artificial Intelligence.

API Application Program Interface.

AVHRR Advanced Very High Resolution Radiometer.

BUFR Binary Universal Form for the Representation of meteorological data. A data format.

CDF Common Data Format.

CDR Common Data Representation.

CEDAR Coupling, Energetics and Dynamics of Atmospheric Regions.

CEOS Committee on Earth Observation Satellites.

CIP Catalog Interoperability Protocol.

CORBA Common Object Request Broker Architecture

CPU Central Processor Unit. Used to refer to a computer.

CZCS Coastal Zone Color Scanner.

DAAC Distribute Active Archive Center.

DAP Data Access Protocol.

DAS Data set Attribute Structure.

DBMS Data Base Management System.

DDS Data Descriptor Structure.

DIF Directory Interchange Format.

DODS Distributed Oceanographic Data System.

ECS EOSDIS Core System.

EOS Earth Observing System.

EOSDIS Earth Observing System Data Information System.

ESIP Earth Science Information Partner.

FGDC Federal Geographic Data Committee. Also refers to a metadata standard.

GCMD Global Change Master Directory.

GMT Generic Mapping Tool. A plotting package based on NetCDF.

GRIB GRid In Binary. A data format used in meteorology.

GSFC Goddard Space Flight Center.

GUI Graphical User Interface.

HAO High Altitude Observatory.

HDF Hierarchical Data Format.

HDF-EOS Hierarchical Data Format - EOS. This is the EOS sanctioned version of HDF.

IDD Internet Data Distribution.

http the hypertext transport protocol. The underlying transfer protocol of the Web.

IDL Interactive Display Language. An analysis package for scientific data developed by Research Systems, Inc.

JGOFS Joint Global Ocean Flux Experiment. Also referencesa data system developed for the project.

JPL Jet Propulsion Laboratory.

LAS Live Access Server. Refers to a server at PMEL providing network access to PMEL data holding.

MIT Massachusetts Institute of Technology.

MATLAB MATLAB. An analysis package for scientific data developed by MATHWORKS, Inc.

MODIS MODerate-resolution Imaging Spectroradiometer.

MTPE Mission To Planet Earth.

NASA National Aeronautics and Space Administration .

NCAR National Center for Atmospheric Research .

NCSA National Center for Supercomputing Applications.

NDBC National Data Buoy Center.

NetCDF NETwork Common Data Format.

NOAA National Oceanic and Atmospheric Administration.

NSCAT NASA SCATterometer.

NSF National Science Foundation.

OSU Oregon State University.

PI Principal Investigator. Often used to refer to the research scientist.

PL Project Leader.

PM Project Member.

PMEL Pacific Marine Environmental Laboratory.

PO-DAAC Physical Oceanography – Distributed Active Archive Center.

RISE Radiative Inputs from Sun to Earth

RMI Remote Method Invocation.

SST Sea Surface Temperature.

UCAR University Corporation for Atmospheric Research.

UMiami University of Miami.

URI University of Rhode Island.

URL Uniform Resource Locator.

WOCE World Ocean Circulation Experiment. A multi-national program to study the circulation of the oceans.

WP-ESIP Working Prototype Earth Science Information Partner.

WWW World Wide Web.

XBT Expendable BathyThermograph.

APPENDIX A

LETTERS OF SUPPORT

APPENDIX B

INSTITUTIONAL RESPONSIBILITY FOR TASKS

AND

TASK BREAKDOWN FOR THE UNIVERSITY OF RHODE ISLAND

B.1 Institutional Responsibility for Tasks

Following is a summary of the role played by each of the 11 institutions involved in this effort.

- The University of Rhode Island will provide coordination for the effort as well as perform a significant fraction of the work associated with enhancements to the DODS core software. It will also develop an expert system in collaboration with the GCMD to automatically populate the GCMD with descriptions of DODS-compliant data sets. (Sections 2.2.1, 2.2.2, 2.2.3, 2.2.4, 2.2.5, 2.2.6, 2.2.8, 2.2.9, 2.2.10, and 5.)
- The JPL's PO-DAAC will provide DODS access to the physical oceanographic data in its on-line archives, will develop, install and maintain the HDF-EOS client and server libraries for DODS, and will work with the DODS core group to provide a seamless interface to EOSDIS. (Sections 2.1.2 and 2.2.5.)
- Unidata will coordinate an effort to integrate their "broadcast" system for meteorological data with DODS, will provide user services support and will develop a feedback mechanism for users willing to provide comments with regard to data quality in the system. (Sections 2.2.10, 2.3.1, and 2.3.2.)
- The Pacific Marine Environmental Laboratory of NOAA will develop an application within Ferret that will permit fusion of gridded and non-gridded data using DODS as the data access mechanism. (Section 2.2.7.)
- The National Center for Supercomputing Applications at the University of Illinois will work with the DODS core group to integrate their "chunking" approach to data compression for use with DODS compliant data sets. They will also develop and maintain HDF-5 client and server libraries. (Sections 2.2.5 and 2.2.8.)
- The High Altitude Observatory (HAO) of UCAR/NCAR will develop, install and maintain DODS servers for their CEDAR data base and will develop MATLAB and IDL GUI's to facilitate access to these data via DODS. (Section 2.1.3.)
- The GCMD of GSFC will install and maintain DODS servers at the Goddard DAAC for the MODIS level 2 and level 3 data stream. They will also maintain a directory of DODS compliant data sets and manage the registration and maintenance process associated with these entries. (Sections 2.1.2 and 2.2.4.)
- The American Geophysical Union will help to work with its membership to obtain feedback with respect to DODS, will install a DODS server for their electronic data depository and will investigate the use of distributed data systems in the publication process in general. (Sections 2.3.3 and 2.3.4.)
- The University of Miami will provide access to the MODIS research data stream via the Unidata IDD and DODS. It will also provide access to the MODIS match-up data set. (Section 2.1.1.)
- Oregon State University will act as a "mirror" site in the IDD context for the MODIS research data stream and will also serve these data via DODS. (Section 2.1.1.)
- The Massachusetts Institute of Technology is part of the DODS core group and will continue to participate in this way, providing input with regard to system design and functionality from the user perspective.

B.2 Detailed Description of Project Milestones

Project Manager:URITask 1

Year 1

1 Qtr Quarterly Progress Report.

2 Qtr Quarterly Progress Report.

3 Qtr Quarterly Progress Report.

4 Qtr Quarterly Progress Report.

Year 2

1 Qtr Quarterly Progress Report.

2 Qtr Quarterly Progress Report.

3 Qtr Quarterly Progress Report.

4 Qtr Quarterly Progress Report.

Year 3

1 Qtr Quarterly Progress Report.

2 Qtr Quarterly Progress Report.

3 Qtr Quarterly Progress Report.

4 Qtr Quarterly Progress Report.
Task Total: 36 months

Catalog Server:URITask 2

Year 1

1 Qtr Design of Catalog Server component which performs multiple DODS server accesses and bundles results into a single DODS data object from transmission back to the client

2 Qtr Develop initial version of Catalog Server Component: retrieves all data from list of URLs

4 Qtr Develop final version of Catalog Server Component: combines all retrieved data into singel DODS data object
Task Total: 13 months

Translation:URITask 3

Year 1

1 Qtr Design of format translator for DODS servers

2 Qtr Implementation into DODS Core of format translator for DODS Grid data types

3 Qtr Implementation into DODS Core of format translator for other DODS data types

4 Qtr Release optimized DODS Core format translator; Document Translator features

Year 2

2 Qtr Release of DODS APIs format translator; Document Format Translation:Technical Writer

3 Qtr Specification of DODS standardized Units translation, Lat/Long and Time

4 Qtr Specification of DODS standardized Units translation, other geophysical parameters; Port to UDunits

Year 3

1 Qtr Design of DODS Core Unit translator for Lat/Long and Time

2 Qtr Release DODS Core Unit translator for Lat/Long and Time; Document Unit Translation: Technical Writer

3 Qtr Design and implementation of DODS Core Unit translator for other geophysical parameters

4 Qtr Release of DODS Core Unit translator for other geophysical parameters; Document DODS Unit Translation:Technical Writer

Task Total: 40 months

Directory Metadata Mining:URITask 4

Year 1

1 Qtr Design of Directory Metadata Mining Tool

Year 3

1 Qtr Expert System implementaion: Implementation of ES based on ES research and design results; Document Expert System Documentation:Technical Writer

Task Total: 5 months

Data Server Development and Upgrades:URITask 5

Year 1

1 Qtr NetCDF 3 server library upgrade; client library upgrade

2 Qtr NetCDF 3 client library upgrade

4 Qtr Design and Implement DODS GRIB abd BUFR servers; Document GRIB and BUFR Servers:Technical Writer

Task Total: 11 months

Java Conversion:URITask 6

Year 1

4 Qtr Complete conversion of DODS Core code to JAVA

Year 2

4 Qtr Complete conversion of DODS API libraries to JAVA; Document JAVA version of DODS:Technical Writer
Task Total: 14 months

System Wide Metrics and Integrated Software Testing:URITask 7

Year 1

1 Qtr DODS version release; Specification of testing suite for DODS servers; Design of Metrics to address Earth Systems Science Goal within the DODS federated system

2 Qtr DODS version release; Specification of testing suite for DODS clients; Design of metrics to address Technology Develop goals; Document DODS client/server Testing Suite:Technical Writer

3 Qtr DODS version release; Implement Technology Development and Earth Systems Science metrics

4 Qtr DODS version release; Preliminary analysis of Technology Development and Earth Systems Science metrics

Year 2

1 Qtr DODS version release; System-wide integrated testing of DODS Core and API libraries tied to DODS version upgrades; Design of WP-Federation metrics in collaboration with other WP-ESIPS

2 Qtr DODS version release; Implementation of WP-Federation metrics

3 Qtr DODS version release; Preliminary analysis of WP-Fed. statistics

4 Qtr DODS version release; 2nd analysis of Technology Development and Earth Systems Science metrics

Year 3

1 Qtr DODS version release

2 Qtr DODS version release

3 Qtr DODS version release; 2nd analysis of WP-Fed. statistics

4 Qtr DODS version release; 3rd analysis of Technology Development and Earth Systems Science metrics
Task Total: 33 months

GCMD Coordinator:URITask 8

Year 1:

1 Qtr Coordination with GCMD and installation of DODS server

2 Qtr Coordination with GCMD and implementation of enhancements to GCMD/DODS Registration service

3 Qtr Coordination and evaluation of GCMD/DODS Registration process; Report on the process and procedures for developing a DIF using the GCMD/DODS Registration Service. report to be basis for Expert System Design

4 Qtr Upgrade DODS servers and registration components at GCMD; Yearly report on GCMD/DODS usage statistics

Year 2:

1 Qtr Update and verify DODS Datasets reference by the GCMD; Coordinate and suppport registration of new DODS datasets and servers

2 Qtr Coordinate and suppport registration of new DODS datasets and servers

3 Qtr Coordinate and suppport registration of new DODS datasets and servers

4 Qtr Coordinate and suppport registration of new DODS datasets and servers

Year 3:

1 Qtr Update and verify DODS Datasets reference by the GCMD; Coordinate and suppport registration of new DODS datasets and servers

2 Qtr Coordinate and suppport registration of new DODS datasets and servers

3 Qtr Coordinate and suppport registration of new DODS datasets and servers

4 Qtr Coordinate and suppport registration of new DODS datasets and servers

Task Total: 36 months

ECS Interface: URITask 9

This task is predicated on the specification of the WP-Federation architecture which will be designated by the WP-ESIPS early in the life of the program. We anticipate that the design and implementation of the WP-Federation interface will not only include the WP-ESIPs but will also interface with EOSDIS.

Year 1

4 Qtr Requirements specification of WP-Federation Interface

Year 2

2 Qtr Design of Prototype WP-Fed Interface

4 Qtr Implementation of Prototype WP-Fed Interface

Year 3

2 Qtr Design of Version 1 WP-Fed Interface

4 Qtr Implementation of Version 1 WP-Fed Interface
Task Total: 26 months

Technical Writer: URITask 10

Tasks for the Technical Writer are in support of DODS development activities and system upgrades. The Technical Writer is responsible for generating User and Developer documentation on DODS. This includes information on the installation and use of DODS software and technical manuals to support independent development efforts using DODS. The listing below includes tasks which are associated with specific project supertask milestones (e.g., Data Server Development and Upgrades) and independent tasks which are not associated with a specific supertask. The supertask associated Technical Writing milestones are designated with an ** and are not tallied in the total below, they have been included in the tally of the supertasks and are provided here to provide a complete picture of the Technical Writers responsibilities during the project.

Year 1

1 Qtr Special Server Installation Guide

2 Qtr **System-wide Metrics and Sftw. Testing-Document Testing suite:; Special Server Installation Guide

3 Qtr Special Server Installation Guide

4 Qtr **Document Catalog Server:;**Document Core Format Translator:; Special Server Installation Guide

Year 2

1 Qtr General update of User Guide and Core Developers Guide

2 Qtr **API Format Translator:;**Data Server Dev-Document GRIB and BUFR Servers:

4 Qtr **Document JAVA Conversion

Year 3

1 Qtr **Metadata Mining-Document Expert System:

2 Qtr **Translation-Document Lat/Long and Time Unit Translator:

4 Qtr **Translation-Document Geophysical Parameter Unit Translator:
Task Total: 4.5 months
** Not counted in Total

NCAR/HAO:NCARTask 1

Year 1

1 Qtr Build and install DODS servers for CEDAR data holdings, with support of DODS core group; Continue migration of the CEDAR data out of theexisting DBMS

2 Qtr Build a Matlab GUI for access to the CEDAR data; Participate in testing of the IDL GUI being developed

3 Qtr Add user login, password verification, transaction and data request logging and summary statistics

4 Qtr Develop netCDF or HDF form of MLSO data; Implement access to RISE and MLSO data

Year 2

2 Qtr Facilitate access to all HAO data by enhancing exisiting web browser based interfaces; Select friendly users for testing each stage of the developed interfaces

4 Qtr Implement registration of datasets with central data server for CEDAR and RISE program and register those datasets with the GCMD

Year 3

1 Qtr For CEDAR, establish DODS servers at remote sites for database distribution

3 Qtr Develop search and query procedures for all HAO datasets

4 Qtr Develop quantitative measures of usage and satisfaction of users of the new interface

Task Total: 36 months

UNIDATA/Push System:UCARTask 1

Year 1

1 Qtr Define a suitable datastream, including the source institution, the "product" segmentation, and the identifier string that will accompany each product through IDD.

2 Qtr Implement a single-sender/single-receiver version of the IDD carrying data as defined above.

3 Qtr Design a DODS server configuration suitable for providing access to a (changing) collection of products from the above datastream.

4 Qtr Implement the DODS server at the IDD receiver node, configured to serve new data as they arrive.

Year 2

1 Qtr Test and refine the various components of this IDD/DODS system in preparation for building a mirrored version.

2 Qtr Test and refine the various components of this IDD/DODS system in preparation for building a mirrored version.

3 Qtr Replicate the IDD receiver/DODS server system at one or two additional sites in preparation for their service as top-level relays.

4 Qtr Extend the system by replicating more of the systems and implementing a relay topology that provides timely delivery to about 8 mirrors.

Year 3

1 Qtr Design and implement a prototype method for automatically updating the Master Directory when a mirror changes state.

2 Qtr Help design and implement a prototype "virtual DODS server" using a notification system similar to that for updating the Master Directory

3 Qtr Test and refine the above two prototypes.

4 Qtr Create a design document that reflects experiences learned and specifies how push-based DODS servers should be employed in future EOS activities.

Task Total: 36 months

UNIDATA/User Support:UCARTask 2

Year 1

1 Qtr Set up a list server and Web pages, including pages for subscribing to and unsubscribing from the lists, and pages for submitting questions, all with suitable links to and from other DODS pages.

2 Qtr Begin fielding questions, interacting with DODS developers, and building a searchable record of support-user interactions.

3 Qtr Create Web pages for searching the support archive. Create an FAQ. Continue fielding questions, interacting with developers, and building the record.

4 Qtr Field all questions that arrive in Q4 (except for the last 3 days). Interact with developers and/or perform development work in order to solve problems with or to enhance the various components of DODS. Add new material to and delete obsolete material from the (publicly) searchable record of support-user transactions. Maintain the FAQ to accurately reflect the state of DODS and its use in Q4. Adapt topology of DODS/IDD push system to conform to changes in usage patterns that occur in current quarter.

Qtrs 5–12 (Continued from Qtr 4)

Task Total: 36 months

PMEL/Data Fusion:PMELTask 1

Year 1

2 Qtr Demonstrate a virtual data base formed by at least three separate operational (not prototypes) LAS systems sharing data sets

3 Qtr Install an operational LAS version at PMEL that is capable of data fusion on local data sets

4 Qtr Demonstrate at least three separate operational LAS systems, each capable of "DODS-fusing" (overlaying and differencing where variables are conformable) its gridded data holdings with its sister servers

Year 2

2 Qtr The all-Java interface installed on at least three servers as an operational interface

3 Qtr The shopping basket functionality installed on at least three servers drawing from the same virtual data base

4 Qtr Demonstrate a data-fusing virtual data base of at least six operational servers

Year 3

2 Qtr Demonstrate operational servers which are capable of fusing non-conformable variables (variables of unmatched spacial or temporal resolutions or units) from distributed locations

3 Qtr Demonstrate an operational server that can fuse gridded data with selected satellite and in-situ data sets from other DODS servers specified "live" by the user

4 Qtr Demonstrate an operational server that includes the following extensions; the application of smoothers and gap fillers along various axes; the computation of monthly and annual climatological fields from time ranges; and performing limited frequency domain analyses

Task Total: 36 months

JPL/HDF-EOS Server:JPLTask 1

Year 1

1 Qtr HDF server development: HDF-EOS to DODS GRID type conversion

2 Qtr HDF server development: HDF-EOS to DODS point type conversion; HDF-EOS to DODS swath type conversion

3 Qtr HDF server development: HDF-EOS to DODS Metadata type conversion; HDF Server maintenance

4 Qtr HDF client development: Conversion of SDS

Year 2

1 Qtr HDF client development: Conversion of VData

2 Qtr HDF client development: Conversion of raster)

3 Qtr HDF client development: Conversion of Metadata; HDF Server Maintenance

4 Qtr HDF client development: Conversion of GRID

Year 3

1 Qtr HDF client development: Conversion of point

2 Qtr HDF client development: Conversion of swath

3 Qtr HDF Server maintenance

4 Qtr HDF client development: Conversion oto HDF-EOS Metadata; HDF Server maintenance

Task Total: 36 months

URI/CS/Directory Metadata Mining:URI/CSTask 1

Year 1

4 Qtr Development of the general strategies for automating the integration of the local catalogs into a central global catalog using the metadata at each site

Year 2

4 Qtr Development of the structure of the general catalog to permit data mining and the development of the user usage patterns tracking system

NCSA/HDF-EOS:NCSATask 1

Year 1

2 Qtr Report on study of of HDF-based storage formats to support efficient access to DODS
Task Total: 4.5 months

NCSA/HDF-EOS:NCSATask 2

Year 1

4 Qtr Report on study of mapping between HDF–EOS and DODS data types
Task Total: 4.5 months

NCSA/HDF 5:NCSATask 3

Year 2

4 Qtr Implementation of HDF 5 API with DODS protocol
Task Total: 12 months

Year 3

4 Qtr Report on maintaining, troubleshooting, and providing user support for the HDF, HDF-EOS and HDF 5 implementations
Task Total: 6 months