A Thematic Data Portal to Satellite-Derived Ocean Surface Properties: Discovery and Access

by

Peter Cornillon	— University of Rhode Island
Tim Armour	— JASON Foundation for Education
Gloria Brown-Simmons	— University of Rhode Island
Andries van Dam	— Brown University
Ben Domenico	— Unidata/University Corporation for Atmospheric Research
James Gallagher	— Open source Project for a Network Data Access Protocol
Steve Hankin	— Pacific Marine Environmental Laboratory
Marty Landsfeld	— New Media Studio, Inc
Rob Raskin	— Jet Propulsion Laboratory

Submitted to

Earth Science Research, Education and Applications Solutions Network
CAN-02-OES-01

REASoN
NASA Peer Review Services Suite 200
500 E Street SW
Washington D.C. 20024-2760

July 30, 2003

Contents

Abstract

A collaboration of seven non-profit institutions and one NASA DAAC led by the University of Rhode Island (URI) propose a five year effort in research and educational areas of interest to NASA. The centerpiece of the proposed effort is the development of a Thematic Data Portal to satellite-derived ocean surface properties. The ocean surface data portal will consist of a suite of five thematic data portals corresponding to sea surface temperature, surface winds, sea surface topography, ocean color and surface precipitation, all routinely measured from satellite-borne sensors.

Technology development activities are proposed which extend the data access protocol with which the data supported by the portal will be accessed. These extensions will facilitate metrics reporting, and will provide a flexible, community-driven mechanism to insure that complete and consistent semantic information is available for diverse, heterogeneous data repositories.

The educational component will focus on making use of data available via the thematic portal in a suite of educational settings ranging from the classroom to popular educational web sites to aquariums. A particular focus of this portion of the project will be on immersive displays that bring the data to life in high traffic public settings. The technology component will involve enhancements to the basic data access infrastructure that will facilitate the tasks associated with the research and educational components.

An open source demonstration project is also proposed. This project will demonstrates the reuse of existing software components, creating gateway applications to provide environmental and public health data to GIS users.

1 Introduction

In response to NASA's REASoN CAN[1], a collaboration of seven non-profit institutions and one NASA DAACs led by the University of Rhode Island (URI) propose a five year effort in **research** and **educational** areas of interest to NASA. The research and educational components of the proposed effort will be supported by a **technology development** component. The effort proposed herein continues work begun by the Distributed Oceanographic Data System (DODS) Earth Science Information Partner (ESIP)[2] under NASA CAN-97-MTPE-01 funding. Each of the three components (research, education and technology development) as well as the relationships between them are detailed separately in subsequent sections. In the remainder of this section we provide a general overview of the project.

1.1 The Data Portal

The centerpiece of the proposed effort is the development of a thematic data portal to satellite-derived ocean surface properties. The ocean surface data portal will actually consist of a suite of five thematic data portals corresponding to five oceanographic parameters routinely measured from satellite-borne sensors: surface temperature, surface topography, near-surface ocean color, surface winds and surface precipitation. Key data sets in each of these areas will be targeted for each portal Each of the thematic portals will follow the same design. They will all be web based, viewable from a standard browser. Each will include the following:

- A description(s) of the basic satellite observation(s); i.e., a brief tutorial on the basic measurements and typical algorithms used to retrieve the geophysical parameter of interest from the satellite observations.
- A list of data sets associated with the portal. For each data set on the list there will be:
  - A description of the data set,
  - A list of references related to the data set, and
  - A discussion of special considerations that differentiate this data set from others on the list.
- Web access to each of the data sets listed in category 2 below via a modified version of the Live Access Server (LAS). LAS will allow the user to not only view the sections in the data set via her/his browser, but also to view the differences between sections from two different data sets.
- Links to public domain tools (e.g., Ferret, GrADS, ncBrowse, IDV, etc.) capable of accessing the data over the web.
- Downloadable Matlab and IDL[3], toolkits, specialized for the data sets described on the portal, that allow users of these packages to access the data directly over the network.

[1]All acronyms are expanded in the Acronym/URL List beginning on Page 27. Items that appear on the Acronym/URL List appear in a slightly different font; if the word looks like THIS it appears on the Acronym/URL List.

[2]http://www.esipfed.org/partners/index.jsp?type=Type2

[3]Commercial analysis packages.

In addition to a browser visualization capability via LAS, all of the data sets associated with a given portal will be directly accessible via the OPeNDAP Data Access Protocol (DAP) (Section 2); i.e., they will all be accessible via DAP-enabled clients such as those listed above. In fact, it is this access that permits LAS to acquire the data for visualization – the vast majority of the data do not reside at an LAS site or at the project's central site, they generally reside at an archive or at the data producer's site.

Supported data sets may be divided into five general categories. These in order of importance to this project are:

1. *Higher level satellite-derived products* – Levels 3[4].
2. Derived products – wind stress curl, sea surface temperature (SST) fronts, etc. – Level 4
3. In situ calibration and/or validation data sets.
4. Other gridded products obtained from numerical models or in situ compilations.
5. Low level satellite observations – Levels 1 and 2.

By far the greatest emphasis will be on data sets in category 2; i.e., funding has been specifically identified to guarantee that a broad range of higher level satellite-derived data products are accessible in each thematic area. Products in the other categories will be included where readily available. More on this in Section 3.

1.2 Why Thematic Data Portals?

Satellite observations are an under-utilized data resource for much ocean science research. In today's research environment an ocean scientist requires specialized knowledge about the satellite measurement and data analysis processes to take advantage of this resource. (S)he must understand the strengths and limitations of the instrument types – the precision, biases, orbits, spatial and temporal resolutions, limiting factors (such as cloud cover), factors that lead to uncertainty (such as atmospheric humidity). The data sets themselves arrive in a baffling range of file formats, often with basic metadata such as coordinates absent from files. (S)he often must have specialized analysis and visualization tools that can properly geo-reference a bewildering collection of projections.

Coupled with these problem is finding the right data set in the first place. For example, a search on *sea surface temperature* at the Global Change Master Directory (GCMD) yields 701 data sets many of which are satellite-derived and many of which are accessible on-line in one form or another. The number of data sets identified for the other variables ranges ranges from 144 to 425.

All of these searches can of course be refined, but even after refinement there often remains a bewildering number of data sets from which to choose and, as noted above these are generally available in different formats from different sites and organized differently. The user wants to either be presented with the "best" product and/or with a clear description of how the products differ along with a capability to preview and compare them.

The Ocean Surface Thematic Data Portal will address these issues and more. The web component of the Portal will provide the descriptive material and the preview capability designed to facilitate data set selection. The access component will provide the tools needed to access the data themselves. This component is based on the OPeNDAP Data Access Protocol (DAP), coupled with an "aggregation server" (Section ??) and "ancillary information service" (Section 5.3) capability.

[4]We use the CODMAC data level definitions.

The research component of this proposal (Section 3) will focus on the development of the data portal, on tools to facilitate acquisition of the associated data from commonly used analysis packages and on tools that may be used to statistically compare the data sets in a web based visualization environment. The educational component (Section 4) will focus on making use of data available via the thematic portal in a suite of educational settings ranging from the classroom to popular educational web sites to aquariums. A particular focus of this portion of the project will be on immersive displays that bring the data to life in high traffic public settings. The technology component (Section 5) will involve enhancements to the basic data access infrastructure that will facilitate the tasks associated with the research and educational components.

Prior to the details associated with the various project components we provide an overview of the OPeNDAP Data Access Protocol (DAP) and the current state of the DAP.

2 The DAP

As noted in the Introduction, the work proposed herein is an extension of work undertaken by the DODS ESIP[5] and continued with National Oceanographic Parternership Program (NOPP) funding[6]. We therefore begin the discussion with a brief overview of OPeNDAP, the organization that has assumed responsibility for the DODS Data Access Protocol.

The Open source Project for a Network Data Access Protocol was established in 2000 for the development and promotion of software that facilitates access to data via the network. OPeNDAP was created to address a growing interest among non-oceanographic disciplines in the Distributed Oceanographic Data System (DODS) data access protocol together with the concern among these same groups that the primary focus of the DODS effort is on oceanographic data system needs. OPeNDAP has a broader mandate than DODS in that it focuses on network access to data in general. Particular attention within OPeNDAP is on the development of a robust data access protocol (referred to as the DAP hereinafter) that addresses a wide variety of data needs.

The DAP is implemented using a multi-tier architecture and features a modular server-side design. On the client side, several different application programming interfaces (APIs) are provided which can be used to interact with the servers. These may be linked into an application thus enabling it to access remote data via a URL. While OPeNDAP supports its own API based on the object-model used by the DAP, OPeNDAP and Unidata also support two versions of the netCDF API which may be used to communicate with servers. The netCDF APIs provide a way for an application program which can read data using netCDF to interact with DAP servers in the same way that it would normally access local data. For example, a request for data from a remote server:

http://dataportal.ucar.edu:8080/dods/pcm-data/temperature?selectheight("sealevel")

returns the temperature at sea level from the OPeNDAP server that this URL is directed toward. OPeNDAP servers provide subsetting and format translation of managed data in response to client requests. Servers for the following data formats or APIs are currently available: netCDF, HDF, HDF-EOS, JGOFS, Matlab, FreeForm, CDF, FITS, CEDAR, GRIB, BUFR and relational data bases accessible via SQL. FreeForm and JGOFS servers have been designed to deal with data sets stored in a non-standard format; i.e., home-grown formats.

The DAP provides its users with a rich set of primitive data structures, structures that may readily be assembled into larger data objects.

The popularity of this architecture is due to several factors: It transparently extends data access APIs without application software modification. It provides servers for several

[5]See http://www.unidata.ucar.edu/packages/dods/archive/proposals/can-97-mtpe-01-html/ for the funded DODS ESIP proposal.

[6]See http://www.unidata.ucar.edu/packages/dods/archive/proposals/nopp-html/ for the funded NOPP proposal.

dataset formats and is readily extensible to support new ones. It enables applications to access subsets of datasets thus minimizing unnecessary network data movement, and it is deployable within existing internet architectures. The DAP as implemented originally in DODS, was designed with heavy emphasis on generality and relied solely on HTTP as the transport protocol. Although HTTP still remains the primary transport protocol for the DAP, an Earth System Grid – II (ESG II) funded project at the National Center for Atmospheric Research (NCAR) is currently extending the DAP to allow for FTP with the GRID enhancements (GridFTP) transfers as well.

2.1 DAP usage

DAP-enabled servers are accessible at approximately 30 sites in the U.S. and an additional 10 sites in Europe, Asia and Australia. These servers provide access to more than 300 data sets. A list of known sites and associated data sets is available from the DODS homepage at: http://unidata.ucar.edu/packages/dods. A pointer to the GCMD DODS portal may also be found on this site. Approximately 2/3 of the data sets listed on the DODS site are described in the GCMD. Metrics of DAP accesses are available at

http://www.po.gso.uri.edu/tracking/metrics/metricsmatrix.html

Usage across the system has been climbing steadily reaching close to 1 million DAP accesses in October 2002. Some care must be taken in interpreting these statistics since scripted accesses are possible. Sites showing a large number of accesses also have a large number of scripted accesses, a program that may come in and acquire several thousand data granules. In addition, these metrics also record metadata accesses. Approximately 15% of the accesses are metadata accesses. The number of users exceeds 400 per month.

An interesting statistic obtained from the Columbia University site is that only about 10% of the web access to their site were DAP accesses for the first quarter of this year, but on the order of 65% of the volume of data moved were from DAP accesses. This is particularly interesting since DAP accesses are generally for a specific subset of data rather than an entire file.

When evaluating these statistics it is important to keep in mind that most DAP accesses are not from someone clicking links in a site, but rather they from explicit accesses often constrained for a specific subset of the data.

3 The Research Component

3.1 Targeted ESE! Research Components

The research proposed herein targets the second research area, "projects contributing to interdisciplinary or process studies". We do not propose gathering variables for a particular interdisciplinary study. The product that we propose is facile access to a suite of variables that are fundamental to many of the interdisciplinary questions cited in the ESE Research Strategy [1]. In addition to these variables, we also propose access to the tools needed to fuse these variables in a meaningful fashion. Although the access that we propose will lighten the burden on the researcher focusing on problems within a discipline, it will be much more useful for researchers working on interdisciplinary problems, problems requiring data from different sources in different formats organized differently from site-to-site.

Although the research portion of this proposal targets the second research area, we believe that it may also play a significant role in the first area, "projects contributing to systematic measurements". In particular, we believe that for each of the variables targeted in this proposal there will likely be a number of systematic data sets produced, some by NASA funded efforts and some by other non-NASA funded efforts. In fact, this is true of some of the variables today. This is not surprising given that most of the retrieval algorithms are tuned in one fashion or another and that in so doing choices must be made. There will inevitably be differences of opinion with regard to these choices and the "correct" choice will almost certainly depend on the use to be made of the data. Consider two choices made by the PI for the NASA funded 1 km Pathfinder SST project[7]. The first choice was to use the coefficients for the 9.28 km Pathfinder project for the regional reprocessing. An alternative would have been to tune the retrieval algorithm to the in situ data for each region. The r.m.s. difference between in situ observations and the satellite-derived values would have been smaller with local tuning, but integrating these data with those from the global Pathfinder effort would have been more difficult. For example, there would have been a mismatch at the boundary of the high resolution data for those interested in using the global data to obtain a basin-wide view of the field with a high resolution coastal component. Second, the decision was made to use the Cayula-Cornillon cloud screening algorithm [2]. This algorithm tends not to reject as many cloud-free pixels in the vicinity of ocean fronts as the cloud clearing algorithm used on the global Pathfinder data, but it also tends to include as cloud-free more cloud-contaminated pixels. The data screened with this algorithm are more appropriate for use in studies of ocean fronts than in studies where accurate estimates of heat flux is important. Processing the entire time series from the three different sites has been a substantial effort, but with three doublings in computer power over the course of the project proposed herein, it is entirely possible that during the proposed project someone will reprocess the entire time series again, making different choices. We believe that convenient access to all such products is desirable together with a mechanism to compare them so that the user can make an informed decision about which product to use. In fact, a high priority data need by the data assimilation community is for uncertainty estimates associated with the variables of interest (Stammer, personal communication). The ability to quickly compare two data sets statistically does not necessarily provide an absolute measure of the uncertainty but it does allow the assimilator to first determine if the data sets appear to differ substantially and secondly to estimate the uncertainty associated with using one compared with the other. Furthermore, if the two data sets are from sensors measuring in different portions of the electromagnetic spectrum or using substantially different retrieval algorithms from different sensors, statistical differences between the data sets will likely bound the error associated with the retrievals of either one. The point is that in addition to stimulating the development of more systematic data sets, ESE researchers will also benefit from consistent access to the data sets that will likely be developed from other sources of funding.

The Data Portal will focus on five variables and higher level products obtained from these variables. The variables are: sea surface temperature, sea surface topography, ocean surface winds, ocean color and ocean precipitation. In the following the common elements of each subportal are described. This is followed by a discussion of the population of the system.

3.2 Data Portal – The Web Page(s)

The web pages associated with the Ocean Surface Data Portal consist of a hierarchy. At the highest level is the site web page: `http://oceanographicdata.org`[8]. Beneath this are the five web pages associated with each of the supported variables. In addition to links to these web pages, the main Portal web page will contain links to the GCMD DODS Portal and to the (**NVODS!**) LAS site; i.e., links to the full suite of OPeNDAP-accessible data sets documented in the GCMD and supported by LAS.

Each of the web pages associated with one of the thematic areas will contain:

- A description(s) of the basic satellite observation(s).

[7] All 5-channel HRPT data from the Wallops Island receiving station, a receiving station at the University of Tokoyo and one in California for the periods 1982-2001, 1985-2000 and 1992-2000, respectively, are being processed at URI with Pathfinder funding.

[8] URI has purchased this site for this project. We also own `http://oceanographicdata.com`.

- A description of each of the supported data sets.
- Web access via a visualization tool to each level 3 or 4 data set listed.
- Links to public domain tools capable of accessing the data over the web.
- Downloadable `Matlab` and `IDL` toolkits, specialized for supported data sets.

The data set descriptions will be arranged in a hierarchy that will allow the user to quickly move to the data of interest. Differences between each level in this hierarchy will be described as will differences between data sets within a level. For example, `SST` products will be available from `AVHRR`, `MODIS`, `TRMM` and possibly `GOES`. Level 1 data from these sensors will be grouped as will Level 2, 3 and 4. The difference between the data in one level and that in an adjacent level will be described. Within a level there will be products from several sensors. The difference between each of the products will be described and the one (or ones) thought to have the lowest r.m.s. variability when compared with in situ data will be flagged – the "best pick" of the level. Depending on the number of data sets, additional levels may be added within a level. For example, coarse resolution Level 3 data will be at a higher level than higher resolution Level 3 data from the same sensor. An attempt will be made to present the hierarchy as a point-and-click tree.

Each web page will also contain a link to an `LAS` configured for that web page; data sets will be arranged in the `LAS` data set list in the same hierarchal structure as they are presented on the web page. See Section 5.5 for a description of existing capabilities as well as capabilities to be added.

The thematic web pages will also serve as a mechanism to gather user feedback associated with the data sets supported by the site. This list will be monitored by the project. Appropriate messages will be posted at the site and will be organized to facilitate subsequent access. Comments associated with a particular data set will be accessible as part of the metadata associated with that data set. In addition, comments related to a suite of data sets will be associated with the element of the data set tree on which these data sets reside.

The five web sites to be developed will be maintained for the duration of the project. At the end of the project they will be transferred to the appropriate `DAAC`, the `PO-DAAC` for winds, surface topography and `SST`, and the Goddard-`DAAC` for ocean color and precipitation.

Table 3.2 provides the schedule for which the various thematic sites will become operational. The times in Table 3.2 refer to the time when emphasis will switch from one variable to another and are not meant to suggest that population of the "completed" site will cease at that time.

Table 1: Dates that the various thematic data portals will be come operational assuming a 1 June 2003 start date

Parameter	Operational
`SST`	June 2004
Surface Winds	March 2005
Ocean Color	December 2006
Surface Topography	September 2006
Precipitation and Ocean	June 2007

3.3 Data Portal – Consistent Access

The main focus of the work to be undertaken in this project will be to provide consistent, semantically complete access to the data sets associated with the Portal. This will involve one or more of the following three tasks for a given data set.

3.3.1 Data Server Installation

Access to data via the Portal will make use of the `OPeNDAP` Data Access Protocol. This provides for format independence in that the `DAP` transforms the data from the format in which they are stored to the format requested by the applications program. This transformation is actually performed in two steps: (1) from the format in which the data are stored to the `DAP` data model, and (2) from the `DAP` data model to the format expected by the analysis package. See Section 2 for a list of `DAP`-enabled clients and servers currently supported.

There already exist servers for a large number of data sets. This portion of the project will initially focus on the addition (read help in the installation) of servers at sites supported with funding as part of the effort proposed herein. `OPeNDAP` support will however not be limited to these sites. For example, a concerted effort will be undertaken to help install servers for the `GODAE` High Resolution Sea Surface Temperature Pilot Project (`GHRSST-PP`)[9].

3.3.2 Aggregation Server Installation

Format independence does not guarantee organizational consistency of the data sets. The distinction between the two is most easily demonstrated by example. Consider a data set that consists of a large number of `SST` fields all in the same projection each representing a different time. At one site, the data might be stored as a number of 2-dimensional (longitude, latitude) files, one per time step, while another site might archive the same data as one 3-dimensional (longitude, latitude, time) file. The data at both sites could be stored in the same format for example `netCDF`. The fact that the data are delivered directly to the application package in a consistent format, for example as a `netCDF` file, substantially reduces the complexity of client side applications as well as the metadata needed to describe the data, but the lack of structural uniformity is still a substantial burden on the clients. For example, in the multi-file 2-d case cited above, the client must deal with an inventory system of some form, while in the 3-d case this is not necessary. In addition to the added complexity that must be built into the client, additional metadata are needed to describe the inventory and associated 2-d data objects in the first case and the 3-d data object in the second.

Experience with the `DAP`-enabled data sets currently being served suggests a broad range of organizational structures for the data of interest to this project. Structural differences range from the order of array indexes, which we refer to as a semantic difference, to the organization of files within the data set, which we refer to as a syntactic difference. An Aggregation Server (`AS`) has been developed by `Unidata` that will combine fields in a multi-file gridded data set in which all of the fields are in the same projection at the same resolution into one n-dimensional object – syntactic aggregation. The `AS` supports several different forms of syntactic aggregation, but not all of the forms encountered to date. *One of the technological development tasks that will be undertaken for this project will be to extend the* TRL
capabilities of the `AS`. This will be done as the data sets with structural problems that can $5 \Rightarrow 6$
not be addressed with the current `AS` are encountered.

For all supported multi-file data sets for which aggregation makes sense, an `AS` will be installed and configured. Data that are not in a standard earth projection will not be aggregated in this way.

[9] See supporting letter from Craig Donlon.

3.3.3 Generation of *use* metadata

When one thinks of metadata they generally think of the metadata that are required to locate a data set in a data set directory service such as the `GCMD` rather than the metadata that are needed to actually use the data. For example, the user does not need to know the range of a data set to make use of the data, but this information is needed to find a data set of interest from a large number of data sets. We refer to the metadata required to make use of the data as *use* metadata. A consistent set of use metadata is absolutely essential to the effective use of the data themselves. The most common use metadata fields are related to the following characteristics of the data:

Variable name. Variables in `DAP`-accessible data sets have the names given to them by the data provider. These names are often idiosyncratic in nature with little meaning to anyone other than the individual who generated them.

Units. Data are often delivered in scaled geophysical units – e.g., $8 \times$the value delivered-4 equals `SST` in °C – or in a system of units with which the user is not familiar or would prefer not to use – furlongs per year (OK, OK, that one might be out there, how about miles per hour?)

Flags. For most measurement based data sets there will be data values in the data set that are flagged for one reason or another, missing value, bad value, land, cloud, etc. In some data sets these are flagged in fields associated with the variable of interest while in others the flags replace the data values. For such data sets it is important to know what the data flags are and their meaning.

The metadata fields associated with each of the above define a transformation of the received value to a new value. For example, the variable *COADSX* might be longitude and the fields for which no data were collected might be flagged with -99. A possible set of transformations in this cases would then be: *COADSX* $\Rightarrow$ longitude and -99 $\Rightarrow$ *Missing value*. Other transformations are however possible; e.g., -99 $\Rightarrow$ *nan*. These transformations may in fact be defined in some of the data sets. Cooperative Ocean-Atmosphere Research Data Standard (`COARDS`)-compliant `netCDF` or Hierarchical Data Format - `EOS` (`HDF-EOS`) data sets generally provide a set of transformations important to the data user.[10] Even in the case of data sets that contain this information there is no guarantee that the mapping is to the same domain from one data set to another. For example, the mapping to a "long name" might be defined but different "long names" might be used from data set to data set.

This portion of the project focuses on the problem outlined above. In particular, target domains that are consistent across all of the data sets will be selected for each *use* metadata parameter. For example, for units, `mks` will be used throughout. Specifying the target domains defines how to fill the various metadata fields. For each of the supported data sets, the metadata that describes the transformation to the target domain will be determined. Because the data provider does not in general want to modify his or her archive with new use metadata, the `DAP` is currently being modified to support access to metadata associated with a data set from sites other than that of the data set. We refer to this capability as the Ancillary Information Service. Associated with the Ancillary Information Service (`AIS`) are `AIS` servers and archives. These can be installed anywhere. One will be maintained by this project for the data sets supported by the Data Portal. This is where the use metadata will be held.

[10] A commonly held mis-perception is that this information resides in Federal Geographic Data Committee (`FGDC`) descriptions of data sets. This is in general not true – `FGDC` descriptions of data sets rarely contain *use* metadata although there are field locations within which the standard to encode it, they simply are not mandatory. This is likely because `FGDC` has been used mainly for data location rather than for data use.

See Section 5.3 for a description of the technology development effort that will be undertaken to further address organizational issues with regard to the data archives.

3.3.4 Putting It Together

The AIS, providing for a consistent semantic view of the data across the entire range of supported data sets, with Aggregation Servers providing for a common structural representation for the data, and the format conversion capability of the DAP translate to a consistent syntactic and semantic view of all Level 3 and higher data sets supported by the Portal. Taken together these services will also greatly facilitate the development of client applications accessing supported data sets. For both Matlab and Interactive Display Language (IDL) we will develop toolkits that can be used with the supported data sets in each of the corresponding analysis packages. These toolkits may also be used as the base for more sophisticated interfaces in the analysis package. We already have a Graphical User Interface (GUI) for Matlab. The toolbox will allow for all supported data sets to be made accessible from the GUI with virtually no effort. Both the Matlab and IDL toolboxes will be made accessible via the Portal.

The steps outlined above will provide the data in a syntactically consistent form with the use metadata needed to manipulate the data with meaning.

3.4 Data Portal – The Data Sets

Level 3 and higher data sets that will be supported for the Portal will be selected in one of three ways. First, the PI has identified some of the more significant sites for satellite-derived ocean data products and arranged with a scientist or a data archivist at each of these sites to joint the project either as a Project Member (PM) or to work with the project as a subcontractor[11]. The data sets at these sites for each of the variables will form the base of the list to be considered for Portal support.

Second, a search will be made for Level 3 and higher DAP-accessible data sets (showing the appropriate surface parameters) and these will be added to the list of data sets. As an aside we note that there are order 25 DAP-accessible Level 3 or higher SST data sets, a similar number of surface wind data sets and fewer for the other variables.

Finally, a broad search will be made for more data sets of potential interest. This search will begin with a search of the GCMD holdings for data sets with the appropriate parameter. The discovered data sets will be added to the running lists. We will build on this list through our knowledge of the field, via web searches and discussions with colleagues. We will then pare this list down to one that includes only those data sets that we believe are good candidates for the Portal - high quality, broad interest, etc. The next step will be to vet this list with experts in the field via e-mail to determine which data sets are the most appropriate for inclusion in the program. We will then identify and contact individuals responsible for these data sets and investigate with them the possibility of installing servers at their site. Finally, these data sets will be described for the Portal web page and where appropriate metadata will be determined and data and Aggregation Servers installed.

In parallel with the search for Level 3 and higher data sets, a search will also be made for validation data sets. As with the gridded data sets, this list will be pared down and vetted. An example of the sort of data sets that we are looking for here is the University of Miami Pathfinder Matchup Data Set. Those selected will be supported on the Data Portal the same way that the Level 3 data are.

[11]The choice between participation as a PM or as a subcontractor was made based on the level of expected participation. Where the individual was simply going to perform a service, install and maintain servers, a contractual arrangement was made.

4 Education

The problem we are proposing to provide a solution for is one of systematic, continuous, access to a rich pipeline of ocean data for public education. Creative education teams working on ocean related topics are asking the question "what ocean data shall we select and how do we use it" during the program design process. The next question is usually "how do we get access to ocean data?" These teams, consisting of exhibit designers, architects, educators, computer graphics experts and hybrid media producers are usually collaborating with an oceanographer. Although there are several installations using remotely sensed Earth science data in exhibitions, it is difficult and costly for most to do so. We are proposing to lower the cost of, to optimize and to ease the integration of ocean data into educational venues in ways that will reach millions of people.

Our team includes "educational users" from the leading companies in tele-presence events, new technology and ocean education, the JASON Foundation for Education (JASON) and the New Media Studio (NMS) (in association with) Planet Earth Science, who have close relationships with oceanographers using in-situ exploration methods to gather their primary research data. The educational products from these companies (classroom materials, teacher training courseware, Web content and immersive events) have been the most successful in communicating scientific results from NASA as well as giving the audience a great sense of "mission" exploration. Another essential "user" on our team is the firm whose architects have pioneered aquarium design – Chermayeff, Sollogub and Poole, Inc. (CSP)[12]. To date, six of CSP's aquariums, including the National Aquarium in Baltimore and the New England Aquarium in Boston, have had over 130 million visitors.

We have structured a unique project to develop a "superstructure" that will operate between these educator's target mediums and ocean data using OPeNDAP.

This superstructure, called the Digital Ocean, will be the backbone that educators will use to deliver ocean data via their vehicles for large scale dissemination of ocean science education that has the potential to reach millions of school children and the lay public. As part of our proposed project, our collaborators will develop three products: first, a series of supplemental educational materials for students and teachers, second, a series of interactive Virtual Reality (VR) Web pages for the general public and third, the Digital Ocean system – a set of tools to interactively choreograph ocean focused informal educational productions for immersive, and other large format, theaters. The third product will include three ocean data intensive "pre-programmed" productions that will be used in aquariums as Digital Ocean immersive theater "shows".

Thus, the education component is a collaboration of oceanographers, educators, architects, exhibition designers, system experts, and visualization specialists. It draws from the technical expertise and resources of their respective institutions: OPeNDAP, University of Rhode Island, the JASON Foundation for Education, New Media Studio, Brown University, and an association with the designers of major international aquariums and Earth science centers - Chermayeff, Sollogub and Poole, Inc. (CSP).

During the first two years we will focus on sea surface temperature, and ocean winds. We will also develop several methods to visualize surface dynamics phenomena such as "cat paws." During the second two years, the parameters for sub-surface depths will be added along with dynamically changing structures such as equatorial upwelling and techniques to visualize three dimensional volumetric data representations. The data and associated processing will become a part of the systematic "Digital Ocean pipeline" superstructure feeding the education venues. During the last years we will explore other datasets as well as continue to refine the tools in the Digital Ocean pipeline as we learn more from the events.

By being based on the underlying science, learning theory, as well as state-of-the- art technology, our team intends to engage imaginations in inspired and perceptive new ways.

[12] See supporting letter from Peter Chermayeff.

4.1 Providing an Educational Context

JASON will encapsulate the Digital Ocean data into educational materials to provide live data from a variety of sources for use, not only by science professionals, but also by students and teachers in middle grade (4-9) classrooms. JASON will be responsible for developing the educational context of this data, providing educational background, exercises and teacher guides, all linked to specific national and state science standards. JASON has fourteen years of experience in bringing real science as it is happening to students and teachers through telepresence technology, written curriculum, videos, the web, and through online and in-person professional development for teachers.

The JASON Project, the main program of the JASON Foundation for Education, now serves over 1,000,000 students and 25,000 teachers in all 50 of the United States, Canada, Mexico, Bermuda and several European countries and Australia. JASON programs are developed as standards-based, hands-on activities and experiences in an exploration paradigm. JASON will incorporate the data to be made available from this project into web-accessible material, with downloadable teachers' guides and opportunity for collaboration among teachers and students who are geographically and demographically diverse.

JASON, along with NMS, will incorporate the Digital Ocean data into web-based applications that are independent of the JASON Project and readily accessible to both the JASON Project participants and other interested students and teachers.

4.2 Web Browsing For The General Public

New Media Studio will develop the Digital Ocean techniques to dynamically update three "canned productions" and present them as Web based virtual worlds. The data stream incorporated into the 3D virtual scenes will be dynamically updated via the proposed Digital Ocean pipeline superstructure built on OPeNDAP.

To facilitate this data stream, virtual worlds will serve as the core visualization structure onto which near real time ocean data sets will be experienced. These re-usable virtual world models will be created with Alias/Wavefront Maya to be utilized within Macromedia's Shockwave. The ability to embed Maya 3D information into a "shocked" Web page lies at the heart of the Digital Ocean Portal, making the pages accessible by the majority of the lay public browsing the Web. Users will be able to download the Shockwave "movie" which will "dynamically update it's frames" as it will have the capability of retrieving the most currently available data via the Digital Ocean interface to OPeNDAP and "draping" the data over the 3D world structure.

The dynamic feature is made possible by employing the Earth Data Multi-media Instrument (EDMI) plug-in to Shockwave. The EDMI will be licensed through Planet Earth Science and gives the Shockwave projector all the functionality of IDL. With IDL as a background engine, running on the users computer, the user will have the ability to utilize the OPeNDAP servers and download data, minimizing the download process and speeding up the data transfers to the user. The EDMI will be able to download the most current data sets and convert them into Shockwave assets that can then be utilized in the 3D environment and interactively explored at the users will.

1) View from Space: The user will view a 3D globe of the earth from space and be able to select from different data sets and see them applied to the globe. The graphical interface will look like a space-vehicle with instrument panels for both data selection and for maneuvering the ship to different positions around the Earth. The user will be able to maneuver to any point to examine the data type both visually and through sampling of the data as if they were scientists on a mission with real instruments.

2) View from the Ocean's surface: The user will have a vehicle designed to ride on the ocean surface to view the data displayed around them. Sea surface height, for instance would dictate ocean height around the user and sea surface temperatures could be mapped onto the sea surface. Alternatively, a thermometer can be provided on board the vessel which would

allow the user to take readings of sea temperature at any given location. Transparency will be used on the sea surface to give a more photo-realistic immersive effect. An ocean floor can be included which might be visible in shallower parts of the ocean or above continental shelves. The vehicle used will also have control panels for navigation and for selecting data.

3) Under the sea: The user will be able to navigate under the ocean in an under- water research vessel equipped with navigational controls. Light diffusion and other spectral effects will be employed to give the feeling of being under the ocean. There will be the ability to either turn on data in the surrounding water (ie. change the water color to reflect temperature) or to have a heads-up-display which shows data relating to the waters surrounding the user and/or the ocean floor below.

4.3 The Digital Ocean – Ocean Data Dynamically Used in Aquariums

There are several research VR environments focused on studying the display technologies relevance to ocean science. Of note are Brown University's CAVE for scientific visualization, the Scripps/UCSD CAT-2 project; Old Dominion University's VR project Chesapeake Bay and their newer CAVE6D. These installations effectively display data for small audiences (2-50 people). We are proposing to develop a Digital Ocean pipeline to drive productions for audiences of 300-400 people many times a day in many public aquariums.

The Digital Ocean pipeline is an advanced operational system of tele-processes capable of near real-time data assimilation, multiple display formats with changeable projection geometry. It is conceived as a seamless access to ocean data via the OPeNDAP data access protocol to give aquariums and other ocean science education venues an easily accessible data rich pipeline. Currently, the JASON Project's immersive theater events are fantastic live televised underwater explorations. Working along with Brown University's Graphics Group to tap into the proposed Digital Ocean data pipeline, JASON proposes to incorporate selected elements of the data and data analysis parts of the project into their regular annual JASON Project (a tele-presence event) where applicable. It will be used, in addition to the JASON Project, to drive other new experiential venues at public aquariums.

One prime aquarium candidate for receiving the Digital Ocean pipeline is the CSP Ocean Forum, an advanced networked tele-structure. "...is a space devoted to observation and manipulation of ocean information, a place where climate change, ocean surface dynamics, ocean health conditions, animal migrations, current events, and myriad other subjects can all be accessed, questioned, debated and explored. Data retrieval from public agencies and myriad private sources via the Internet and global monitoring systems will be state of the art, continually updated and expanded, and utilize advanced methods of visualization, including computer enhanced images, maps, projected video and photography, and digital modeling... ...The Forum will be a place of dynamic science, accessible to all, a place for full engagement of the public in the ocean, for the ocean, and for all of life on Earth."

CSP proposes to demonstrate the system in their Ocean Forum as a source of continuous ocean data programming in addition to providing audiences and study groups with "canned" programs that highlight key concepts supported by the world of ocean data. The Digital Ocean system will be designed to systematically provide the Ocean Forum with ocean data, scientific visualization of data, and in diverse media applications including utilization of digital images including pre-recorded as well as down loaded live real-time video, for continuous display in several interactive modes. "CSP contemplates dialog occurring in the Ocean Forum between scientists and observers in the field (including those working underwater) and an aquarium based audience, calling on streams of data as needed to address chosen subjects. Dialog is also contemplated between scientists and audiences at sister institutions such as a network of public aquariums."[13]

[13]The Visitor Experience – A Walk-through of the New Bedford Oceanarium, CSP. The information con-

Projects currently being developed by CSP that offer an Ocean Forum possibility within three years include the New Bedford Oceanarium, in New Bedford, Massachusetts which has ties to ocean science at the University of Massachusetts, and the Tucson Aquatic Center in Tucson, Arizona, which plans to have ties to ocean science at the University of Arizona. CSP has master planned a new National Museum of Ocean Science and Technology for Taiwan, where a major application could occur, and has planned an expansion for the Lisbon Oceanarium in Portugal where a planned special auditorium will include an Ocean Forum. CSP is interested to include the work being done at NMS and Brown (i.e. simulations and visualizations) and JASON Project events and technology. Tele-presence events would occur together with JASON events, as well as independently, perhaps in concert with national and international conferences.

Brown University, with contextual guidance from JASON, NMS and CSP, will prototype immersive experiences based on the ocean data pipeline, develop techniques for merging data visualization models with real scenes, as well as design and implement the display geometry solution for the immersive and large format theaters. Multiple visualization and user interaction challenges must be solved in order to create the immersive components of the proposed Digital Ocean pipeline superstructure.

The Brown University Graphics Group has two decades of experience researching and developing interactive 3D computer graphics. Brown will build on their experience developing user interfaces and visualization techniques for scientific visualization in immersive visualization environments as well as extensive experience building electronic books. Furthermore, Brown University's visualization and computing infrastructure housed at the Technology Center for Advanced Scientific Computing and Visualization (TCASCV) consists of multiple display configurations ranging from the conventional desktop to PowerWalls to a fully immersive Cave–this capability will help us more quickly converge on the most effective display formats for different tasks. Brown will also contribute in terms of state-of-the-art rendering for immersive environments. Brown's 4-node commodity cluster drives their Cave at about ten-times the performance of out-dated SGI Onyx2 hardware and their upcoming 48-node IBM visualization cluster promises significantly higher performance.

The proposed Digital Ocean pipeline superstructure built upon the OPeNDAP data acquisition backbone is the most forward-looking educational data delivery technology infrastructure to date. By easing the delivery of data and value added data streams for educational venues, educators and production designers can think seamlessly about the ocean and the data from a host of remote ocean sensing instruments. They can focus better on the context and not how to build a system to get the data. This advanced approach to public education will truly allow the audience to get "their mind into" the content.

5 Technology Development

5.1 Maintenance of the OPeNDAP software

TRL $5 \Rightarrow 7$

When OPeNDAP's work started, it was called the *DODS project* and the total code base was small. There was a C++ toolkit, two servers a single client and a handful of Perl software. Now the code base uses Java, C++, Perl, Tcl/Expect, the CppUnit test framework and others; it consists of two libraries, seven servers, four clients and four general-use server

tained in the pages of this proposal constitutes a trade secret and/or information that is commercial or financial and confidential or privileged. It is furnished in confidence with the understanding that it will not, without permission of the offerer, be used or disclosed other than for evaluation purposes; provided, however, that in the event a contract (or other agreement) is awarded on the basis of this proposal the right to use and disclose this information to the extent provided in the contract (or other agreement) is allowed. This restriction does not limit the right to use or disclose this information if obtained from another source without restriction.

modules. We spend more time than ever before on software maintenance. To manage this increase, we have changed the way some tasks are performed. We follow the tenants of the Open Source software movement which exposes our software to more developers and we run routine nightly maintenance builds on multiple hardware platforms with multiple compilers. These practices reduce the overall costs of the major software maintenance activities, but the inescapable fact is that the success of the `DODS` and now `OPeNDAP` efforts has lead to an increase in effort spent on software maintenance.

OPeNDAP, as its name implies was chartered on the foundation of open source software development. While the original data access protocol was developed by a small number of software developers, and singly funded by the NASA ESIP Federation, over time the number of contributing software developers and sponsors for the technology has grown steadily. Alongside them is a burgeoning user and provider community, spanning multiple disciplines. Multiple communities have recognized the value of the underlying software technology and embraced it to meet their data access requirements.

To achieve a scalable solution for both software development and maintenance, we propose to implement the infrastructure required to make the OPeNDAP protocol truly open source. By that, defining a process to foster open participation in the evolution of the DAP specifications, and providing complete, and well-defined specifications for each of the DAP components so that contributing software developers can easily add to the suite of available clients, servers, and services available for OPeNDAP-enabled data. This REASoN CAN provides a unique, and timely opportunity to apply SEEDS principles for open source development and software reuse, to transform a software technology originating in the ESIP Federation and enjoying widespread acceptance into an open source effort, insuring it adheres to underlying SEEDS principles.

5.2 Metrics

TRL $3 \Rightarrow 5$

We have developed a prototype system for gathering metrics for our software. The testing and limited field use of this has taught us much about what really needs to be done to learn important information about how our software is used. There are three issues that must be dealt with by the metrics gathering software. First, we need to know both where our software is being installed and how each of those installations are used. Second, the software that gathers and reports this information must be non-intrusive and its use must be an active choice by each data provider. Finally, because there is a long time constant associated with the deployment of this type of software, most analysis of the data must be carried out by us using raw information gathered at the server sites. This means that the information gathered by the servers must be very basic but also fairly comprehensive.

We know that not every person who downloads our software winds up using it to serve data. Some people are just curious, others find that it does not meet their needs. One crucial statistic is the number and location of running `OPeNDAP` servers.[14] This provides us with information about the number of downloads versus actual use. We will use the installation script which now includes a prototype registration mechanism as a basis for a more robust scheme. The new design and implementation will provide a simple way for data providers to register their servers if they so choose.

We also need to know how the servers are being used. In the early days of the project, there were only a handful of servers running and it was possible to ask server administrators for log files and scan those to see important use patterns. However, now there are many sites where we have never had contact with the administrator(s) as well as sites with ten thousand or more data requests per month. Clearly, the collection and analysis of this data must be automated. We will build on our prototype system so that it provides easy access to basic information. In practice this information is very similar to what most web daemons currently

[14]It is possible to use a search engine such as Goggle to look for `OPeNDAP` servers, but since most of our servers use the `CGI` mechanism, it is likely that only fraction of these that exist will be found.

log. However, we have found that the logs maintained by web daemons vary in form and content, something that complicates their analysis. We will implement a simple logging subsystem specific to our servers. In addition, we will build on the remote access capabilities of the current servers so that, with the explicit permission of individual administrators, those logs may be accessed by OPeNDAP. The combination of logs which contain just information about our server and remote access to that information will enable us to see important patterns of use for individual data sources and the system as a whole.

It is crucial, however, that this system is implemented so server administrators are in control of their resources. The metrics sub-system will be designed so that by default these features are disabled and administrators can readily enable them if they wish.

5.3 The Ancillary Information Service

TRL $3 \Rightarrow 5$

The AIS is vital to the ongoing development of the OPeNDAP Data Access Protocol. We have experimented with a prototype version of the AIS. It provides the basis for creation of data sources which exhibit the uniformity needed for complex analysis and fusion of information. While there are other ways to achieve this goal, what makes the AIS special is that the original data stores do not require modification. Beyond this, the AIS provides a mechanism by which one data source may be mapped onto multiple metadata conventions, thus supporting use of that data source by several programs which have different semantic requirements about how data are presented.

We propose to expand on the current prototype AIS to build a full-fledged system by reimplementing it with three major additions. The AIS will be expanded to support additions to the variables held in a data source. The prototype supports only additions to the attributes present in a data source. We will also implement the AIS using the DAP version 3.0 software. The DAP 3.0 specification is currently under development.[15] We will also build a special AIS server which will enable users to publish AIS resources so that others can use them in the context of thematic collections.

The prototype AIS provides a way to add attributes to a data source. This is a powerful capability because missing metadata values can be added and existing values can be aliased to new names. This provides support for metadata conventions which rely solely on attribute information. However, while most (all?) search metadata standards fall into this category, most data *use* conventions also depend on certain 'well known' variable name and/or organizational schemes.[16] We will extend the AIS so that variables can be added to a data source and so that existing variables can be aliased to new names. Clearly, adding variables to a data source must be handled with the utmost care. The AIS will implement this feature so that all such additions are absolutely clear to the end user and/or are authorized by the provider of the original data source. These additional capabilities will greatly enhance the breadth of data sources and data conventions over which the AIS can be used.

We will change the prototype AIS so that it works with the DAP version 3 protocol that is currently under development. The DAP is an evolving protocol which has now been independently implemented by outside groups.[17] In response to requests and the continuing evolution of the Internet, version 3.0 of the DAP is being designed. This new version of the protocol will simplify reuse by groups other than OPeNDAP and will contain some features specific to the AIS.[18]

[15] This version is largely based on version 2.0, which we have implemented in both Java and C++. The new version uses XML as a persistent encoding for objects and contains many changes requested by the user community.

[16] By *well known* we mean known to a community of practitioners who use the standard or convention.

[17] Including Columbia University's *Ingrid* system, developed by Benno Blumenthal and PMEL's *ncbrowse* developed by Don Denbo.

[18] Although the AIS itself will actually be a separate protocol.

The AIS as envisioned at the January 2002 DODS developer's meeting contains a component not implemented by the prototype system. Part of the power of the AIS is that it provides a way for users to write local AIS files which augment a data source they use and then *publish* that information so that others can use it too. As part of this idea, special servers, called *AIS Servers*, will be developed to enable this publish capability. These AIS Servers will also serve as nodes which automatically bind AIS resources to data sources (normally the process is carried out in DAP-aware clients) so that even simple web browsers will be able to take advantage of the extra information available using the AIS. The AIS servers will function as a focal point for thematic collections of data sources and AIS resources. The AIS prototype lacks these AIS servers; we propose to build them as part of this effort.

5.4 Enhancements to the Grid selection function

TRL $4 \Rightarrow 6$

The Grid data type relates a single N-dimensional array with N map vectors. The data type provides a way to provide real-number indexes for array data. The OPeNDAP implementation of DAP toolkit contains a server-side function which provides a way to sub-set Grid variables using the values of the map vectors. We propose to extend this server-side function so that it is aware of the modulo nature of some types of map vectors.

The `grid()` server-side function currently uses map vectors in the simplest of possible ways. It does not understand that requests may 'wrap around' the ends of the vector in certain situations. This case occurs when the maps hold information about latitude or longitude, for example. Our extensions to this function will enable it to be used with relational expressions that, in fact, treat the map vectors as cyclic. This feature will greatly simplify the development of user software because the servers will be able to handle requests made using representations of the data as the users typically think about them, not as they are actually stored.

5.5 Live Access Server

5.5.1 The Need for a Web-based Satellite Evaluation Environment

In creating the Thematic Server we propose to overcome many of the barriers faced by the researcher wanting to make use of satellite derived data. The most basic class of barriers – challenges of locating the data, of navigating hierarchies of files, of deciphering formats and determining and associating missing use metadata – will be handled through OPeNDAP. Despite the distribution of the data sets across many institutions, utilizing distinct formats, directory structures and metadata conventions, the satellite fields will available to the user through a single Web portal. The fields will be made complete and uniform with respect to use metadata and will be "aggregated" to appear as dozens of data sets instead of thousands of files. They will be available to "open" as if they were locally resident on the user's computer in common desktop research tools such as Matlab, IDL, Ferret, and GrADS.

A subtler class of barriers to effective usage arises from issues of data content and interpretation, rather than issues of format and accessibility. These barriers cannot be broken down through simplified access to the data values, alone. Human-readable documentation, which summarizes and contrasts the characteristics of each satellite field – start and end dates, repeat intervals, spatial resolution, nominal accuracy, known limitations – can help somewhat with this class of problem and will be provided through this portal. To progress to the next stage of interpretation, evaluation, and comparison of the fields, however, requires tools that touch the data themselves.

For the (satellite non-specialist) researcher to gain the comfort level required to use satellite data with confidence (s)he must be in possession of tools with which (s)he can formulate idiosyncratic questions. The scientist must be able to focus in on the particular region of space and time, or the particular phenomenon which is her/his special interest.

(S)he must be able quickly and with ease to evaluate the effective coverage of the data, the gappiness of the data, spatial averages and average values over intervals of time, the variability in space and time. (S)he must be able to compare ensembles of different satellite estimates to separate the outliers from the more consistent measurements. Tools are needed that can quickly resolve differences in resolution and projection in order to render the satellite fields comparable, for example binning daily-average fields into 3-day average bins in order to compare to 3-day fields, and analogously to compare high resolution spatial fields with lower resolutions. (S)he should be able quickly to compare the fields to climatological data products and in-situ data. (S)he should also be able to quickly examine the relationship between variables as a "sanity check".

In many cases the type of analysis that would be desired may be impractical given the size of the data sets, the limitations of modern network and computer power, and the desire for a quick response. (A design requirement of the Web portal would be to impose limits on the level of compute or networking resources it would provide.) In such cases the tools should provide statistical modes that approximate the desired result through decimation or random sampling.

We propose to provide the user of the Thematic Web portal with exactly such a workbench. A key characteristic of the work bench is that the user interface must be simple to use and "friendly" to the non-specialist. If a user attempts an analysis that is inappropriate (e.g. comparing the time-variance of a 1-day field with a 3-day field) the workbench must warn the user that this is a questionable operation.

We enumerate below some of the obvious candidates for features and tools that should be included in this workbench. We note, however, that the process of assembling the right collection of tools will be evolutionary and will be refined and augmented based upon feedback from users throughout the duration of the project.

5.5.2 Workbench Tools

- The Workbench User Interface (UI) should permit the user to specify geophysical variable(s) of interest (.e.g., specific satellite data products). The UI should support selection based on space/time regions of interest, or optionally representative space or time scales, as well as map-based selection, and named regions (.e.g., Tropical Pacific) or by specifying latitude/longitude extents.

- Graphing tools are required that provide time series plots, XY maps (.e.g., images and colored contour plots), and scatter plots.

- Basic transformations and analyses such as time series averaged over spatial areas, lat/long regions averaged over time intervals, and similar capabilities for other statistics: variance, extrema, should be provided by the workbench. The workbench should match differing resolutions by binning (i.e., averaging) higher resolution fields in space or time to match lower resolution fields, as well as provide tools to regrid between differing spatial reference projections.

- The workbench should provide reference data sets and the ability to compare the satellite fields to these data, such as real-time in-situ observations (available via the U.S. GODAE server). Historical in-situ surface measurement archives and climatological surface fields such as COADS data should be provided by the workbench as well.

5.5.3 The Live Access Server: A Foundation on which to Build the Workbench

`LAS` is a configurable scientific data "product" server. `LAS` offers a friendly user interface to gridded and in-situ data sets through standard Web browsers, providing customizable graphics and formatted data subsets. `LAS` is a well-proved system, originally released in 1994 and now in version 6, with on the order of 50 sites installed internationally.

LAS does not create products, itself. Rather it redirects the users' requests for product creation to "back end" functionality, which may be provided by various applications and services. The Ferret application is the default back end with in LAS, but individual sites have added alternatives such as GIS via ESRI ArcIMS (at NOAA/NGDC), the NCAR Control Language (at NCAR), and the Climate Data Analysis Tools (at DOE/Lawrence Livermore Laboratory.) The ability to utilize a range of applications and services will permit the Web portal to exploit a broad range of analysis and visualization tools. For example, the use of IDL as a back end within LAS will leverage extensive libraries of specialized scripts and programs for the workbench. Note that in the LAS architecture the user interface is fully de-coupled from the product server. Thus there is broad latitude to build a specialized user interface for the proposed workbench within the standard design constraints of the LAS server.

6 SEEDS Involvement

Within the scope of this work we propose to contribute to several SEEDS working groups: Standards and Interfaces, Technology Infusion and Software. OPeNDAP is well suited to support these groups based on the partners' decade-plus experience in developing data access and transport technology in an open source environment. DODS and now OPeNDAP have had extensive experience in the development of middle-ware technology, which by its very design, must be adopted by other groups in order to succeed. To date many groups have used both our software and DAP protocol to build their own data systems or to build software which functions in the DODS/NVODS data system.[19] In addition to experience developing and promoting our own technology, we have extensive experience in integrating outside technology, both concrete implementations and standards. The complexity of this activity is easy to underestimate and doing so may cause groups to feel 'burned' by their initial experiences. We can help SEEDS to develop guidelines that reduce this risk. Another critical activity that must be tackled by an open source project is to integrate modifications to the project's own software that have been made by outside developers. The ability to tap into the resources of developers outside the formal bounds of the project is one of the key strengths of open source development. But this, too, is a tricky process which must be carefully managed. Many people outside of the small development team base at URI (now OPeNDAP) have contributed to the code base of our project and we can share our knowledge of this process through SEEDS.

We proposed to contribute 0.25 FTE to the Standards and Interfaces Working group and 0.25 FTE to be split between the Technology Infusion and Software Working groups.

7 Metrics

Metrics are measures of the state or usage of individual components of a system, or of a combination of components, including the system as a whole. Metrics in the context of the CAN, serve two basic complementary functions, similar to grades in school. The first is to give NASA a quantifiable view of progress of a project. The second is to give that project a means of self examination and a way to look for improvements.

Two of the standard metrics currently reported are the number of bytes available through a given system or available at a given site, and the number of monthly "hits" at that site. Such metrics in and of themselves will not provide good measures of progress on the development or use of the Thematic Data Portal. More appropriate metrics would be the number and variety (distribution in space, time, and sea-surface property variables) of datasets linked to the portal, and, by looking at the logs from the datasets themselves,

[19] An example of the former is Columbia's *Ingrid*, an example of the latter is COLA's *GDS*.

determining how many data requests were made and filled through the portal. These metrics would give NASA an idea of population and use of the portal, once it was operational, and they would provide the Portal Project with information about who uses the portal, how regularly they use it, and if the portal were being used more for discovery, routine access, analysis, or for educational purposes. Single user access to multiple datasets with compatible variables (datasets of sea surface temperature at various locations determined at different times, for example) might indicate the need for an aggregation server to allow access to those datasets with a single request, increasing the efficiency of the whole system. Therefore one of the goals of the proposal is to develop metrics and metric analysis software "on the back end" that will truly reflect the growth of the portal. A final major use of metrics analysis is to apply the information about the way users access which data, to feed back to the users (directly when we know who they are and indirectly through user groups and our web sites) more efficient ways for them to access their data, and let them know about other data sets with similar or related content.

Implementation of automatic metric analysis software will depend on implementation of the uniform metrics logging mechanism already described for the servers. Currently, the OPeNDAP metrics are derived from a variety of log file formats, require some hand manipulation and estimation in some cases, and are often incomplete for any given dataset. The implementation of uniform logging and access will remedy that situation, so that the metrics effort can concentrate on analysis of the metrics rather than production of the raw numbers.

Support of the Overall Metrics Effort: We plan three concurrent phases:

1. First Supplying metrics in response to the CAN Appendix B section 3.7.

   - Project Inputs – Capital Expenditures, Human Resources used in the various components, Allocation of Software and hardware resources, Data acquired.
   - Project Outputs – Data Sets Served (a means to an end), Portal Development milestones, Portal Connectivity to Datasets, Portal Usage determined by internal metrics analysis.
   - Project Outcome – Portal Capabilities vs Projections with time. Portal use vs original intent. Different modes of portal use – Education, Research, Analysis, Management. Success of infrastructure developed.
   - Project Impact – Lessons Learned. How do portal capabilities differ from original design/expectations? How does the portal use differ from the original concept? Is the portal successful as a route to discovery and access for sea surface properties? How can one design and implement portals along the same lines for other themes/disciplines, particularly within ESE? (With a view toward NASA adoption of development of further portals at the end of the CAN period.)

2. Second SEEDS Study Team participation:

   We will participate by providing a regular member (up to 0.25 FTE) of the Metrics Planning and Reporting Study Team. We expect to bring the experience of gathering metrics from Distributed Datasets with rates reaching more than half a million requests in a single month.

   `http://www.po.gso.uri.edu/tracking/metrics/metricsmatrix.html`

   We expect to share our experiences in evaluating various metrics for reporting requirements, user support, and internal improvement with the ESE community through the SEEDS Working Group. While we understand that the primary goal of the SEEDS metrics activity is to monitor and demonstrate the usefulness and successes of the CAN program directly, we expect to show how to use the usefulness of the metrics internally and for user enhancement, to vastly increase that demonstration to NASA and the world.

3. Third To monitor the development and use of the portal in all its aspects, and to develop metrics analysis that will demonstrate monitor the growth of the discovery, access, analysis, management, and educational uses to which the portal is put.

8 Schedule Milestones

Deliverables fall into two categories: regular and event driven.

Regular deliverables may also be divided into two categories: mandated reports and general support services.

Mandated reports include: Initial Plans and Reports, Monthly Financial Reports, Quarterly Technical Reports, Interim Review, Annual and Final Reviews and Annual and Final Report(s) – OPeNDAP .08 FTE

General Support services include: Code Maintenance such as bug fixes and system upgrades – OPeNDAP
user and AS support – Unidata

Following are the more specific milestones.

6 months after receipt of funding

- Metrics A: Incorporate Registration in OPeNDAP Servers – OPeNDAP
- AIS A: XML Conversion of Prototype – OPeNDAP

9 months after receipt of funding

- Metrics C: Initial "Smart" Analysis Software running – OPeNDAP
- GRID() function enhancement complete – OPeNDAP

12 months after receipt of funding

- URI Development Year 1: URI Generate SST data list of available OPeNDAP SST datasets. Design Top Level Sea Surface Property Portal Web Page Design generic Portal Subpage with SST as Use-Case. Develop use metadata for SST datasets to allow interoperability in the portal.
- **opendap!** Development Year 1.0 opendap AIS A: Design, Implement, Test Extensible Markup Language (XML) conversion of prototype Metrics A: Develop server registration as part of installation Metrics C: Begin Automation of Metrics Gathering
- LAS Year 1: – PMEL Procure and install server hardware, initial LAS. Configure SST fields, accomodating projections. Configure Basic browse (time series maps, Hofmuller plots) and subsets (netCDF, ASCII, IEEE). Basic comparison (difference and overlay). Add averaging in space and time (e.g. time series of an area average)
- Data access to NASA Satellite Data at JPL, Year 1: – JPL Provide access to the Near Real Time data through OPeNDAP servers. Server statistics will be provided on a monthly basis.
- Brown Development Year 1: – Brown University The general educational context for the data will be determined. The engineering the distribution mechanisms for classroom use will be determined. Initial selection of ocean data and exercises will be available. Design of IDL Tools to be used with the EDMI Prototypes of IDL Tools for opendap access Begin 3D and 4D ocean data visualization prototyping Begin computer graphic modules to display data in various projection formats for public theaters
- JASON Year 1: – JASON Develop the general educational context for this data as well as engineer the distribution mechanisms. Data and exercises will be available in Year One.
- NMS Development Year 1: – NMS Design of IDL Tools to be used with the EDMI. Prototypes of IDL Tools for opendap access

18 months after receipt of funding

- URI Development Year 1.5: URI Complete Development of SST Sub-page. Begin Design of sea surface wind Sub-page. Develop use metadata for Sea Surface wind Sub-page. Generate sea surface wind data list of available opendap SS-Wind datasets.
- opendap Development Year 1.5 opendap AIS B: Design Support for variables in AIS resources.

24 months after receipt of funding

- URI Development Year 2: URI Complete Development of sea surface wind Sub-page. Begin Design of sea surface color Sub-page. Develop use metadata for ocean color Sub-page. Generate ocean color data list of available opendap ocean color datasets.
- opendap Development Year 2 opendap Metrics B: Design server logging and access to server log. AIS B: Implement and Test Support for Variables in AIS resources. Metrics C: Design Metrics Analysis for user-dataset distribution.
- LAS Yr 2: – PMEL Configure 2nd variable (as determined by Principal Investigator (PI)). Integrate LAS Portal with GCMD searches. Augomatic decimation of large requests. Add graphical and analysis tools to characterize ensembles of geo-physical variables. Add property-property plots and correlations. Add HDF and DODS URL outputs.
- Data access to NASA Satellite Data at JPL, Year 2: – JPL Provide access to MODIS SST / SeaWinds / JASON-1 data
- Brown Development Year 2: – Brown University Updating the exercises for changes in the data. Determine the promotion of the availability of the educational resources Determine the distribution of educational materials for classroom use. Final versions of IDL Tools Design of Satellite View Interface and 3D world model Completion of Satellite View Interface and 3D world model Design of Educational Components with JASON Project Evaluate prototype scientific visualizations for the immersive theater. Construct computer models of various theaters with different projection configurations. Test display models in simulation.
- JASON Year 2: – JASON Update the exercises to take advantage of data made available since the previous yearly milestone, and promote the availability of this resource and its distribution. Begin a plan to evaluate the usefulness of the distribution mechanism.
- NMS Development Year 2: – NMS Final versions of IDL Tools Design of Satellite View Interface and 3D world model Completion of Satellite View Interface and 3D world model Design of Educational Components with JASON Project
- Open source Demonstration Year 2: Planning. Develop gateway modification designs for WCS/WFS conversions to existing WMS gateway. Develop plan to specify and document reusable assets.

30 months after receipt of funding

- URI Development Year 2.5: URI Complete Development of ocean color Sub-page.
- opendap Development Year 2.5 opendap Metrics B: Implement and Test server metrics logging and access to logs.

36 months after receipt of funding

- URI Development Year 3: URI Begin Design of sea surface Height Sub-page. Develop use metadata for sea surface Height Sub-page. Generate sea surface Height data list of available opendap sea surface Height datasets.
- opendap Development Year 3 opendap AIS B and C: Develop AIS capability within the Portal. Metrics C: Implement Metrics analysis to determine user-dataset access with a view toward enhancing user support.
- LAS Year 3: – PMEL Configure 3rd variable (as determined by PI). Redesign user interface to include context-sensitive documentation and informative messages for inappropriate analyses. Add self-testing/robustness assurance, automatic limiting (choking) of resources, and summary/analysis of logs to produce metrics.

- Data access to NASA Satellite Data at JPL, Year 3: – JPL Provide access to Quick Scatterometer (QuikSCAT) / TOPEX/Poseidon / AVHRR Pathfinder data
- Brown Development Year 3: – Brown University Updating the exercises for changes in the data. Manage the promotion of the availability of this resource and its distribution for classroom use. Begin evaluation of the educational materials developed. Integration of all components for Satellite View Interface Design of Sea Surface Interface and 3D world model Completion of Sea Surface Interface and 3D world model Begin merging immersive display techniques with real feeds from JASON Display the first set of scientific visualizations in an immersive theater and/or large format theater in (1-3) national public aquariums.
- JASON Year 3: – JASON, URI Continue to improve the exercises based on new data availability. Test the plan to evaluate the usefulness of the distribution mechanism, based on metrics obtained from the opendap servers.
- NMS Development Year 3: – NMS Integration of all components for Satellite View Interface Design of Sea Surface Interface and 3D world model Completion of Sea Surface Interface and 3D world model Design of Educational Components with JASON Project
- Open source Demonstration Year 3: Develop complete specification documents, document all source for reusable assets and provide through open source distribution network. Extend existing WMS gateway to support WCS requests, provide complete Javadoc-like source documentation. Develop installation/technical documentation for WCS. Develop AIS repository for raster-based ocean data provided by Thematic Data Portal.

42 months after receipt of funding

- URI Development Year 3.5: URI Complete Development of sea surface Height Sub-page
- OPeNDAP Development Year 3.5 OPeNDAP Test and refine AIS capability based on intial (beta) usage. Begin design of aggregation server enhancements for data sets served into the Portal. Design Grid() function enhancements.

48 months after receipt of funding

- URI Development Year 4: URI Begin Design of Rainfall Sub-page Develop use metadata for Rainfall Sub-page. Generate Rainfall data list of available opendap Rainfall datasets.
- opendap Development Year 4 opendap Begin Installation of AIS on servers accessed by the Portal. Implement and test Grid() function enhancments Metrics C: Determine user access demographics through the various Sub-portals.
- LAS Year 4: – PMEL Configure 4th variable (as determined by PI). Set up mirroring at high power server at URI. Perform usability testing and iterative improvement to user interface. Incorporate in-situ data and climatological reference fields. Add Geographic Information Systems (GIS)-compatible output (geoTIFF) and ability to compute variance and extrema in space and time.
- Data access to NASA Satellite Data at JPL, Year 4: – JPL Provide access to GRACE / NSCAT / ASTER data
- Brown Development Year 4: – Brown University Updating the exercises for changes in the data. Continue to promote educational resource and distribution for classroom use. An evaluation of the educational material continues. Integration of all components for Sea Surface Interface Design of Subsurface Ocean Interface and 3D world model Completion of Subsurface Ocean Interface and 3D world model Present three "canned" programs to public audience. Evaluate immersive and/or large format display audiences.
- JASON Year 4 – JASON, URI Continue to improve exercises based on new data available. (See JPL milestones above, e.g.) Implement plan for evaluating the operation of the educational component. Begin making evaluations based, in part, on metrics collected from the opendap servers, and analysed with respect to logged requests from known education users.

- NMS Development Year 4: – NMS Integration of all components for Sea Surface Interface Design of Subsurface Ocean Interface and 3D world model Completion of Subsurface Ocean Interface and 3D world model
- Open source Demonstration Year 4: Develop WFS gateway using WCS/WFS templates, provide complete Javadoc-equivalent documentation for source. Develop installation/technical documentation for WFS. Develop AIS repository for in-situ ocean data provided by Thematic Data Portal. Meet with public health data organizations to identify target data products for making accessible via gateway prototypes.

48 months after receipt of funding

- URI Development Year 4.5: URI Complete Development of Rainfall Sub-page
- opendap Development Year 4.5 opendap Implement aggregation server enhancements for cross-dimensional datasets.

60 months after receipt of funding

- URI Development Year 5: URI Complete development of full Portal Capabilities. Determine usage of the different Sub-portals from analysis of server logs.
- LAS Year 5: – PMEL Configure 5th variable (as determined by PI). Add automated statistical sampling of very large fields. Automated reprojection to match differing projections of fields. Arbitrary regridding to match differing resolutions.
- Data access to NASA Satellite Data at JPL, Year 5: – JPL Transition the web pages responsibilities at the end of the REASoN project.
- Brown Development Year 5: – Brown University Updating the exercises for changes in the data. Continue to promote educational resource and distribution for classroom use. An evaluation of the educational material continues. Integration of all components for Subsurface Ocean Interface Present interactive "live" ocean data immersion theater using interactive tools.Present combined JASON/Ocean Data tele-presence events
- JASON Year 5: – JASON, URI Complete integration of available datasets into the educational resources. Complete the distribution within the project, and prepare a plan for further distribution and enhancement of the resources beyond the end of this CAN. Conclude the evaluation of the usefulness and growth of the educational resources developed by the CAN.
- Open source Demonstration Year 5: Install OPeNDAP servers, AIS, and Aggregation Servers at public health data organization. Develop AIS repository for public health data targeted for demonstration. Develop end-to-end demonstration providing OGC-compliant access to public health data. Extend demonstration to combine NASA earth science data product with public health data served via gateway prototypes.
- NMS Development Year 5: – NMS Integration of all components for Subsurface Ocean Interface

9 Open-Source Prototype Demonstration

To demonstrate reusable asset deployment, gateway applications will be prototyped which will adhere to the OGC web coverage server (WCS) and web feature server (WFS) implementation specifications. Coupled with the existing web map server (WMS) gateway these will provide the complete set of Open GIS Consortium (OGC) web servers for OPeNDAP-accessible data to GIS user communities.

The prototype effort will demonstrate the reusability of several OPeNDAP components; the core DAP classes, the Ancillary Information Service (AIS), and Aggregation Server (AS). Gateway prototypes will be deployed as part of the Thematic Data Portal. We've identified, and contacted researchers and information technologists at the Centers for Disease Control (CDC), the National Association of Health Data Organizations (NAHDO), and the Harvard

School of Public Health. Groups within each of these organizations have expressed an interest in the software technology as a means to make their data readily accessible by GIS users. Each group has described a current requirement to provide distributed access to their data, and the desire to combine NASA earth science data products with their data, providing the combined products to GIS users.

Many problems associated with providing data to GIS applications are common across the ocean science and the public health communities. Both communities utilize diverse data storage formats, often with minimal and loosely coupled or incomplete metadata. Their respective data archives often contain large volumes of data, and maintain data collections comprising numerous individual granules, where each granule may store from one to many individual measurements. This presents a number of problems to GIS applications. GIS applications require that the input data correspond to a limited set of well-known formats and contain a predefined, consistent set of metadata, minimally representing the explicit geo-referencing information required to operate on the data in the GIS. Additionally, data in GIS is abstracted to coverages and features, which may represent one or several individual measurements from a science data granule.

To achieve interoperability the OGC has defined a set of abstract specifications, these include the Grid Coverage, Simple Features, and Coordinate Transformation Services specifications, among others. Utilizing the abstract specifications, the OGC develops and publishes open implementation specifications to facilitate development of standards-based software products for geo-spatial data. The OGC implementation specifications include the WMS, WFS and WCS specifications. The OGC specifications, both abstract and implementation, facilitate the development of GIS client applications by requiring that the products returned by OGC-compliant servers strictly conform to the underlying specifications for the requested data product, i.e., coverage or feature. For data providers, the OGC specifications provide a description of what their data must conform to, achieving conformance becomes the responsibility of the individual server implementation, and ultimately the data provider. For provider communities where diverse data storage formats with incomplete metadata are common, generating OGC-compliant products can become difficult and burdensome. Data providers are required to provide specialized OGC server implementations for their specific data storage formats, and potentially for particular data collections. Optionally, providers may reprocess their data into storage formats that maintain the required geo-referencing information in a form consistent with existing OGC server implementations to make their data accessible to GIS applications. Many organizations will find either option prohibitive, which ultimately acts to limit data accessibility to GIS even though data providers acknowledge the importance of this application.

The prototype gateways will demonstrate the reuse of OPeNDAP components to provide a scaleable solution for generating OGC-compliant products from data originating in diverse storage formats, with loosely coupled, incomplete or inconsistent metadata, and from large, multi-granule data collections.

The DAP classes and servers form the foundation for the stack. Together they operate to provide a flexible mechanism to provide access to data stored in diverse, heterogeneous storage formats. At its core, the DAP provides a data model based on standard computer-science data types, similar to those employed by all storage formats. The DAP classes are the basis of a suite of existing client/server applications developed as part of the DODS ESIP. DAPenabled servers exist for the following storage formats; netCDF, HDF4, HDF-EOS, HDF5, CDF, FITS, GRIB, BUFR, DBMS/JDBC, JGOFS, as well as simple, free-format binary and ASCII data. DAPenabled servers read data from the underlying storage format, converting the requested data into an intermediate representation referenced to the data model. Each DAP data type supports subsetting operations consistent with its type, optimizing the amount of information transmitted over the network. Additionally, the DAP provides the necessary mechanism to transport the data over the network, transparently, to client applications.

The DAP also provides a mechanism to represent and transport ancillary information, referred to as attributes, corresponding to individual elements within the data. DAPenabled servers will intern all of the ancillary information available from the underlying storage format. However, it is not uncommon for data storage formats to contain minimal, or an incomplete set of ancillary information, or that the information is inconsistent with the client application's requirements. The AIS component has been developed to insure that the metadata corresponding to specific data elements is complete and consistent for domain-specific applications. The gateway prototypes will use AIS to ensure that the relevant geo-referencing, and any required use metadata are tightly coupled to the data, and are complete and consistent for GIS interoperability. For example, in the GIS domain, spatial reference systems for coverages and features contain both input data elements, (.i.e., values in rasters, at points, and lines) and metadata (.i.e., datums, projections, and geo-spatial bounding extents) which are employed in concert to instantiate a spatial reference system for particular geo-referenced data. Using this information, geo-spatial applications can operate on the data elements to perform accurate geo-referencing operations to meet the varying needs of geo-spatial data applications. Without complete and consistent spatial reference information GIS application cannot operate on the data. To insure that the input data contains this information, the AIS is used to augment the geo-referencing information coupled to the input data, insuring that a complete and consistent set of geo-references is available with the input data. The AIS provides a flexible, client-driven mechanism to couple ancillary information to data, in a form consistent with the client application's requirements.

An integral problem facing GIS applications is access to large, multi-granule data collections. In GIS, the representation and interaction with multi-granule data collections can be problematic. To address this issue, the OGC implementation specifications employ the semantic terminology of sample dimensions to provide a mechanism to describe the independent variables corresponding to specific data elements. To facilitate access to multi-granule data collections Aggregation Servers will be used to provide a monolithic, n-dimensional representation for the data collection. Using AS, each sample dimension as defined in the OGC implementation specifications exist as one of the dimensions in the aggregated, n-dimensional virtual dataset.

Layering one OPeNDAP component upon another, the gateway prototype operates to resolve the common but separate problems encountered when generating OGC-compliant data products from ocean science and public health data repositories. The open-source prototype demonstration will accomplish several objectives. First, it will demonstrate the ability to build semantically aware applications through mechanisms that allow information communities, in cooperation with data providers, to couple application specific ancillary information and semantic concepts to a flexible, network accessible data model. Second, it will provide a complete set of OGC Web Servers for use in the ocean science community, in particular providing GIS access to in-situ measurements, represented as map features, which is both important and desirable to the ocean science community. Working with identified public health organizations, the open-source demonstration effort will provide insight gained through NASA's ESIP technology development in distributed data access to the public health community, fostering open-source, standards-based access to public health data in general, and for GIS in particular. Finally, the open-source demonstration project will ensure that the software technology originating in the DODS ESIP development effort, and evolving through OPeNDAP, will meet the guidelines and reuse principles envisioned in SEEDS.

To accomplish the requirements set out in the Open-source Prototype Demonstration description the project will be delineated into 3 functional areas; mining the technology assets and readying them for open-source dissemination; building prototype OGC gateways based on the mined assets; and working with public health data organizations to install the OGC gateways, and assist in the installation and configuration of OPeNDAP servers, AIS and AS for public health data organizations.

References

[1] **ESE's Research Strategy for the Next Decade, NASA Earth Science Enterprise Research Strategy for 2000-2010**, NASA, available at: http://www.earth.nasa.gov/visions/researchstrat/Research_Strategy.htm 64pp, 2000.

[2] Cayula, J.-F., P. Cornillon. Cloud Detection from a sequence of SST images. *Remote Sens. Environ.*, 55, 80-88, 1996.

10 Acronyms/URLs

AIS Ancillary Information Service.

AIST Advanced Information Systems Techonology.

API Application Program Interface.

ASCII American Standard Code for Information Interchange.

AS Aggregation Server.

ASTER Advanced Spaceborne Thermal Emission and Reflection Radiometer – *http://asterweb.jpl.nasa.gov*

AVHRR Advanced Very High Resolution Radiometer.

BUFR Binary Universal Form for the Representation of meteorological data – *http://www.wmo.ch/web/www/WDM/Guides/Guide-on-DataMgt-1.htm*

CAN Cooperative Agreement Notice.

CDC Centers for Disease Control – *http://www.cdc.gov*

CDF Common Data Format.

CEDAR Coupling, Energetics and Dynamics of Atmospheric Regions.

CGI Common Gateway Interface.

COARDS Cooperative Ocean-Atmosphere Research Data Standard.

CODMAC Committee on Data Management, Archiving, and Computing.

COLA Center for Ocean-Land-Atmosphere studies

CSP Chermayeff, Sollogub and Poole, Inc. – *http://csp-architects.com/contact.htm*

DAAC Distribute Active Archive Center.

DAP Data Access Protocol.

DBMS Data Base Management System.

DODS Distributed Oceanographic Data System – *http://www.unidata.ucar.edu/packages/dods*

DOE Department of Energy – *http://www.energy.gov*

EDMI Earth Data Multi-media Instrument – *http://www.newmediastudio.org/Homepage/TNMSHomeFramset.htm*

EOS Earth Observing System.

ESG II Earth System Grid – II – *http://www.earthsystemgrid.org*

ESIP Earth Science Information Partner – *http://www.esipfed.org*

ESRI Environmental Systems Research Institute– *http://www.esri.com*

Ferret – *http://ferret.pmel.noaa.gov/Ferret*

FGDC Federal Geographic Data Committee – *http://www.fgdc.gov*

FITS Flexible Image (or Interchange) Transport System.

FreeForm – *http://www.ngdc.noaa.gov/seg/freeform/freeform.shtml*

FTP File Transport Protocol.

GCMD Global Change Master Directory – *http://gcmd.nasa.gov*

GHRSST-PP `GODAE` High Resolution Sea Surface Temperature Pilot Project – *http://www.ghrsst-pp.org*

GIS Geographic Information Systems – *http://www.gis.com*

GODAE Global Ocean Data Assimilation Experiment – *http://www.bom.gov.au/bmrc/ocean/GODAE*

GOES Geostationary Operational Environmental Satellites – *http://www.oso.noaa.gov/goes*

GRACE Gravity Recovery and Climate Experiment – *http://www.csr.utexas.edu/grace/spacecraft/config.html*

GRIB GRid In Binary – *http://www.wmo.ch/web/www/WDM/Guides/Guide-on-DataMgt-1.htm*

GrADS Grid Analysis and Display System – *http://grads.iges.org/grads/index.html*

GridFTP `FTP` with the GRID enhancements

GSFC Goddard Space Flight Center – *http://www.gsfc.nasa.gov*

GUI Graphical User Interface.

HDF Hierarchical Data Format.

HDF-EOS Hierarchical Data Format - `EOS` – *http://hdfeos.gsfc.nasa.gov*

HRPT High Resolution Picture Transmission.

HTTP Hyper Text Transfer Protocol.

IBM Internation Business Machines – *http://www.ibm.com/us*

IDE Integrated Development Environment.

IDL Interactive Display Language – *http://www.rsinc.com/idl/index.asp*

IDV Integrated Data Viewer – *http://my.unidata.ucar.edu/content/software/metapps/index.html*

IEEE Institute (of) Electrical (and) Electronic Engineers – *http://www.ieee.org/portal/index.jsp*

JASON JASON Foundation for Education – *http://www.jason.org*

JDBC Java Database Connectivity

JGOFS Joint Global Ocean Flux Experiment – *http://puddle.mit.edu/datasys/jgsys.html*

JPL Jet Propulsion Laboratory – *http://www.jpl.nasa.gov*

LAS Live Access Server – *http://www.ferret.noaa.gov/nopp/main.pl?*

Matlab – *http://mathworks.com*

mks meters, kilograms, seconds.

MODIS MODerate-resolution Imaging Spectroradiometer – *http://modis.gsfc.nasa.gov*

NAHDO National Association of Health Data Organizations – *http://www.nahdo.org*

NASA National Aeronautics and Space Administration – *http://www.nasa.gov*

NCAR National Center for Atmospheric Research – *http://www.ncar.ucar.edu/ncar/index.html*

ncBrowse – *http://www.epic.noaa.gov/java/ncBrowse*

netCDF NETwork Common Data Format.

NGDC National Geophysical Data Center – *http://www.ngdc.noaa.gov*

NMS New Media Studio – *http://newmediastudio.org/Homepage/TNMSHomeFramset.htm*

NOAA National Oceanic and Atmospheric Administration – *http://www.noaa.gov*

NOPP National Oceanographic Parternership Program – *http://www.nopp.org*

NSCAT NASA SCATterometer – *http://winds.jpl.nasa.gov*

OGC Open GIS Consortium – *http://www.opengis.org*

OPeNDAP Open source Project for a Network Data Access Protocol – *http://opendap.org*

PI Principal Investigator.

PM Project Member.

PMEL Pacific Marine Environmental Laboratory – *http://pmel.noaa.gov*

PO-DAAC Physical Oceanography – Distributed Active Archive Center – *http://podaac.jpl.nasa.gov*

QuikSCAT Quick Scatterometer – *http://winds.jpl.nasa.gov/missions/quikscat/quikindex.html*

REASoN Research, Education and Applications Solutions Network.

RSS Remote Sensing Sytems –

SEEDS Strategic Evolution of ESE Data Systems – *http://lennier.gsfc.nasa.gov/seeds*

SGI Silican Graphics Incorporated – *http://www.sgi.com*

SQL Structured Query Language.

SST sea surface temperature.

TCASCV Technology Center for Advanced Scientific Computing and Visualization – *http://www.cascv.brown.edu/aboutus.html*

TCP/IP Transmission Control Protocol/Internet Protocol.

THREDDS Thematic Realtime Environmental Data Distributed Services – *http://www.unidata.ucar.edu/projects/THREDDS*

TRL Technology Readiness Level.

TRMM Tropical Rainfall Measuring Mission – *http://trmm.gsfc.nasa.gov/*

UCAR University Corporation for Atmospheric Research – *http://www.ucar.edu*

UCSD University of California, San Diego

Unidata – *http://unidata.ucar.edu*

URI University of Rhode Island – *http://www.uri.edu*

URL Uniform Resource Locator.

VR Virtual Reality.

WCS web coverage server – *http://www.ogcnetwork.org*

WFS web feature server – *http://www.ogcnetwork.org*

WMS web map server – *http://www.ogcnetwork.org*

WIMP Windows, Icons, Menus, and Pointers

XML Extensible Markup Language – *http://www.w3.org/XML*

11 Budget

Appendix A. Letters of Support

Due to time constraints we were not able to obtain letters of committement from all advisory group members. We have however obtained verbal confirmation of an intention to participate and can, if needed, obtain written confirmation.

Appendix B. Brown University Subcontract

Appendix C. Goddard–DAAC/Goddard Space Flight Center (GSFC) Subcontract

Appendix D. JASON Foundation for Education Subcontract

Appendix E. New Media Studio Subcontract

Appendix F. Open source Project for a Network Data Access Protocol (`OPeNDAP`) Subcontract

Appendix G. Pacific Marine Environmental Laboratory (PMEL) Subcontract

Appendix H. Physical Oceanography – Distributed Active Archive Center (PO-DAAC)/Jet Propulsion Laboratory Subcontract

Appendix I. Unidata/University Corporation for Atmospheric Research (UCAR) Subcontract