

Joe Wielgosz*, Brian Doty, and Jennifer Adams
Center for Ocean-Land-Atmosphere Studies

1. INTRODUCTION

The GrADS-DODS Server (GDS) combines GrADS (the Grid Analysis and Display System) (Doty and Kinter, 1992) and OPeNDAP (the Open source Project for a Network Data Access Protocol, formerly DODS - Distributed Oceanographic Data System) (Gallagher and Milkowski 1995; Davis and Gallagher 1999) to create a powerful open-source solution for serving distributed scientific data (Wielgosz et al, 2001).

This paper will discuss the many new capabilities implemented since the initial release of the GDS. These include the addition of station data support, including WMO BUFR format; the ability to perform analysis using both remote and local datasets as input; and a rich set of features designed to enhance reliability, scalability, security, and ease-of-use for both the administrator and the users of the server.

2. OVERVIEW OF THE GRADS-DODS SERVER

The GDS provides a wide range of clients with remote dataset access via the OPeNDAP protocol. It can serve data in any GrADS-readable format, including netCDF, HDF, WMO GRIB and now WMO BUFR as well. Any analysis tool which has been OPeNDAP-enabled - for example, GrADS, Ferret, Matlab, IDL, or the Unidata Integrated Data Viewer - can read the data on a GDS as if the data were stored locally. Metadata and subsets are retrieved transparently from the server as needed.

Web users can access GDS datasets in several ways. The GDS has its own Web interface which allows users to browse dataset directories and metadata, and obtain subsets in ASCII text format. Additionally, the GDS can serve data to a middleware application such as the Live Action Server, providing extensive Web-based visualization and analysis capabilities.

In addition to supporting the OPeNDAP subsetting protocol, the GDS has a powerful server-side analysis capability. Analysis tasks using GrADS functions can be encoded into URLs, which users can then pass into any OPeNDAP client application. When the GDS receives such a URL, it performs the encoded analysis task and stores the results as a new dataset,

which is fully described and accessible via the OPeNDAP protocol. The client can then access the analysis results just like any other dataset on the GDS. Results can even be used as input into further analysis tasks, allowing multi-stage analysis using intermediate calculations.

Finally, the GDS has been designed to be easy to install and use. All of its software components are pure Java, with the exception of GrADS which is ANSI C, providing extremely high portability and stability. As distributed, it is essentially ready to run - its only external software requirement is the Java runtime environment. All server settings are contained in a single configuration file.

Downloads for GrADS software, and current information about the GrADS project, are available at the GrADS

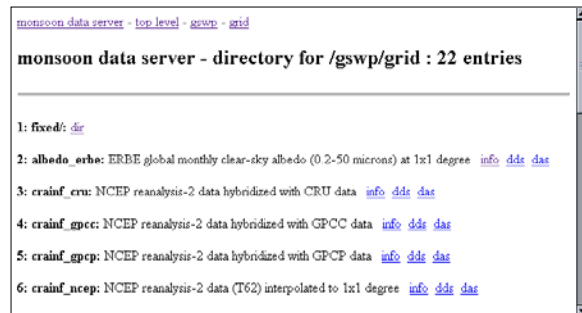


Figure 1: GDS Web interface: Browsing OPeNDAP-accessible datasets (taken from <http://www.monsoondata.org/>)



Figure 2: GDS Web interface: Browsing a dataset summary on the same site

* Joseph Wielgosz, COLA, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705; email joew@cola.iges.org

home page:

<http://www.iges.org/grads>.

Downloads and information for the GrADS-DODS Server can be found at:

<http://www.iges.org/grads/gds>.

OPeNDAP information and software are available at:

<http://unidata.ucar.edu/packages/dods>.

Source code for all software components is freely distributed, and user participation in the development process is encouraged.

3. NEW CAPABILITIES IN GDS VERSION 1.2

The second major release of the GDS, version 1.2, incorporates a number of features which improve the power and flexibility of the software. This release represents a thorough refactoring of the GDS internal design. The architecture of GDS version 1.2 provides significantly higher performance, both during initialization and in handling client requests. It is also highly modularized, allowing for extensibility in the future. The range of configuration options has been expanded, and includes a rich set of resource management tools. Lastly, the GDS can now access remote datasets for use as input into analysis tasks.

3.1 Modular design

Internally, the majority of the GDS 1.2 software consists of generic modules, collectively nicknamed Anagram, which perform such functions as handling the servlet interface, reading configuration information, and managing security restrictions. The Anagram modules are connected via a generic Java interface to a module containing GrADS-specific implementations of OPeNDAP subsetting and analysis functions.

Because the interface between Anagram and the GrADS-specific code consists of a small, carefully defined set of methods, the potential exists for future development of interfaces to analysis tools other than GrADS. The resulting data servers could provide OPeNDAP clients with robust access to new ranges of data formats and analysis capabilities.

3.2 XML Configuration

Configuration is now done via an XML text file, and includes a wide range of additional options. For example:

- Administrators have the ability to customize text and hyperlinks on the GDS Web interface.
- Large collections of datasets can now be served using only a single XML tag, which specifies the directory tree where the data collection is to be found.

- Custom metadata can be added to datasets, including hyperlinks to documentation resources, as well as arbitrary descriptive attributes.
- Logging options include multiple detail levels, rotating filenames, and customizable time stamping.
- Custom security restrictions can be set on groups of datasets for browsing, subsetting and analysis; these restrictions can then be mapped to multiple IP address ranges.

The modular design of the server allows it to re-configure on-the-fly, without shutting down and restarting. Re-configuration is triggered via a URL-based administrative interface, allowing the administrator to control the server without taking it offline.

3.3 Resource management

A design feature which distinguishes the GDS from many distributed data servers is its emphasis on administrative control of resource usage. The administrator can set a wide range of configuration parameters in order to limit the disk, memory, and CPU resources consumed by the server, and set specific levels of resource-usage privileges for individual users. These include customizable limits on:

- maximum simultaneous client connections
- CPU time available per analysis task
- maximum requestable subset size
- disk space for caching of analysis results

On the World Wide Web, malicious denial-of-service attacks have caused significant problems for content providers. The same potential problems exist for distributed data providers. In fact, because clients in a distributed data system may frequently be operating without human supervision, there is a potential for even a well-intentioned user to generate an unmanageable volume of requests to a data server. For example, a script containing a conditional loop might fail to terminate properly, and thus generate an endless series of subset requests.

The GDS has the ability to block IP addresses for a fixed period of time after they individually exceed a hit threshold set by the administrator. A client which has been blocked receives an error message explaining that they have been blocked and indicating the duration for which the block is in effect. This mechanism provides GDS administrators with a defense against intentional or unintentional denial-of-service conditions.

3.4 Intermediary to remote OPeNDAP datasets

Because all data requests in the GDS are passed to GrADS, and GrADS is itself an OPeNDAP client, it was a natural step to allow the server administrator to provide URLs for remote OPeNDAP datasets, as well as local file names, in the GDS configuration file. The GDS thus can act as an intermediary to other OPeNDAP data servers

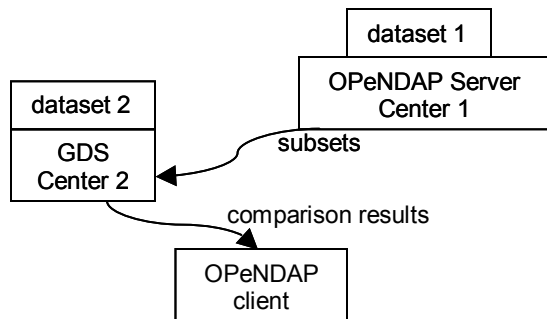


Figure 3: Potential use of a GDS to do comparative analysis between two data centers

anywhere on the Internet. When clients send data requests to the GDS for such datasets, the requests are passed off to the appropriate OPeNDAP server elsewhere.

The power of this feature lies in the ability of the GDS to perform analysis tasks using remote datasets exactly as for local ones. Administrators can therefore use this feature to add analysis capability to any existing OPeNDAP data server, or to enable server-side comparison between their own data and a collaborator's data.

For example, two data centers with a high-bandwidth connection might set up GrADS-DODS Servers as intermediaries to each other's data.

Under this arrangement, clients could use either center's GDS to perform comparative analysis between the two centers' data holdings. The necessary analysis inputs would be automatically transferred across the high-bandwidth connection from one data center to the other. The client would thus avoid transferring either of the input datasets to their own machine; they would only need to download the final result of the analysis.

4. STATION DATA SUPPORT

The terms in-situ, observational, profile, or station data are often used to refer to data which does not conform to a gridded representation. In GrADS terminology, such datasets are referred to as station data.

4.1 GrADS and GDS enhancements

GrADS version 1.9 includes significantly expanded station data support. In particular, two key

enhancements have been made: support for the WMO BUFR format, and for accessing station data using the OPeNDAP protocol.

Simultaneously, the GDS has been enhanced with the capability to serve any station data supported by GrADS. GDS administrators can therefore provide remote access to data stored in either WMO BUFR or GrADS station data format, or act as an intermediary to station data on another OPeNDAP server.

The GDS allows clients to subset station data by providing either geospatial and time ranges or textual station identifiers. In addition, the OPeNDAP constraint language can be used to form complex constraints based on the station data itself; for example, retrieving only those weather stations where surface temperature is below freezing.

GrADS provides the ability to intercompare gridded and station datasets. With GrADS 1.9, a user can therefore access remote station data from a GDS, and compare it with their own gridded model output, stored either locally or remotely on an OPeNDAP server.

This intercomparison capability is also available for server-side use as well. Thus station datasets can be used along with gridded datasets in server-side analysis tasks.

With the features available in GrADS 1.9 and GDS 1.2, GrADS and the GDS now form an end-to-end distributed data system that handles both gridded and station data side-by-side.

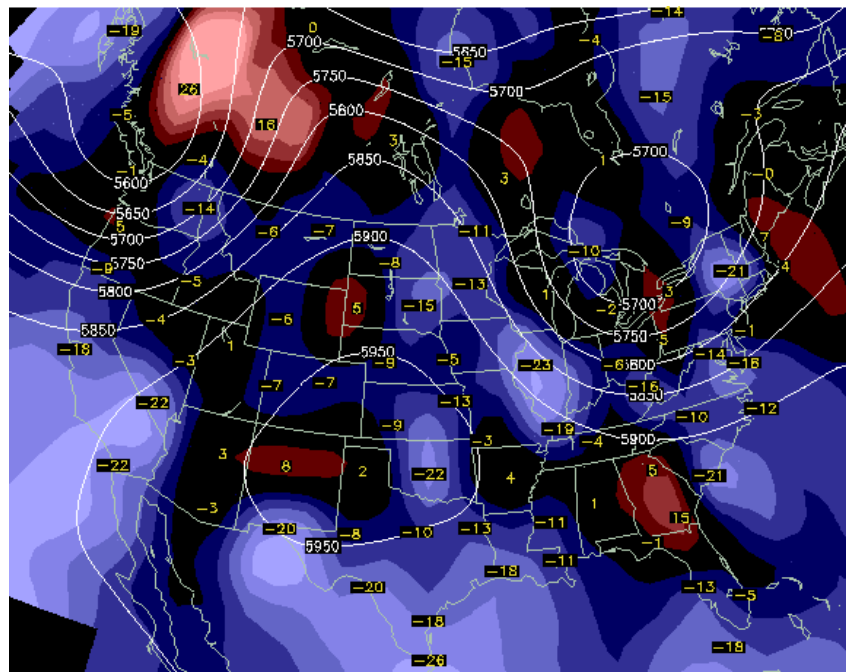


Figure 4: A shaded contour plot in GrADS, showing the difference between station observations, and values from a gridded field interpolated to the station locations. White contours show the original gridded field. Both the station and gridded datasets are stored on a GrADS-DODS Server. The script which produced this plot is online at <http://www.iges.org/grads/gds/oberr.gs> – try it yourself!

4.2 Client API's for OPeNDAP station data

Efforts have also been made to guarantee interoperability with other OPeNDAP-based clients and servers.

GrADS 1.9 incorporates a library developed at COLA, called gadods, which simplifies the task of accessing station data over OPeNDAP. The gadods library is an ANSI C compatible wrapper around the OPeNDAP core libraries. It provides a simple, documented API that is specifically designed for querying station datasets, and thus reduces the burden on client application programmers for supporting station data access via OPeNDAP.

In addition, COLA and the OPeNDAP team have collaborated on the design of a translation mechanism for the OPeNDAP-enabled netCDF library. This translation mechanism will provide access to station data even in clients that support exclusively gridded data types, by representing station data subsets as gridded variables and serving them through the netCDF API. A number of parameters can be used to control the translation, thus allowing the user to choose the translation options that work best for their client.

5. USAGE OF THE GDS

The GDS has achieved significant operational usage, both at COLA and in other groups in the modeling community. Production use at COLA has demonstrated its ability to handle high data volumes and client demand levels.

COLA's operational GDS provides public access to approximately 200GB of continuously updated gridded data, including NCEP MRF, ETA, and AVN model outputs, and the Global Offline Land surface Dataset (GOLD). This data server has consistently received over 300,000 hits per month since September of 2001 (see <http://www.iges.org/stats/gds> for detailed site statistics).

The GDS is also one of the core technologies for NOMADS (NOAA Operational Model Archive and Distribution System, <http://nomads.gfdl.noaa.gov/>) (Rutledge et al., 2002), and has been successfully installed at collaborating centers such as the National Climate Data Center (NCDC), the NOAA-CIRES Climate Diagnostics Center (CDC), the Princeton University Geophysical Fluid Dynamics Laboratory (GFDL), and the National Center for Environmental Prediction (NCEP).

Other groups which are using or considering the GDS for data distribution include the National Virtual Ocean Data System (NVODS, <http://www.nvods.org/>), the Goddard Space Flight Center Land Information System (LIS, <http://lis.gsfc.nasa.gov/>), the Global Ocean Data Assimilation Experiment (GODAE, <http://usgodae.org/>), and the Goddard Space Flight Center Distributed Active Archive Center (GDAAC).

6. FUTURE DEVELOPMENT

Near-term plans for GDS development include adding a facility to interact with mass storage devices, allowing near-line data to be served via OPeNDAP; and the addition of a login-based authentication mechanism. Planned enhancements to GrADS include extended support for netCDF and HDF data, and automatic support for and interpolation between a wide variety of projected geospatial grids.

REFERENCES

- Davis, E. R., and J. Gallagher, 1999: Using DODS to Access and Deliver Remote Data, Fifteenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, Dallas, TX.
- Doty, B., and J. L. Kinter III, 1995: Geophysical Data Analysis and Visualization Using GrADS. Visualization Techniques in Space and Atmospheric Sciences, eds. E. P. Szuszczewicz and J. H. Bredekamp. (NASA, Washington, D.C.), 209-219.
- Gallagher, J., and G. Milkowski, 1995: Data Transport within the Distributed Oceanographic Data System, World Wide Web Journal, Fourth International World Wide Web Proceedings, 691-702.
- Rutledge, G.K., D. Williams, R. Stouffer, J. Alpert, L. Buja, B. Doty, S. Hankin, B. Domenico, M. Kafatos, 2002, The NOAA Operational Model Archive and Distribution System (NOMADS), Proceedings 13th Symposium on Global Change and Climate Variations, American Meteorological Society, Orlando FL. J76 pp.
- Wielgosz, J., B. Doty, J. Gallagher, and D. Holloway, 2001: GrADS and DODS, Seventeenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, Albuquerque, NM.