A Gateway to Support Interoperability of OPeNDAP and OGC Protocols

Kenneth McDonald[1], Yonsook Enloe[2], Liping Di[3], Daniel Holloway[4]
[1]NASA Goddard Space Flight Center, Code 423, Greenbelt, MD 20771
[2]SGT Inc., 7600 Greenbelt Road, Suite 400, Greenbelt, MD 20770
[3]Center for Geospatial Information Science and Systems (CSISS), George Mason University, 6301 Ivy Lane, Suite 620, Greenbelt, MD 20770
[4]OPeNDAP Inc., 165 Dean Knauss Drive, Narragansett, RI 02882-1124

Abstract- **Data access and analysis tools that are developed within specific disciplines and the protocols that they are built upon provide valuable services to their respective users but can actually be a barrier to the integration of data from a broad set of data sources. An example of this is data supported by OPeNDAP that is widely used in the ocean and atmospheric sciences, and data provided through the interface specifications of the Open Geospatial Consortium (OGC) that typically serves the land science community. This paper describes a project that is developing a gateway to bridge these two data system infrastructures, in response to a specific need expressed by CEOP, an international science program.**

I. Introduction

The Coordinated Enhanced Observing Period (CEOP) is an element of the World Climate Research Program (WCCRP) and was initiated by the Global Energy and Water Cycle Experiment (GEWEX) to focus on the measurement, understanding and modeling of water and energy cycles within the climate system [1]. It has identified a network of thirty-six reference sites and several enhanced observing periods (EOPs) for which it collects and assembles field observations and satellite data and also model and data assimilation output products. To meet its objectives, the CEOP science community requires data integration services that allow it to access and inter-compare these diverse data types from multiple sources.

The CEOP Program initiated a discussion with the Committee on Earth Observation Satellites (CEOS) Working Group on Information System and Services (WGISS) to determine if WGISS could assist in the development of advanced tools to access the various data collections with the data service needed to support data integration. Two of the WGISS agencies, the Japanese Aerospace Exploration Agency (JAXA) and NASA volunteered to work with CEOP in addressing these services. The NASA team chose to focus on the enhancement of access to its satellite data resources.

II. Project Definition

From early interactions with a number of CEOP scientists, NASA learned that they were currently using client/server applications based on the Open-source Project for a Network Data Access Protocol (OPeNDAP) [2, 3] to access and analyze the field data and the model/assimilation products. The NASA team had been developing a set of data system components called the NASA Web-based GIS Services (NWGISS) [4] that could apply a number of data services to satellite data products, including georectification, subsetting, resampling and reprojection, and makes the resulting products *accessible via standard interfaces specified by OGC [5]*. These data services are the same type of services needed to support data integration. The scientists saw the value of the NWGISS capabilities, but they needed the products to be served via OPeNDAP to be truly useful for data inter-comparisons of the diverse data types.

To address this issue, the NASA team proposed the development of a gateway that would allow an OPeNDAP client to access the data and services provided by an OGC-compliant web coverage server (WCS). With assistance from OPeNDAP Inc., a proof-of-concept demonstration was developed to allow a user of an OPeNDAP client (e.g., GrADS) to access a satellite data product from NWGISS and to overlay a model output acquired from an OPeNDAP server [6]. While the demonstration was far short of a fully functional, generic capability, it did begin to address some of the challenges of translating between the two environments.

Building on the experience of the demonstration, a team of NASA, GMU, SGT Inc. and OPeNDAP Inc. wrote a successful proposal to NASA's Advancing Collaborative Connections for Earth-Sun System Science (ACCESS) Program. The proposed effort will integrate catalog services with the gateway and provide additional flexibility to allow a user *to fully specify the desired services to be applied to the* satellite data. This effort was also merged with a second proposal to enable data held by OPeNDAP, or what was referred to as geoscience servers, to be accessed by OGC clients, thus providing interoperability in the opposite direction.

III. The Gateway Design

While this ACCESS project is conceptually addressing two-way interoperability between OPeNDAP and OGC clients and servers, doing this requires the development of two gateways. The first one is called the CEOP Satellite Data Server (SDS) and the second gateway the OGC-Geoscience Gateway.

A. CEOP Satellite Data Server (SDS)

The CEOP SDS leverages the geospatial processing capabilities of the WCS with the transparent access of the OPeNDAP data access protocol or DAP, to expose a single, standardized data representation of a satellite data product consistent with other CEOP data sources. The project is using GrADS, a widely-used analysis and visualization application to demonstrate access to satellite data products, in particular data from NASA's EOS missions, for comparison with field measurements and model output. GrADS is an excellent example of a pre-existing application that has been OPeNDAP-enabled, meaning it has the additional capability to make requests to, and retrieve data from OPeNDAP servers located anywhere on the Internet.

There are two fundamental elements required to integrate these two data access technologies such that existing applications, like GrADS, can readily access the data. The first is a standardized representation or data model for satellite data that is consistent with other CEOP data sources and into which the server will transform the native satellite data product. The second is a set of software components that will transform elements of a DAP request into a WCS request, issue that request to a WCS instance and transform the WCS response into a DAP data object consistent with the standardized representation, and return that DAP object to the requesting client. Figure 1. shows the design of the server.

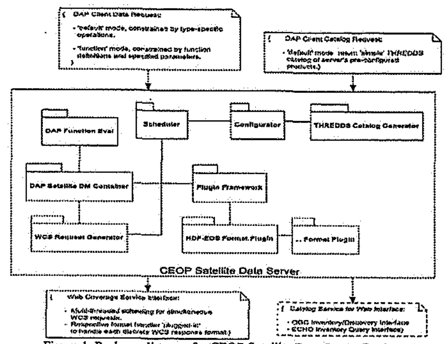

Figure 1. Package diagram for CEOP Satellite Data Server Design

A.1 Data Model

To facilitate interoperability with CEOP model data the server utilizes a CF-based (Climate and Forecast) metadata multi-dimensional grid structure for representing the requested satellite data. The CF standards use the dimension ordering of (t,z,y,x) to regularize the shape of multi-dimensional data sources. Additionally, CF implements and utilizes extensible naming conventions with regard to geophysical quantities, their units of measure and associated attributes to facilitate data access and use within analysis applications. Initially, the server implementation will focus on the outermost two dimensions of the CF grid structure (y,x), and not attempt to interpolate within the innermost dimensions (t,z). Aggregating multiple granules temporally, from a given collection of satellite data products into a multi-dimensioned grid representation consistent with that used in the CEOP numerical modeling community would be an important additional capability to facilitate model and satellite data inter-comparison but the focus of this effort is to provide consistency at the outermost (y,x) dimensions. To facilitate existing client applications the server will express, whenever possible, the appropriate (t,z) independent dimensions in the multi-dimensional grid structure.

A.2 Software Components

The server comprises components that evaluate the OPeNDAP client request, generate valid OpenGIS Web Coverage Service (WCS) requests, and intern the response from those services into the structural Grid representation prescribed by the common data model. Each component can be envisioned as a package consisting of classes operating together to provide a functional element for the server. The following paragraphs provide a brief introduction of each component.

- DAP Satellite Data Model Container: This component instantiates a DAP data object conforming to the CF-based representation of the satellite product, relative to the default grid parameters specified for the product in the external configuration document. This component is the primary DAP element when the server operates in the non-function based, or default DAP mode.
- DAP Function Evaluator: This component evaluates any server-side functions that are contained in the client request and creates a DAP data object consistent with the results of that functional request. As described earlier, the functional interface will consist of a predefined, extensible, set of function names with well-defined argument parameters. The set will initially consist of functions to specify geo-spatial selection with latitude-longitude bounding-box or center latitude-longitude coordinates, grid cell resolution and interpolation methods. As a result of function evaluation a DAP Satellite Data Model Container will be instantiated with the parameterization necessary to represent the result of the functional request. That container will then operate to create the WCS request and process the WCS response into an appropriate DAP response to the request.
- WCS Request Generator: This component combines the WCS request elements translated by the DAP Satellite Data Model Container and/or DAP Function Evaluator, with additional WCS request elements from the Configurator component to generate a valid WCS request for processing the satellite data product.
- WCS Response Format Handler Plug-in: The WCS Implementation Specification allows a WCS to return, minimally, any one of five well-known binary file formats, and optionally any other file format it chooses for a 'Layer' it advertises. To support multiple response formats from a WCS, the server utilizes a plug-in framework to instantiate the proper format handler for the

resulting response format from the WCS. The plug-in can be viewed as a specialization of a base class that operates to read WCS well-known binary formats, creating a standardized data structure that can be used by the OPeNDAP interface module to form the response back to the requesting client. This component defines the interfaces for those base classes.

- Scheduler: This component is the primary engine of the server. It operates to coordinate the activities of the packages comprising the server. Activity management is provided to support multiple threads for handling simultaneous WCS requests, and the potential asynchronous nature of their completion. The component determines the required Format Handler from the Configurator component and instantiates the appropriate plug-in module. Additionally, this component will support caching to provide enhanced throughput for particular DAP usage patterns.
- Configurator: Initially this component will utilize an external XML configuration document to identify the satellite data products available to the server from a WCS. The configuration parameters include WCS layer information for each satellite product available for transformation by this server. Included in the configuration is the default grid resolution and interpolation parameters to use for transforming the satellite product from its native representation to the server's DAP representation. The configuration document also stores which WCS response format to request, and the format handler to plug-in for processing the WCS response for this layer. Further development is envisioned to support integration with OGC CS/W services for automating the configuration of new products into the server.
- THREDDS Catalog Response Generator: The OPeNDAP-enabled servers and clients generally support the THREDDS catalog interface. To expose the available satellite data products from the server to the OPeNDAP community of users the server will provide a 'simple' THREDDS catalog response. This component translates the relevant elements of the external XML configuration document into a THREDDS catalog. THREDDS catalogs are themselves XML documents that can describe both directory and inventory level information for a data source. The resulting THREDDS catalogs will include the *specific information OPeNDAP URLs to the satellite* products available from the server, as well as the geospatial and temporal extent of the individual satellite products, and information available from the WCS describing the satellite data. THREDDS catalogs are becoming generally available from data providers using OPeNDAP servers and the suite of OPeNDAP-enabled client applications are migrating toward this catalog technology.

The CEOP SDS will undergo extensive testing to prepare for the operational use. The software will be made available to other interested CEOS agencies (e.g. JAXA, ESA) willing to host the software with Web Coverage Server access or OPeNDAP server access. After the initial deployment of this gateway, the CEOP science community will test the SDS and provide feedback on ideas for iterative enhancements.

B. OGC-Geoscience Gateway

The *OGC-Geoscience Gateway* uses a set of interface specifications developed by the Open Geospatial Consortium (OGC) to access geospatial data served by geoscience protocols such as OPeNDAP and THREDDS. The major work of the geoscience gateway is to rapidly develop and deploy an OGC-geoscience gateway for facilitating the interoperability from the OGC catalog and data access protocols (CS/W [7], WCS [8], and WFS [9]) to the catalog and data access protocols used in the geoscience communities (THREDDS [10], OPeNDAP, and netCDF/http). The gateway will enable OGC clients to search and access data served through those geoscience protocols. Figure 2 show the overall architecture of the gateway.

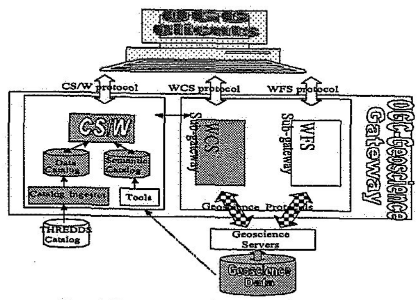

Figure 2. The Architecture of OGC-Geoscience Gateway

The shaded components in the figure are those that the prototypes already exist. This project only needs to update, enhance, assemble, and test them in the OGC-Geoscience Gateway instead of developing them from scratch. As shown in Figure 2, many of the components already exist. This allows us reusing those components to rapidly develop and deploy the gateway. The following paragraphs describe the major new developments for the three sub-gateways: CS/W, WCS, and WFS.

- CS/W to THREDDS sub-gateway: This sub-gateway will provide CS/W interface to OGC clients. The NWGISS CS/W server will be reused in this sub-gateway. The major development in the sub-gateway is to develop the catalog ingestor that reads the *THREDDS XML* data catalog document and ingests it into the CS/W searchable mySQL database. Tools also need to be developed for facilitating the creation of semantic catalog. The CS/W will be enhanced so that it can combine the information in the data catalog and the semantic catalog to generate valid CS/W search results that can be used by OGC clients to formulate valid WCS or WFS data retrieval requests.
- WCS to geoscience-servers sub-gateway: The THREDDS prototype WCS server developed by Unidata will be used as the starting point for the sub-gateway. The major work for the sub-gateway is to enhance the prototype so that it becomes a fully compliant, fully functional OGC WCS server, from the viewpoint of OGC WCS clients. The enhancement includes providing services of map-coordinate based subsetting, resampling, reprojection, reformatting, etc to coverage data before the data being sent to OGC clients. Those services will be exposed to the WCS clients through the capabilities description.
- WFS to geoscience-servers sub-gateway: Currently, there are no *components available for this sub-gateway.* Therefore, the sub-gateway will need to be developed from scratch. The similar architecture as the WCS sub-gateway will be used. Beside the protocol translation, the major work for this sub-gateway will be the development of constraint converter and implementation of feature data selection function.

To design the gateway for facilitating the protocol interoperability, the general pattern for data access in OGC client-server model was considered. A typical OGC session for accessing geospatial data starts with a client searching a data catalog at a CS/W server (data discovery step). Once the requested dataset is found, the client can obtain the dataset by issuing a getCoverage request to the WCS server or a getFeature request to the WFS server which hosts the dataset (data retrieval step). The design and implementation of the Geoscience gateway will require an analysis of the metadata used by the OGC CS/W and the THREDDS protocols to determine what additional metadata is needed for CS/W access to THREDDS catalogs and also what metadata is needed in the catalog search results to construct a valid WCS or WFS access.

Under the current ACCESS award, the project will develop the CS/W-THREDDS subgateway (the catalog subgateway) in the first and the second years. The WCS-OPeNDAP/netCDF subgateway (the WCS subgateway) will be developed in the optional third year, if it is funded.

IV. CONCLUSION

The successful implementation and deployment of the gateway will immediately provide easy access to a large amount of NASA EOSDIS data from OPeNDAP clients and to large amounts of Geoscience data from OGC clients. Currently, a small subset of NASA EOSDIS data is available to CEOP scientists only via the pre-processed datasets for a limited CEOP reference sites. The gateway will also greatly facilitate the utilization of satellite data in the CEOP modeling efforts. The gateway will provide many of the data services needed to support data integration and enable the quick and *effective inter-comparison between model outputs and satellite* observations and in situ reference data. The gateway will significantly improve the scientific productivity of CEOP scientists. The data services needed to support data integration that facilitate inter-comparison of satellite data with model output data and in situ reference data are also needed by other scientists outside the CEOP community. The gateway could be offered to other science communities needing the capabilities to inter-compare satellite data with model output and in situ data. The gateway will also significantly reduce the cost for CEOP data operation and provide much better data services to CEOP scientists since all the pre-processing for making CEOP datasets will be automated and the CEOP scientists can get personalized, on-demand data products generated on their specifications.

The gateway will positively impact not only the CEOP program, but also all other programs that use OPeNDAP protocols for data access and want to use the satellite data in *their efforts. The CEOP Satellite Data Server developed in this* project will use the standard WCS protocols to communicate with any OGC WCS compliant server. Therefore, any WCS server can act as an OPeNDAP server by deploying this server which acts as a gateway server. Currently, most space agencies around the world are developing or deploying WCS servers. The availability of such a gateway server will make all satellite data in those space centers immediately available to Earth science research and modeling community who commonly use OPeNDAP protocols for data access. Finally, this project also facilitates the interoperability from geospatial community to the Earth science research and modeling community and makes the use of geospatial technology in the Earth science community.

ACKNOWLEDGEMENT

This research is supported by NASA ROSES 2005 ACCESS program (05-ACCESS/05-28).

REFERENCES

[1] T. Koike, "The Coordinated Enhanced Observing Period – an initial step for integrated global water cycle observation," WMO Bulletin 53 No. 2, April 2004.

[2] Gallagher, J, Potter, N., Sgouros, T., Hankin, S., Flierl, G., 2005, The Data Access Protocol – DAP 2.0, http://opendap.org/archive/design/dap/dap2/dap_2_data_model.pdf

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

[4] Di, L., and McDonald K. 2005. "The NASA HDF-EOS Web GIS Software Suite (NWGISS)." In Qu, J. etc eds, *Earth Science Satellite Remote Sensing, Volume II: Data, Computational Processing and Tools.* Springer-Verlag.

[5] Di, L., 2005. "The Open GIS Web Service Specifications for Interoperable Access and Services of NASA EOS Data." In Qu, J. etc eds, *Earth Science Satellite Remote Sensing Volume II: Data, Computational Processing and Tools.* Springer-Verlag.

[6] Deng D., L. Di, D. Holloway, Y. Enloe, and K. McDonald, 2004. "The Development of an OPeNDAP Satellite Data Server for CEOP". Poster presentation at AGU Fall meeting, December 2004. San Francisco, CA.

[7] Nebert, D. and Whiteside A., (eds.), 2004, OpenGIS Catalog Service Specification 2.0, OGC Document 04-021r2.

[8] Evans, J., 2003, The OGC Web Coverage Service Specification, OGC-03-065-r6, http://www.opengeospatial.org/.

[9] Vretanos, P., (editor), 2004, Web Feature Service Implementation Specification V1.1, OGC Document OGC-04-094.

[10] Unidata, 2005, THREDDS, http://my.unidata.ucar.edu/content/projects/THREDDS/index.html