A DISTRIBUTED ENHANCED SERVER FOR MULTIDIMENSIONAL SCIENTIFIC DATA

Faced with large and growing data sets from diverse sources (including general-purpose search engines that provide unwieldy results), the authors combine metadata search and data access concepts to create a prototype enhanced server tuned to scientific data search and accessibility.

The dawn of the 21st century has seen an unprecedented amount of remote-sensing data collected. Around the world, national agencies have launched a series of Earth-observing and remote-sensing missions, all of which contribute to the amount of data available for scientific analysis. NASA's Earth Observing System satellite alone contributes more than half a Tbyte of data to various databases every day.[1]

In most cases, Earth science data comes in multidimensional arrays, but many variations in data format exist (such as map projections and spatial and temporal resolutions). Information volume and complexity are continuing to grow, together with requirements and usage stringencies imposed by different science communities in different environments. Using traditional data access and analysis methods, Earth science researchers will find it increasingly difficult to retrieve, analyze, and present results from such massive amounts of data. Scientists need distributed data information systems with effective search and access capabilities to perform their research more efficiently.

Finding useful data sets in the first place is difficult. Commonly, people use metadata to search data sets. A major problem with current metadata systems serving the Earth science community, however, is that the search engines—usually large and supported by national centers—contain too much information. Scientists get too many hits for a specific search, and some of the data links are out of date.

Under the Seasonal to Interannual Earth Science Information Partnership (SIESIP) program,[2] we developed an XML-based distributed metadata server (called DIMES)[3] comprising a flexible metadata model, search software, and Web-based interfaces to support multiple-level metadata accesses. The major difference between our metadata model and others' is that during development, we kept small data providers in mind by requiring minimum metadata standard.

DIMES is a flexible system that can be tailored easily for specific uses and integrated with

1521-9615/03/$17.00 © 2003 IEEE
Copublished by the IEEE CS and the AIP

RUIXIN YANG AND MENAS KAFATOS
George Mason University

BRIAN DOTY AND JAMES L. KINTER III
Center for Ocean-Land-Atmosphere Studies

LONG PHAM
NASA Goddard Earth Science Data and Information Services Center, Distributed Active Archive Center

other systems. To overcome the central metadata servers' large data volumes and the data servers' search requirements, we integrated the metadata and data systems to give users consistent access to both data and metadata. Following this strategy, we designed an enhanced server based on DIMES and GDS.[4] This ES combines metadata and data interoperability solutions to enhance data accessibility.

The emerging resource-sharing Grid technologies will provide a more general mechanism on data sharing over the network. Nevertheless, the system described here is based on mature and well-accepted technology in the specific science communities. Domain scientists trust only proven environment and technologies for conducting serious research.

Glossary of Terms

DIMES	Distributed metadata server
DODS	Distributed Oceanographic Data System
ES	Enhanced Server
GDS	GrADS/DODS Server
GrADS	Grid Analysis and Display System
GRIB	GRIdded Binary
SIESIP	Seasonal to Interannual Earth Science Information Partnership

Our Data Interoperability Access

The SIESIP consortium, which grew from one of 12 original, science-focused ESIP projects created by NASA,[5] consists of three main distributed sites:

- George Mason University provides expertise in information technology, data searches and analysis, and interdisciplinary Earth systems science
- The Center for Ocean-Land-Atmosphere Studies (COLA) provides expertise in seasonal to interannual (S-I) science, user services, and tools
- NASA Goddard Earth Science Data and Information Services Center, Distributed Active Archive Center provides expertise in data management, data archiving, and user services

Under SIESIP, we developed a three-phase data search and access model to provide data and information to users. In phase 1, the user browses the data holdings via the metadata provided by the SIESIP system. We provide information-rich products, incorporating knowledge about data products in the system. In phase 2, the user gets a quick estimate of the type and quality of data found in phase 1. We integrated analytical tools with statistical functions and visualization algorithms into a Web-based SIESIP online data-analysis system. With this interface, SIESIP users can browse data and perform certain predefined analysis functions on data. Finally, in phase 3, the user has located the data sets of interest and is ready to order. If the data are available through SIESIP, it will handle the data order; otherwise, SIESIP issues an order to the appropriate data provider on the user's behalf or forwards necessary information to the user to complete the order.

Although each phase is equally important for data users, the data access–order phase plays a special role because the goal of most science users accessing a data system is to locate and access data to perform their research.

Users generally access data via one of three levels of data interoperability. The first level is simple data transmission such as FTP or basic Web capabilities. Users of this level of data services can get results quickly and easily but with little metadata or choices on the resultant data types and formats. In the next data interoperability level, servers provide data and some metadata in a general way. Users can make some subset and data format choices on the resultant data and integrate those results with other data. The final level of data interoperability is a "stateless" analysis server. Users issue analysis requests, the server processes the data on the server side, and then the server returns the resultant data. Data servers with this level of service let users distill the data in greater detail while significantly reducing network traffic.

Suppose that we have a data set called *sea surface temperature anomaly* (SSTA) that has monthly temporal resolution and a $1.0^o \times 1.0^o$ (longitude and latitude) spatial resolution. Furthermore, let's assume that the data cover the globe for 50 years. If we use IEEE binary for each SSTA value, we will have 4 bytes $\times$ 360 $\times$ 180 = 259,200 bytes for each month for a total of about 155 Mbytes of data for all values of this parameter (assuming missing values are filled in with special 4-byte values such as IEEE Not a Number (NaN).

Now, suppose a user wants to have a 50-year time series of SSTA averaged between $5^oS – 5^oN$ latitude and $150^oW – 90^oW$ longitude (the Niño 3

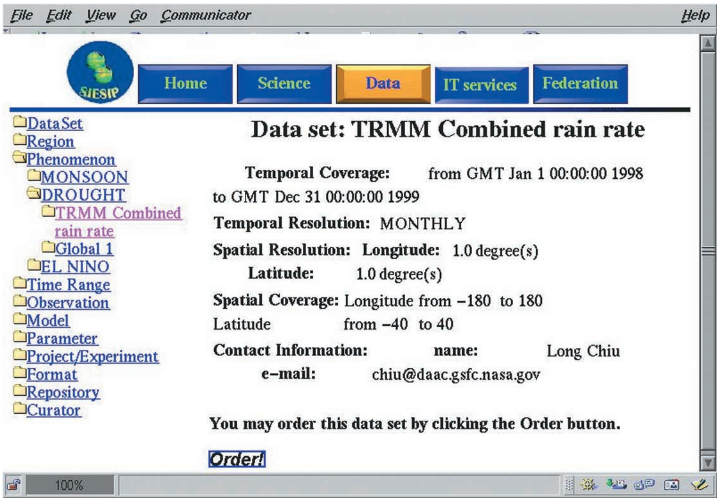

Figure 1. The distributed metadata server (DIMES) navigation interface. Web browser users start their searches by clicking on folders, words, or buttons.

area). The user must process a subset of the original data, approximately 1.4 Mbyte (4 bytes × 60 longitude × 10 latitude × 50 year × 12 month). The analysis server result is a 600-month time series of data in 4-byte format that is only 2.4 Kbytes in length. If, however, a server doesn't support analysis, the user must download the entire 155 Mbytes of data. If no server-side data manipulation support exits, the user still must move 1.4 Mbytes to the desktop and process the data to obtain the time series. With a "stateless" data analysis server, the user could reduce the transferred data volume from hundreds of Mbytes to a few Kbytes.

Another potential data interoperability level consists of a "session-oriented" analysis server. On such servers, users can keep a session open (as long as they are using the same server) to reuse previous results and save work completed in one session and use it in another one. Servers with sessions let users pursue a relatively large data processing project on distributed servers. The "session-oriented" server is useful for science communities but needs more complicated resource management. We are unaware of a successful session-oriented server in the seasonal-to-interannual research community.

In the SIESIP project, we developed systems to support all three phases of data accesses with data interoperabilities at the first three levels.

Our Distributed Metadata Server

A common weakness of many existing Earth science distributed information systems is the lack of metadata interoperability support. To support metadata interoperability, the system should facilitate metadata searches from different sources and providers without requiring major, if any, developmental effort from the user. This requires a flexible metadata model capable of capturing diverse sources and formats of metadata. XML is one solution for this, so we developed our metadata model using XML technology.

DIMES uses a flexible metadata model, and consists of a metadata depository and a software package for querying the metadata. In the DIMES metadata model, metadata concepts are treated as XML nodes, with IDs assigned to nodes. We used XML structures and special links such as `refer_to`, `type_instances`, and `node_types` to preserve and create relations among nodes.[4] Users can define the relations, so domain knowledge could be included in DIMES for metadata search and navigation.

The XML Query Engine

In addition to the XML-based metadata model and the corresponding XML metadata documents, another major part of DIMES is an XML query engine that is responsible for answering queries against the XML metadata. DIMES's principal design goal is flexibility, so we designed the query engine using that same strategy.

Instead of defining fixed queries for specific purposes, we identified a few fundamental types of queries and used them to answer complicated user requests. The fundamental queries include the simplest one for obtaining a node's content via its ID as well as basic queries for searching nodes based on standard conditions (such as spatial resolution, temporal resolution, spatial coverage, temporal coverage, and textual conditions, including keywords and free-text search).

Search Mechanism

In DIMES, we also introduce a powerful, yet simple search mechanism that exploits the XML structure and our extended metadata semantics. We call it the nearest-neighbor (or shortest-path) search.[6] As its name implies, if we consider each node as a concept in the metadata, the nearest-neighbor search finds the closest neighbor concepts for the one specified. This is especially useful to support metadata interoperability—that is, when incorporating various kinds of metadata from diverse data sources, the system provides a way for the user to discover metadata information that is closely related to the user's known knowledge or focus point.

Because the nearest-neighbor search does not require any a priori knowledge about the metadata, it provides a good starting point for ordinary users. It also supplies data providers with a tool to accommodate special metadata and to make them searchable without modifying the metadata structure. A sample query using the nearest-neighbor search is to find all data sets that are related to phenomenon "El Niño."

A special variation of the nearest-neighbor search is a `tree expand` query. As we mentioned earlier, DIMES metadata contains a set of interlinked nodes. Each node has some nearest-neighbor nodes. If we choose one node as a root, all its nearest neighbors as the first level branches, and so on, we generate a tree-like presentation of all the nodes. In other words, the `tree expand` query reorganizes a graph (nodes linked with different relations) into a tree with a specified root node.

The Java-based DIMES search engine uses the TCP/IP socket as its base communication mechanism to use agents. Because we could not expect anyone to use the search engine based on only a socket implementation, we added a layer to accept XML queries through the HTTP protocol.

Navigational Interface

One of the interfaces we developed is for regular searches against DIMES. The interface is similar to other Earth science data search interfaces, which let users type in or select spatial–temporal resolution, spatial–temporal coverage, and field and free text searches. In addition to this, we use the fundamental queries just discussed—in particular, the `tree expand` query—to build a metadata navigation system.

We designed the navigation system based on the results produced from extensive and iterative interactions with several Earth scientists, so the system reflects how scientists navigate and browse in real life and how they query Earth science metadata and data sets.

The relations among the nodes in metadata are organized into a metadata tree for users to browse through metadata. The tree structure groups nodes such as "data sets" by various categories. Each category is considered as a dimension in a multidimensional database, and users can start their browsing through any dimension.

The initial (default) page is a list of categories, and there is a folder symbol in front of each category name (see Figure 1). The category names and the folder symbols are clickable, so Web browser users can start their metadata journey from there. Each click on the clickable items (folder, word, or button) is a request to the server.

An Online Data Analysis System

To support phase 2 data access, the SIESIP team developed an online data analysis system. The system's client-server architecture uses the World Wide Web and expands on a previous workable prototype system.[7] The system is science driven and serves science communities. It provides data and information products as well as data visualization and analysis and user-support capabilities. The integrated system consists of components of data products, a database management system, communication protocols, data analysis tools, and user interface modules.

The data sets held in this system are multidimensional Earth science data with uniform spatial and temporal resolutions. Typical examples are the monthly mean data of sea surface temperature and the Normalized Difference Vegetation Index over global scales, both from NASA's Climatology Interdisciplinary Data Collection,[8] and air temperature and precipitation for South America. The data are saved in the file system as binary files (one for each month and each parameter) for online data analysis and data ordering. The corresponding statistical summary data and holding catalogue are inserted into the database management system (DBMS).

We use a client-server architecture to support the online data analysis. The server program listens to commands from the client and performs the necessary steps to satisfy requests. After receiving a request, the server parses the command, queries the DBMS for the right files, and invokes GrADS,[9] a free, public-domain tool that does the data analysis and and image generation. The tool is fully automatic and transparent to the extent that users do not know and don't need to know what tool the system is using other than the analysis functions used.

The system's client is a Web page containing a GUI applet based on the Java Swing technology (see the left part of Figure 2 on next page). The workplace menu list gives users control on choosing working data sets for further analysis and other functions. A user starts the system at the SIESIP URL (http://esip.gmu.edu/siesip/gui/current). Instead of using the default settings, users can choose the "Workplace" menu list to find a specific parameter. Then, they can access "region" and "time" panels to restrict spatial and

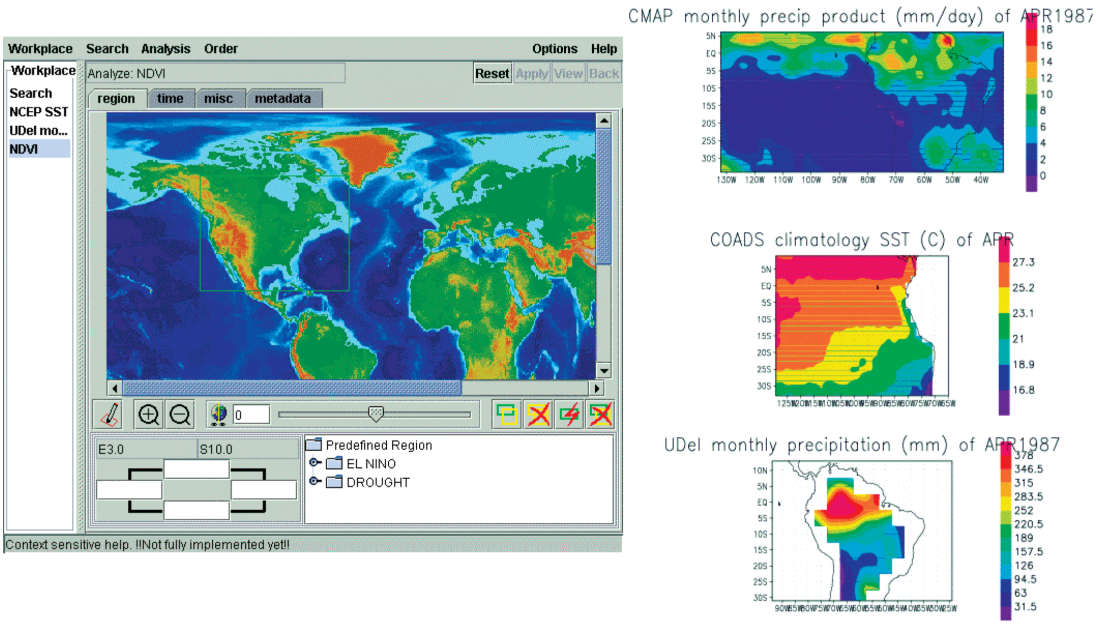

Figure 2. The online data analysis system interface and sample results. The right part of the figure is a sample browsing result from the SIESIP online system. These are simply browse images generated on the fly. However, the three images are generated based on real distributed data sets. With one command through the interface, the user can get three figures in one Web page without knowing the physical locations and transmission protocols of the three data sets.

temporal coverage. Finally, users can select from the "Analysis" menu list to submit an analysis request, such as "Browse Image." After the server completes the requested job, it will send back the necessary information (a URL) to the client, which will automatically pull another Web page displaying the result, as shown in the right part of Figure 2.

A Data Analysis Server

The Distributed Oceanographic Data System (www.unidata.ucar.edu/package/dods) is a robust, client-server data transport protocol based on HTTP. With a DODS-enabled application program such as Matlab, a scientist can open a data set with a URL instead of a local file name. Think of DODS-enabled application programs as special Web browsers.

These special browsers receive data though HTTP and handle the data based on the applications programs' capabilities, such as data analysis and visualization. DODS's simple design and ease of use have led to its widespread adoption by many Earth science data providers.

As we mentioned earlier, GrADS is an open-source desktop tool that integrates the analysis and display of a variety of geophysical data.[9] Under SIESIP, we combined DODS with GrADS to apply GrADS over the Internet. This combined GrADS/DODS server, called GDS,[10] extends DODS to include additional data formats and enables on-the-fly, server-side data analysis and manipulation. Scientists who use GDS can overcome the delays inherent on the relatively low-bandwidth Internet by distilling data to its essence at the server.

With the GDS, any data set that GrADS can handle can also be served over the network. Of particular note is the full support for GRIdded binary (GRIB), the World Meteorological Organization's format for the storage of weather information and the exchange of weather product messages in gridded binary form. GrADS handles GRIB data as a five-dimensional data set (longitude, latitude, vertical level, time, and variable) and provides the ability to efficiently retrieve arbitrary subsets. A DODS client can access subsets of GRIB data without even being aware that the data is in that format. GrADS also provides extensive support for treating multiple disk files as a single data set. Because of this, the data administrator can split files into conveniently sized chunks but still treat them as one data set from the user's perspective.

Enhanced Server

Separately, the DIMES server, the SIESIP online analysis system, and GDS support metadata search, data browsing, and data access, respectively. Together, they match the three-phase data search and access model defined by the SIESIP team.

This combination satisfies scientists' need to search through large numbers of multidimensional data sets to generate and verify scientific hypotheses. However, scientists do not really care much about the details of data format, location, storage media, and so on as long as they can access and use the data effectively and with ease. Without a distributed data system, scientists must find, download, and struggle with data formats and other details to finally use the data. A data system should free scientists from such nonscientific tasks and provide data interoperability.

One natural solution is to integrate the separate software components to give scientists a distributed data system with seamless data access. For that purpose, we integrated DIMES and GDS to form an enhanced server.[4] ES supports interactive access to both metadata and data. Its power is reflected by the consistency between the data and metadata servers and by their close coupling. Our goal is to maintain constant consistency between the data server holdings and the DIMES metadata contents. As a result, we can guarantee that you will find a data set through the metadata engine and that all data sets you find through that engine will be accessible through the data server. Figure 3 illustrates the ES system architecture. GDS and DIMES are in prototype operation but the ingest toolbox is under development as indicated by the dotted line.

ES supports both metadata search and data access—that is, phase 1 and phase 3 data access. To connect the two procedures smoothly, we used phase 2 data access to "glue" them together. As we claimed before, GDS lets users leverage GrADS's power to manipulate data on the server side before ordering the final products.

The GDS URLs, which play the file name role to users, are quite complex. For GrADS users who are familiar with DODS, writing a GDS URL is straightforward. However, for non-GrADS DODS client users, the GDS URLs are too complex to be used easily. To help novice GrADS users, we developed a Web-based GDS URL generator with predefined functionalities. On the GDS URL generator interface, we display the browse images based on users' selection of physical parameters and let users make further data selections based on the browsing image. By doing that, we effectively use phase 2 data access to connect phase 1 and phase 3.

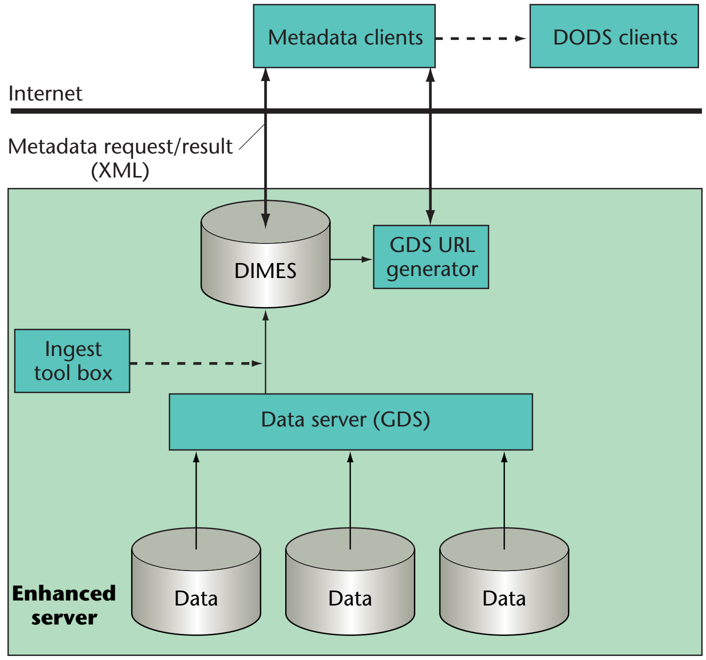

Figure 3. Enhanced Server system architecture. The dashed arrow line between the metadata and DODS clients indicates that manual actions are required to use the ES at present.

The number of predefined functions is always limited. The current GDS URL generator prototype includes only two functions: spatial average and temporal average. We expect that the generator will help GDS users get familiar with the GDS URLs and then modify the sample URLs to create more specific ones.

A user first accesses ES by visiting a Web page with a specific URL to a DIMES page (for example, http://spring.scs.gmu.edu:8199/servlet/SiesipDataTree). The user could browse the metadata until he or she finds interesting data sets; remember, scientific data are usually multidimensional, with the most common five dimensions previously discussed (longitude, latitude, vertical level, time, and variable).

DIMES lets a user find data through any metadata dimension, which might match a data dimension directly or indirectly or might not match any at all. For example, data sources, a satellite instrument, or a numerical model from which the data were retrieved are useful metadata information. Because DIMES is flexible, we can also integrate scientific knowledge, such as what natural phenomenon is related to a specific data set, into

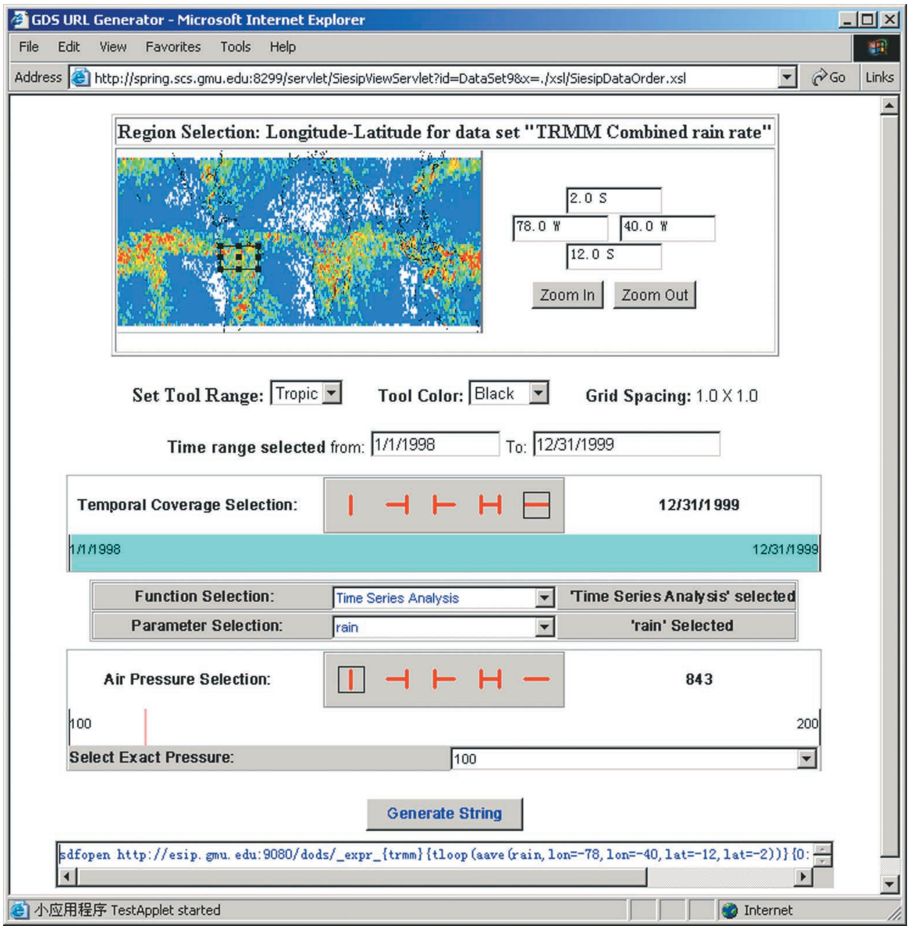

Figure 4. The GrADS–DOD server (GDS) URL generator interface. The interface lets users make area selections (using a rubber band or by typing), and select a time or a time interval, a physical variable, and a function before generating a GDS URL by clicking the "Generate String" button.

DIMES. Suppose in our use case, the user browses the "Phenomenon" folder and then the "DROUGHT" phenomenon. From this browsing point, the user can find the "TRMM Combined rain rate" data set in Figure 1.

Figure 1 gives users basic information about the data set such as spatial–temporal coverage and resolutions. The user is on the data page already, so ES lets him or her order the found data by simply clicking the "Order" button, which leads to the GDS URL generator Web page in Figure 4. This page is preloaded with the metadata information from the search result in DIMES to better guide users to create their data requests. The most prominent preloading information is the browse image from the corresponding data set. The user can view the browse image for an overview of the data as well as use the image as a background to make further spatial selections.

We based the spatial selection portion of the interface on the LiveMap.[11] A user can utilize a "rubber band" tool to select an area of interest based on the background image from the real data. Suppose the user is interested in a high rain rate area. By viewing the rain rate image of January 1998, the user utilizes the rubber band to select the area displayed in Figure 4 (between 12°S – 2°S latitude and 78°W – 40°W longitude), which is in the Amazon jungle.

Similarly, a user can select a time or time period in the binding temporal coverage range. The interface also lets users select a physical parameter in this data set and select a function. Figure 4 shows that the user has selected the temporal coverage from January 1998 to December 1999 and the time series of rain rate. After the user makes all the selections, the user clicks the "Generate String" button in the interface. The GDS URL is then created based on the user's selections.

In Figure 4, we also include the GrADS command for opening data through DODS protocol, although this is not part of the URL. Today, most computers will let users cut and paste the generated URL into their DODS client and handle the data sets opened in this way the same as any other data sets opened locally.

The following GrADS commands generate the time series curve in Figure 5:

```
ga->sdfopen
http://esip.gmu.edu:9080/dods/_expr_
 {trmm{tloop(aave(rain,lon=-78,lon=-
  40,lat=-12,lat=-2))}{0:0,0:0,0:0,
 jan1998:dec1999}
ga-> set display color white
ga-> set t 1 24
ga-> display result
```

The first five lines are a one-line command in the GrADS environment. The GDS URL generator creates the URL; now other DODS-enabled clients can use the same URL to access the same time series. The `display result` GrADS command actually passes the command to the GDS, which processes data on the server side based on the information it receives.

The resultant 24-value time series is returned to the client and is displayed as the result in Figure 5. Keep in mind that the user never directly downloaded or saved any data to the desktop—he or she simply provided a few line commands to get the data and the figure. If the user turns off the desktop, all data will be lost.

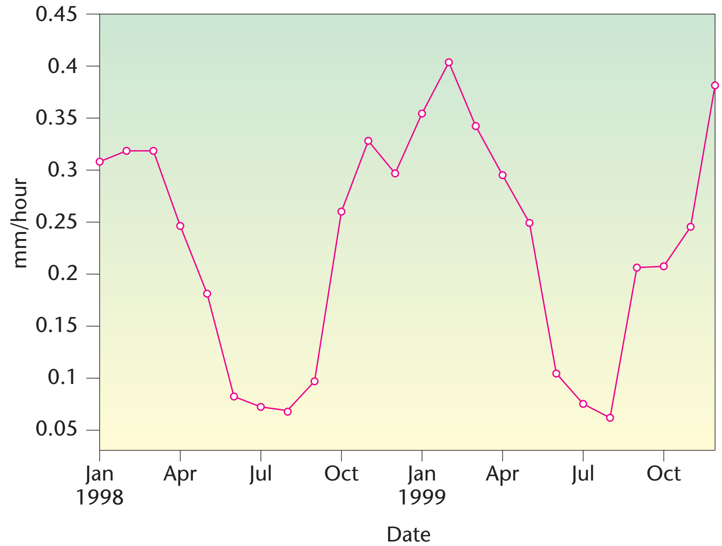

Figure 5. Time series of rainfall rate over Amazonian area. The curve shows obvious seasonal changes in rainfall rate. The key part of our technology is that to create this curve, you need only 24 4-byte values provided seamlessly by our system from the original 2.7 Mbytes data on a remote site.

As the growth of scientific data volume accelerates, distributed data systems for delivering multidimensional and other scientific data to desktops takes on increased importance. We've demonstrated the possibility of such a system that uses existing components to create an integrated prototype. The sample case we described demonstrates a seamless search and use of multidimensional data sets via our ES. Scientists with access to such data systems can save significant amounts of time on the data interoperability problem, especially those who use data sets routinely produced by other centers or individuals.

As you might have noticed, we assume that user agents for DIMES are regular Web browsers and that the data analysis and visualization tools differ. Therefore, we use a cut-and-paste process in data locating, downloading, and displaying, which is not very efficient in some cases. Our ultimate goal is to build an end-to-end, one-stop data information system serving multidimensional science data. With such a system, users can use their familiar data analysis and visualization tools to find, browse, and use distributed data without concerns for data format, location, and type. The ES project is a solid step toward that end.

Another ES improvement we'd like to see is on the server side, to develop the ingest toolbox and integrate it into the ES. Metadata ingestion from a data server to a metadata server may be much more complicated than a simple format conversion because metadata may contain accumulated knowledge that was not with the original data. We have developed several software components for metadata harvesting, cleaning, and merging based on Java and XML technologies. We expect to deliver a full Java-based, platform-independent ES package after finishing development and system integration.

Acknowledgments

We acknowledge partial funding support from the NASA Earth Science Enterprise WP-ESIP CAN program as well as from George Mason University. Other members of the SIESIP team and its advisory board contributed to this work in various ways: Paul Chan, *Chris Lynnes, and George Serafino (NASA Goddard Earth Science Data and Information Services Center, Distributed Active Archive Center); Long Chiu, Xinhua Deng, Zuotao Li, Zhong Liu, James McManus, Yixiang Nie, Paul Schopf, Dong-Bin Shin, Jearanai Vongsaard, Changzhou Wang, Xiaoyang Wang, Yujie Zhao, and Daniel Ziskin (George Mason University); Joe Wielgosz (Center for Ocean-Land-Atmosphere Studies); Cort Willmott and Kenji Matsuura (University of Delaware); and Hank Wolf.*

References

1. G. Asrar and R. Greenstone, eds., *EOS Reference Handbook*, NASA, 1999.
2. M. Kafatos et al., "Information Technology Implementation for a Distributed Data System Serving Earth Scientists: Seasonal to Interannual ESIP," *Proc. 10th Int'l Conf. Scientific and Statistical Database Management*, M. Rafanelli and M. Jarke, eds., IEEE CS Press, 1998, pp. 210–215.
3. R. Yang et al., "An XML-Based Distributed Metadata Server (DIMES) Supporting Earth Science Metadata," *Proc. 13th Int'l Conf. Scientific and Statistical Database Management*, L. Kerschberg and M. Kafatos, eds., IEEE CS Press, 2001, pp. 251–256.
4. R. Yang, M. Kafatos, and X. Wang, "Managing Scientific Metadata Using XML," *IEEE Internet Computing*, vol. 6, no. 4, 2002, pp. 52–59.
5. "NASA Selects Earth Science Information Partners," NASA Press Release, 2 Dec. 1997, www.nasa.gov/releases/1997.
6. R. Goldman et al., "Proximity Search in Databases," *Proc. 24th Very Large Database Conf.*, 1998, www.vldb.org.
7. M. Kafatos et al., "The Virtual Domain Application Data Center: Serving Interdisciplinary Earth Scientists," *Proc. 9th Int'l. Scientific and Statistical Database Management*, IEEE CS Press, 1997, pp. 264–276.
8. H.L. Kyle et al., "Climatology Interdisciplinary Data Collection, Volume 4 for Climate Studies," NASA Goddard DAAC Science Series, *Earth Science Enterprise*, NASA, NP-1998(06)-029-GSFC (CD-ROM).
9. B.E. Doty et al., "The Grid Analysis and Display System (GrADS): An Update for 1997," *13th Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Am. Meteorological Soc., 1997, pp. 356–358.
10. J. Wielgosz et al., "GrADS and DODS," *17th Int'l Conf. on Interactive Information and Processing Systems, AMS,* 2001 (CD-ROM).
11. J. Callahan, "LiveMap Version 3.0," 2000, http://tmap.pmel.noaa.gov/~callahan/JAVA/map_v3.0/LiveMap_v3.0.html.

Ruixin Yang is an assistant professor in the School of Computational Sciences at George Mason University. His research interests include remote-sensing application and information technology for Earth science data. He received his PhD in aerospace engineering from University of Southern California. He is a member of the IEEE and AGU. Contact him at Ctr. for Earth Observing & Space Research, MS 5C3, School of Computational Sciences, George Mason Univ., Fairfax, VA 22030-4444; ryang@gmu.edu.

Menas Kafatos is a professor at George Mason University. He also is dean of the School of Computational Sciences and director of the Center for Earth Observing and Space Research at GMU. His research interests include distributed data systems, interdisciplinary studies in Earth system science (such as coupling of vegetation and climate phenomena), data mining, black-hole theories, and cosmology. He received his PhD from the Massachusetts Institute of Technology. He is a member of the IEEE and AGU. Contact him at the Ctr. for Earth Observing & Space Research, MS 5C3, School of Computational Sciences, George Mason Univ., Fairfax, VA 22030-4444; mkafatos@gmu.edu.

Brian Doty is a research scientist at the Center for Ocean-Land-Atmosphere Studies, Institute of Global Environment and Society. His interests include data analysis and visualization systems for the scientific evaluation of large atmospheric and oceanic data sets. He has a BS from Northern Illinois University. He is a member of the AMS. Contact him at the Ctr. for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Rd., Ste. 302, Calverton, MD 20705-3106; doty@cola.iges.org.

James L. Kinter III is an associate research scientist and executive director at the Center for Ocean-Land-Atmosphere Studies. His research interests include seasonal to interannual climate variability, particularly regarding the interaction between the Indian monsoon and El Niño and the Southern Oscillation. He has a PhD from Princeton University. He is a member of the AMS and AGU. Contact him at the Ctr. for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Rd., Ste. 302, Calverton, MD 20705-3106; kinter@cola.iges.org.

Long Pham is a computer engineer at NASA-Goddard Space Flight Center in Greenbelt, Maryland. He is currently working with different facets of interoperability, including cross-site catalog, data access issues, and scientific data mining. He has a BS in computer engineering from the University of New Mexico, Albuquerque. Contact him at the NASA Goddard Space Flight Ctr., Data Active Archive Ctr., Code 902, Greenbelt, MD 20771; long.b.pham@nasa.gov.