

## P2.9 A METADATA INTEGRATED DATA ANALYSIS SERVER (MIDAS): ENHANCEMENTS AND UPDATES

Yujie Zhao, Ruixin Yang\*, Menas Kafatos

Center for Earth Observing & Space Research (CEOSR)  
George Mason University (GMU)

### ABSTRACT

A metadata server, DIMES (DIstributed METadata Server), and a data and analysis server, GDS (GrADS/DODS Server), are integrated to form a Metadata Integrated Data Analysis Server (MIDAS) for interactive access to both metadata and data. The power of MIDAS is reflected by the consistency between the data server and the metadata server and by the closely coupling between them. Middle layers are added to combine the server systems and to help users to generate GDS URLs. Therefore, MIDAS provides seamless way to search, access, and analyze the ever-increasing distributed Earth observing and model data.

MIDAS is planned to be used in an operational mode. Various modifications and enhancements on performance, on query implementations, and on interfaces are carried out based on users' recommendations. In this article, we will summarize the recent enhancements on MIDAS, and discuss existing issues for a potential operational usage.

### 1. INTRODUCTION

As more and more data for Earth science research are available from Earth observing, in particular, satellite remote sensing, and from models, efficient mechanisms for finding and delivering distributed data become necessary. One well-known data delivery infrastructure is the Distributed Oceanographic Data System (DODS) originating in the oceanography community (DODS, 2004), or OPeNDAP as recently renamed (OPeNDAP, 2004). Plain DODS was enhanced by being combined with the data analysis capability of

GrADS (the Grid Analysis and Display System) (Doty *et al.*, 1997) to form GDS (GrADS/DODS Server or GrADS Data Server) (Wielgosz *et al.*, 2001), which allows users to define operations performed on the server side and to obtain the resultant information (processed data) via the Internet.

The XML-based DIMES (Yang *et al.*, 2001) implemented a flexible metadata model and web-based metadata navigation interfaces to support various level metadata accesses. A Metadata Integrated Data Analysis Server (**MIDAS**) was developed by combining DIMES with GDS, initially named a Scientific Data and Information Super Server (SDISS) (Yang, Kafatos & Wang, 2002; Yang *et al.*, 2003). In the following, we will review the MIDAS architecture and discuss the recent enhancements, especially on the search interface and on serving the operational NOMADS (NOAA Operational Model Archive and Distribution System) nodes at NCEP (National Centers for Environmental Prediction).

### 2. MIDAS ARCHITECTURE AND MAJOR COMPONENTS

Figure 1 is the high-level system architecture of MIDAS. Major components of the system are certainly GDS and DIMES. A GDS URL Generator is included in the architecture to help users to build the relatively complex GDS URLs through a GUI. The MIDAS is designed to be a distributed system, and therefore a register can be used to record all available MIDAS. Certainly, the MIDAS register itself can be distributed providing information on each server or even residing on the client side. However, the centralized register will make it easier to reach broader audiences by

---

\* Corresponding author address: Ruixin Yang, MS 5C3, School of Computational Sciences, George Mason University, Fairfax, VA 22030; e-mail: [ryang@gmu.edu](mailto:ryang@gmu.edu).

leveraging the existing centralized metadata search engines such as GCMD (Olsen & Major, 1996). On the server side, setting up a MIDAS starts both the data server and the

metadata server. Metadata in a MIDAS will be ingested and checked to reflect the changes on the data holdings served by the current data server via the ingest tool box.

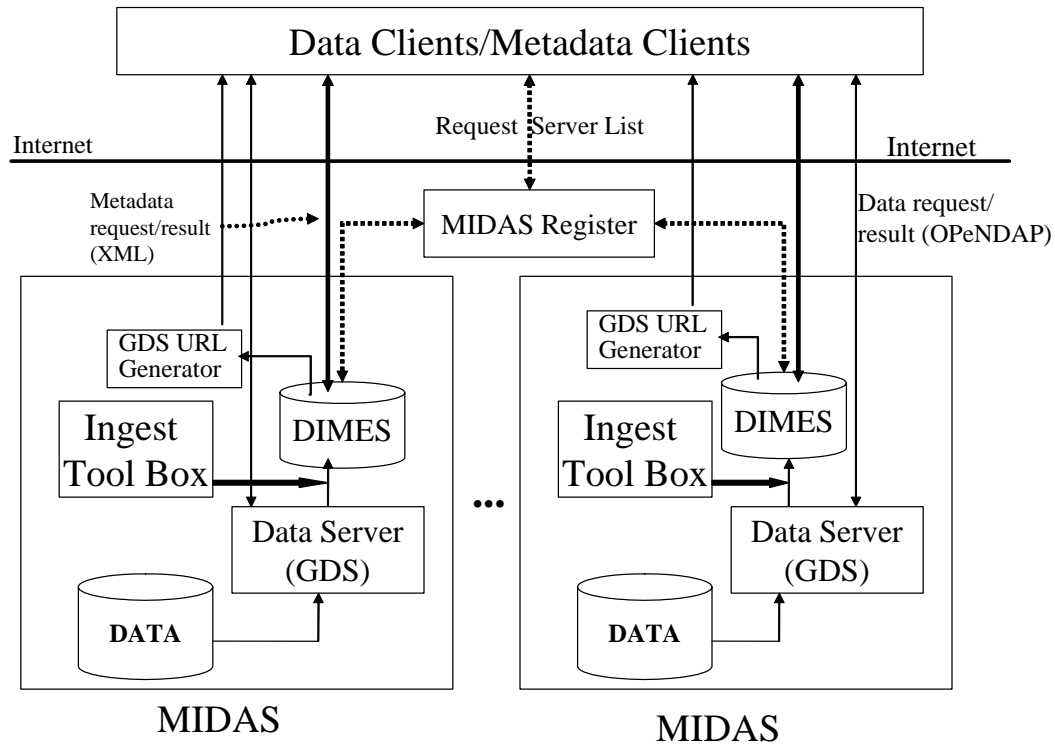


Figure 1. The high-level system architecture of MIDAS.

### 3. METADATA SEARCH INTERFACE

One important function of a DIMES server is that users can retrieve the metadata managed by it. A DIMES server may manage the metadata of thousands of data sets. One example is the DIMES server for NOMAD2 node at NCEP, which manages the metadata of about 2,500 data sets. And the retrieval function of DIMES facilitates users to find out the metadata they want. There are five basic criteria that users can follow to search metadata through a DIMES server: temporal resolution, spatial resolution, temporal coverage, spatial coverage, and text. Besides these, users can build up knowledge information into the metadata, and DIMES servers support the search of them, too.

The metadata search interface is the entry point to retrieve the metadata in a DIMES server. Figure 2 is a metadata search interface

used by the DIMES server at GMU (<http://spring.scs.gmu.edu:8199/servlet/VASearchInterface>). The first box, "Customize Search Panel," summarizes the primary search criteria. Users can select one or several of them to tune their search criteria. After selecting the search criteria, users can click the "Refresh" button, and then the special boxes for the selected criteria will be generated. For example, in Figure 2, three boxes, "Search\_Space" (for spatial coverage), "Search\_Time" (for temporal coverage), and "Search\_Text" (for text), are generated. Users can fill out the search interface with the parameters that confine the metadata they want, and then click the "Submit" button to send a query to the DIMES server. For the temporal coverage and spatial coverage, we set up five options as the pull-down menu in Figure 2, and users can use any of them during their retrieval.

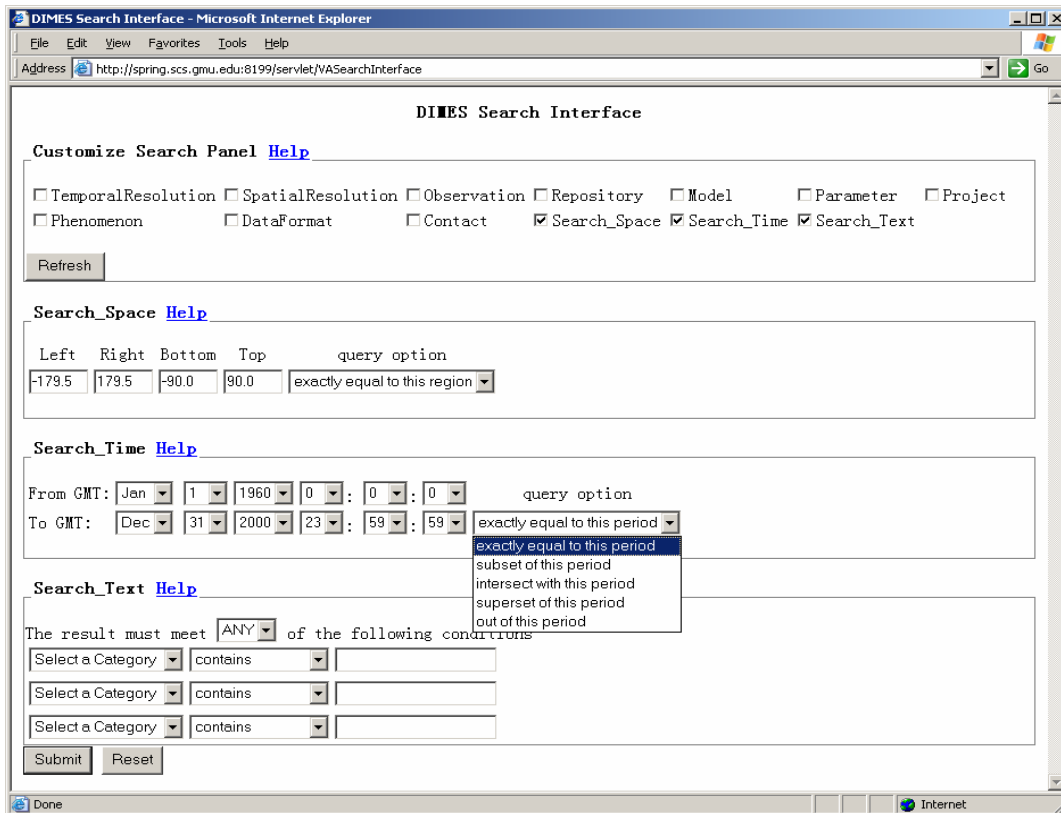


Figure 2. Metadata Search Interface of DIMES.

NOMADS Server	Dataset Number	The First Ingestion time	Daily Update Time
Nomad1	About 1,100	About 13 min	About 1 min
Nomad2	About 2,500	About 38 min	About 6 min
Nomad3	About 1,200	About 40 min	About 5 min

Table 1. NOMADS GDS Servers and their ingestion time

#### 4. DIMES SERVERS FOR NOMADS

NCEP plans to set up five NOMADS GDS servers, and three of them are active now. We set up three DIMES servers corresponding to each of them. The web page <http://spring.scs.gmu.edu:8199/nomad/index.html> summarizes all the URLs of the DIMES servers for NOMADS GDS servers.

Combining with the NOMADS GDS servers is an important practical application of DIMES. Every NOMADS GDS server has a large amount of data sets (see Table 1), which requires that DIMES servers should have high

ingestion and query performance. We significantly improved the ingestion and query efficiency of DIMES in this enhancement phase to make DIMES satisfy this requirement.

NOMADS GDS servers add new data sets and delete old data sets every day. In order to keep DIMES servers consistent with NOMADS GDS servers automatically, we added an automatic update part to DIMES so that we can set up the update time points and DIMES servers can update themselves automatically at these time points.

Table 1 is a statistics of NOMADS GDS servers and related ingestion time. From the table, we can see that daily update time is short although the first ingestion time is long.

## 5. FUTURE WORK

In the future, we have two potential directions to improve DIMES. The first one is to improve the display of the query result generated by the metadata search interface. Now, a query result is returned as a long list, and usually needs tens of web pages to display. In order to find out the needed data sets, a user has to browse the entire list. If the query result could be classified into categories and organized into tree structure based on these categories, the query result would be displayed in one web page and a user could go to the right categories directly to get the needed data.

The second one is to enhance the retrieval function of the metadata search interface by combining semantic information. GDS uses abbreviations, e.g. "tmp" for "temperature". Because DIMES ingests metadata from GDS, DIMES includes these abbreviations in it, too. One disadvantage of these abbreviations is that it affects the retrieval function. For example, if a user uses the key word "temperature" to search metadata, then he/she will get nothing because DIMES does not have "temperature" in it. If we could set up a semantic corresponding table to match "temperature" with "tmp," then DIMES would use both "temperature" and "tmp" to search metadata when a user searches for "temperature."

## ACKNOWLEDGMENTS

The authors would like to thank Jordan Alpert and June Wang of NCEP for their suggestions on improving the system.

## REFERENCES

DODS 2004: "Distributed Oceanographic Data System." <http://www.unidata.ucar.edu/packages/dods/> (last accessed on October 27, 2004).

Doty, B. E., J. L. Kinter III, M. Fiorino, D. Hooper, R. Budich, K. Winger, U. Schulzweide, L. Calori, T. Hol, and K. Meier, 1997: "The Grid Analysis and Display System (GrADS): An update for 1997: " 13<sup>th</sup> Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, pages 356-358 (American Meteorological Society, Boston).

Olsen, L.M., G. R. Major, "Global Change Master Directory Enhances Search for Earth Science Data," 1996, Transactions of the EOS, American Geophysical Union, [http://www.agu.org/eos\\_elec/95127e.html](http://www.agu.org/eos_elec/95127e.html) (last accessed on October 17, 2003).

OPeNDAP, 2004: "What is OPeNDAP?" <http://www.opendap.org/> (last accessed on October 27, 2004).

Wielgosz, J., B. E. Doty, J. Gallagher, and D. Holloway, 2001: "GrADS and DODS," 17th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Jan. 2001.

Yang, R., X. Deng, M. Kafatos, C. Wang and X. Wang, 2001: "An XML-Based Distributed Metadata Server (DIMES) Supporting Earth Science Metadata," in Proceedings of the 13th International Conference on Scientific and Statistical Database Management (L. Kerschberg and M. Kafatos, eds.), pages 251-256, IEEE, Computer Society.

Yang, R., M. Kafatos, and X. Wang, 2002: "Managing Scientific Metadata Using XML," *IEEE Internet Computing*, v6, no.4, pp. 52-59.

Yang, R., X. Wang, Y. Nie, Y. Zhao and M. Kafatos, 2003: "A Web-Based Scientific Data and Information Super Server with A Flexible XML Metadata Support," in Proceedings of the 19th International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society, *Long Beach, California, February 9-13, 2003 (CD-ROM)*.