Towards Digital Earth
— Proceedings of the International Symposium on Digital Earth
Science Press 1999

Earth Observing and Distributed Data Information System for Earth Systems Sciences

Menas Kafatos[1] Ruixin Yang[1] David Wong[2] X. Sean Wang[3]

[1] *Center for Earth Observing & Space Research*
George Mason University
Fairfax, VA 22030
USA
E-mail: mkafatos@gmu.edu, ryang@gmu.edu
[2] *Geography & Earth Science*
George Mason University
Fairfax, VA 22030
USA
E-mail: dwong2@gmu.edu
[3] *Information and Software Engineering*
George Mason University
Fairfax, VA 22030
USA
E-mail: xywang@gmu.edu

ABSTRACT George Mason University's Center for Earth Observing and Space Research (CEOSR) has active programs in Earth observing science data and associated data information systems and remote sensing programs. In particular, a consortium led by George Mason University has been funded by NASA's Earth Science Information Partner (ESIP) program to develop, implement, and operate a distributed data and information system. The system supports the data and information needs of seasonal to interannual scientists whose research focus includes phenomena such as El Niño, monsoons and associated climate studies. The system implementation involves several institutions using a multitiered client-server architecture, distributing tasks in the areas of user services, access to data, archiving, and other aspects enabled by a low-cost, scalable information technology implementation. The integrated system is composed of remote sensing data, a database management system (DBMS), communication protocols, data analysis tools, and user interfaces through a web-based Java graphical user interface (GUI). Users can search the DBMS for metadata information, conduct content-based searches, perform some initial analysis and issue an order for the selected data. Additional GMU remote sensing and related information technology research activities, tools and capabilities are described here. Related data cover different spatial scales, from regional to global as well as different spectral types from multispectral to hyperspectral. These are of use and benefit to earth science research and applications, as well for the development and application of Digital Earth.
KEY WORDS remote sensing; data and information resources; data mining; Internet applications

1. Introduction

Remote sensing data and associated information systems at George Mason University (GMU) are supporting research and applications areas, from regional to global coverage, at different spatial and spectral resolutions. In addition to global coverage data collected by NASA and NOAA missions, innovative technologies in gathering data, such as hyperspectral remote sensing imaging and innovative technologies such as data mining, content-based searching and data interoperability among different physical nodes, are being utilized.

Earth system science research developments in the last two decades have made extensive usage of remote sensing and other Earth observation data as well as associated distributed data systems and easy access to those systems. Satellites (Kramer 1996) provide global and repeatable observations of regions of the Earth to address important issues of interest to changes of the Earth's state, the subject of the U.S Global Change Research Program (USGCRP) (The National Science and Technology Council, 1998) and similar international efforts. Science areas of the global change programs include the study of long-term, decadal climate variabilities as well as seasonal-to-interannual variabilities. These require continuous coverage with satellite observations. Owing to the complexity, interdependencies, large, and diverse volumes of data (Kafatos et. al., 1997), Earth system scientists require easy access to data products and distributed information. Data centers (e.g. NASA's Distributed Active Archive Centers, DAACs) are conceived to serve many different user communities. Data usage is, however, focused to particular scientific communities, with their own requirements and information needs. Due to large costs associated with both large space platforms and associated data systems, particularly the Earth Observing System Data Information System (EOSDIS), alternatives have been proposed such as distributed federated systems between diverse data providers (Kafatos et al., 1994).

In contrast to global coverage data sets and associated information systems, applications for a variety of regional projects including environmental monitoring, inventories, forestry applications and transportation needs require data at much higher spatial and spectral resolutions. New technologies such as hyperspectral imaging and synthetic aperture radar (SAR) are becoming available. GMU is in the process of entering into cooperative programs with Virginia area industries in support of user communities which need these data, including other companies, State agencies and applications users. Geographic Information Systems (GIS) tools integrated with remote sensing products, relevant algorithms and laboratory spectral measurements are all required. At GMU we are pursuing development of several of these techniques and tools. Moreover, innovative information technology products and techniques in computer science, statistics and visualization can enhance remote sensing tools and applications.

George Mason University, together with Goddard Space Flight Center DAAC (GDAAC) and the Center for Ocean-Land-Atmosphere Studies (COLA) were awarded an Earth Science Information Partner (ESIP) NASA program (NASA Press release, 1997) designed to focus on seasonal to interannual processes (Deser & Wallace, 1987; CLIVAR Scientific Steering Group, 1995; TOGA Panel, 1996; Kirtman et al. 1997), by providing data and information products to seasonal to interannual (S-I) scientists. The GMU-led ESIP system termed SIESIP (**Seasonal to Interannual Earth Science Information Partner)** forms a member of the federation originally recommended by the US National Research Council in its 1995 U.S. Global Change Research Program study (Board on Sustainable Development, 1995).

We present here the information technology implementation of the SIESIP distributed information system and give examples of the system usage. SIESIP allows integration of information across many domains and in a distributed fashion, which in turn requires sharing of data, information technology, science, and applications, forming a real-life federated system. The integrated system is composed of data, database management system (DBMS), communication protocols, data analysis tools, and an user interface. Through a web-based Java graphical user interface (GUI), users can search the DBMS for metadata information, conduct content-based searches, perform some initial analysis and issue an order for the selected data.

2. A Distributed Data Information System

The S-I research areas include the El Niño/ Southern Oscillation (ENSO) phenomenon; monsoons; large-scale precipitation and wind patterns; the Intertropical Convergence Zone; the Tropical Biennial Oscillation (TBO) as well as associated influences in the tropics and extratropic regions. Over the years, S-I scientists have improved their ability to predict rainfall, sea surface temperature (SST) variations in the tropics and other associated geophysical climate variability. Even though forecasts (ranging up to a year in advance) by S-I scientists are still experimental, they are increasingly being used by planners in the tropics and even in the U.S to mitigate harmful effects, such as floods, droughts and shifting weather.

SIESIP's overall goal is to assist S-I climate scientists with both *data and information* solutions in support of the above science areas. In practical terms, the SIESIP consortium consists of three main distributed sites: George Mason University with expertise in information technology, data searches and analysis and interdisciplinary Earth system science; the Center for Ocean-Land-Atmosphere Studies (COLA) with expertise in S-I science, user services and tools; and the NASA Goddard DAAC with expertise in data management, data archiving and user services. Each consortium member delivers services and products or has developed working prototypes. The overall system is designed to enhance the current consortium capabilities to serve the specific Earth science S-I community and to provide an innovative information technology query engine and implementation of a working federation.

SIESIP promotes ease of use by deploying innovative products and information technology and allowing users to find and obtain data easily. Moreover, SIESIP is designed to assist students by collecting relevant data sets into a single point of access, integrating complementary data sets to enhance information, and producing needed products. As much as possible, easily available analysis tools are applied across diverse data sets, creating ease of use and compatible data interuse. The integration of products and information technology extends from data discovery, search, browse, selection, and access, to elementary analysis making these aspects easier for users. Specifically, SIESIP is developing selection tools (e.g., content based search, data mining) to make data selection and access easy and to streamline data use, by achieving interoperability for the user activity that counts most--**data interoperability.** Unlike large, conventional information systems, which often adopt a system orientation, SIESIP adopts a product and services orientation and focuses on the information inherent in the data.

A fundamental component of SIESIP enabling S-I research is the close integration of data products and services with analysis software tools. The analysis engine at the heart of the SIESIP engine is the *Grid Analysis and Display System* (GrADS) (Doty & Kinter, 1995). GrADS can be used on the user's own platform to carry out extensive scientific analysis; or it can be used in a client-server mode to perform on-line analysis to obtain information about relevant data sets of use to the user.

This SIESIP consortium is building on the existing GMU *Virtual Domain Application Data Center* (VDADC) search engine (Kafatos et al., 1997; Yang, Li & Kafatos, 1998) and current user services and data at COLA and the DAAC. Innovative features include a new GUI, on-line analysis capabilities and data access and query scenarios. The SIESIP consortium's information technology implementation is innovative and scalable, makes substantial use of new technology, and relies on WWW with a multi-tiered client-server system architecture that starts with easy access for users.

2.1. Siesip Data

The main goal of the distributed data system is to enable S-I climate researchers with both data and information solutions for their research. SIESIP achieves this by creating new products based on existing and new satellite and station data and models (http://www.siesip.gmu.edu/data.html). These products will serve the broad user base involved in S-I research, *Tropical Rainfall Measuring Mission* (TRMM) scientists, experiment users such as the South China Sea Monsoon Experiment (SCSMEX), as well as application users such as agriculture.

To design the data access and query implementations, data access and usage scenarios have to be constructed. We have examples of several such scenarios (see current VDADC prototype, http://www.ceosr.gmu.edu/~vdadcp). For example, SIESIP data sets are to enable users to explore phenomena such as teleconnections between El Niños and vegetation cover in Africa, by plotting time series (e.g., Eastern Pacific SST anomalies and Sahelian precipitation) and correlations of relevant parameters (e.g., Eastern Pacific SST anomalies and seasonal area extent of Sahara).

SIESIP data sets include new gridded TRMM data products; event-driven subsets (such as hurricane events, see http://esip.gmu.edu/siesip/orbit3D/html/orbit3D.html); TRMM high-resolution rainfall data over land; TRMM coincidence subsets; data in support of the SCSMEX project; new five-day mean interdisciplinary climate data; and diverse NOAA data sets prepared at a single site (COLA). We have several data sets available via the Web and introducing new ones all the time. Moreover, one particular useful data collection is the Climatology Interdisciplinary Data Collection in 4 free CD's (Kyle et al., 1998) which can be ordered (see also http://siesip.gmu.edu/data.html).

2.2. Three-Phase Access Model

SIESIP provides a search and analysis engine that will allow users to obtain data whether they do or do not know exactly what to retrieve and let them identify, among available data, significant correlations, trends worthy of further analysis and assess the data to be retrieved. Moreover, this support will allow additional communities such as process scientists and applications users to access the SIESIP data and can, therefore, provide scalable usage lessons to NASA and information technology communities. Based on this assumption, we developed a **three-phase** user data search model for SIESIP.

Phase 1: Using the metadata and browse images provided by the SIESIP system, the user browses the data holdings. Organizing knowledge, such as phenomenon related parameters and spatial regions, is incorporated in the system (producing information-rich products). We distinguish different types of metadata, *description metadata* (which describe the available data sets, their location, parameters, etc.); *content-based and statistical metadata* (which describe the content of the data for data mining and quick browsing); and *text annotations* (which can be supplied by the system or added by users).

Phase 2: The user gets a quick estimate of the type and quality of data found in phase 1. Analytical tools such as GrADS are applied with a variety of on-line capabilities including statistical functions and visualization algorithms. The SIESIP interface will also incorporate a spectrum of statistical data mining algorithms. We have begun to implement tools for finding positive correlations providing realistic, human-aided data mining capability. We have applied this and other data mining systems (Li, Kafatos & Michalski, 1997) to ENSO teleconnections with possible results in identifying anticorrelations with vegetation in tropical Africa and in the NE coastal U.S. (Li & Kafatos, 1999).

Phase 3: The user has located the data sets of interest and is ready to order. If the data are available through SIESIP, the system will handle the data order; otherwise, an order will be issued to the appropriate data provider on behalf of the user, or necessary information will be forwarded to the user for this task.

2.3. Siesip System Architecture

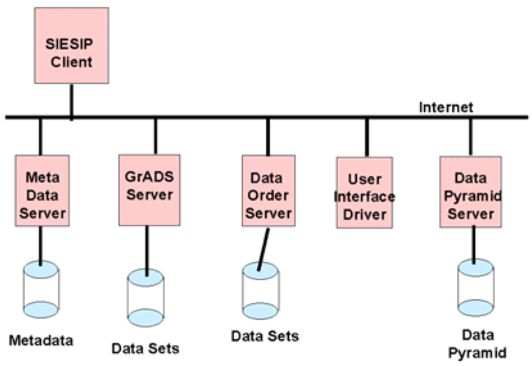

Figure 1

The architecture design (Kafatos et al., 1998) of the SIESIP mini-federation is to support the queries in all three phases in a modular fashion. Specifically, there exist three types of servers, each serving queries in one phase. Figure 1 is the architecture diagram. A "Metadata Server" is for phase 1 (or metadata) queries, a "GrADS Server" is for phase 2 (or analysis) queries, and a "Data Order Server" is for phase 3 queries (or data set requests). There is also an experimental "Data Pyramid Server" which will be responsible for data Figure 1.

mining queries. The Data Pyramid stores low-resolution data as well as some precomputed statistics for fast processing of data mining (content-based) queries (Li et al., 1998).

The system supports two-tiered as well as three-tiered client-server architectures. In a two-tiered framework, the SIESIP Client will talk directly with the servers. In a three-tiered framework, a "User Interface Driver" module is used in between the SIESIP client and the servers. Such a three-tiered framework is more suitable to build a "thin" SIESIP client.

In most cases, a server may be able to answer a query issued to it, using the data located in the disks associated with the server. However, the SIESIP architecture is designed to support queries that may need data that are located in different physical disks. In this case, two or more servers may be needed to answer a single query, where the server that the query is issued to behaves as the "master node" for the query. The master node will request data from all relevant servers, or dispatch subqueries to these servers. An implementation of the SIESIP architecture can have many servers, and user interface drivers, located at different physical locations, while a physical system may host a number of different servers and interface drivers.

The current workable SIESIP distributed data information system consists of components of data products, a database management system, communication protocols, data analysis tools, and user interface modules.

Holding of data products is one of the major components of the system. Presently, the system holds monthly mean data of Sea Surface Temperature (SST) and Normalized Difference Vegetation Index (NDVI) over global scales and air temperature and precipitation covering South America (Udel Climate Station Data Product, 1998). The raw data files in the system are used for analysis and ordering. The corresponding statistical summary data and holding catalogue are inserted into the DBMS for supporting the content-base queries. In addition to local data holdings, SIESIP can also access data sets on remote date sites through different data transfer protocols. Right now, SIESIP integrates predefined data sets through ftp protocol and *Distributed Oceanographic Data System* (DODS) protocol (Gallagher & Milkowski, 1995).

A database management system is used for the system to handle catalogue metadata and statistical summary data. Two major kinds of queries will be supported by the database system. One is used to find right data files for analysis and ordering based on catalogue metadata only. The other is queries on the data contents which are supported by the statistical summary data. For example, finding SST for given spatial and temporal ranges is a query belonging to the first group. However, a search for regions and time periods over which the values of mean SST are in a given value range is a query of the latter group.

To analyze data on-line, analysis and visualization tools must be incorporated into the system. A free (public domain) software, GrADS, developed by one of our consortium members, COLA, is chosen for the system. The usage of the tool is fully automatic and transparent. Users actually do not know and do not need to know what tool the system is using other than the analysis functions used. With the current modular design, other tools such as IDL can be easily integrated into the system.

The user interface is possibly the most important component of the system since users usually evaluate a system based on it. The main web page of current SIESIP contains a GUI applet based on Java Swing technology. A snapshot of the GUI is given in Figure 2. There are three menu lists for handling the three phase queries, respectively. The workplace menu list gives users control on choosing working data product objects for further analysis and other functions.

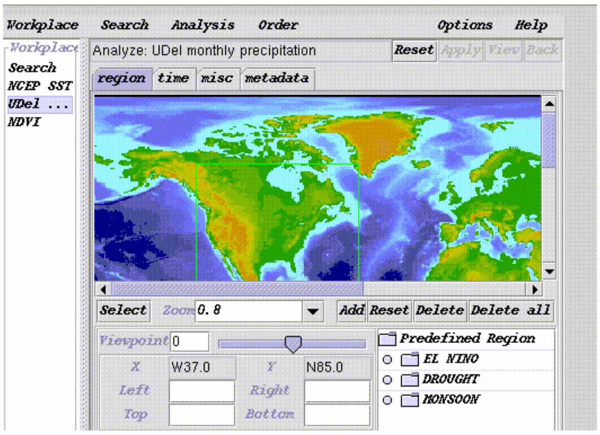

Figure 2

2.4. Siesip Functionality

The SIESIP supports many functionalities for on-line data search, analysis and order. Search can be performed based on regular metadata or based on data contents. Another major feature of the SIESIP system is analysis utilizing the World Wide Web. There are many analysis functionalities which are supported. Some of the analyses were prototyped in the VDADC (Kafatos et. al., 1997), a precursor of the SIESIP.

The most simple and straightforward functionality is creating browse images for given space, time and parameters. Since SIESIP supports data pyramids, a specific resolution could be selected for browsing. With the multiple resolution data, users can start from low resolution images and drill down to higher resolution ones. In this way, users can browse the data covering large spatial and temporal ranges and then focus on particular, small interesting ranges based on previous browsing results. Because users use data of different resolution each time, as the users drill deeper and deeper, the data volume does not increase rapidly. Therefore, what we call **content-based browsing** (or "manual data-mining") on large volumes of data can be performed at reasonable speeds. SIESIP also supports some time series analysis such as temporal correlation of spatially average geophysical values.

3. Example of The Distributed Data System Usage

Since SIESIP is a web based system, the only requirement to use it is a web browser supporting Java and, of course, a network connection. The three phase queries are orthogonal to each other. Therefore, users can go to any queries directly. One procedure to use the system is described here.

A user starts the system by visiting the SIESIP URL. Instead of using the default settings, the user may use the "Workplace" menu list to find a specific parameter he/she is interested in. Then he/she can use "region" and "time" panels to restrict spatial and temporal coverage. Finally, the user can use 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 (an URL) to the client which will automatically pull another web page displaying the result as shown in Figure 3. Based on the browse results, the user can then proceed with other on-line analysis functions or end this phase and proceed to order the data.

It is worth noticing that the three images on figure 3 were created dynamically based on data from three different sites. The first image is based on CMAP precipitation data. The data were obtained from the COLA site through the ftp protocol at the user's request. The COADS climatology SST data for the second image were transferred to the GMU SIESIP site through the DODS protocol. The third one is based on local data set at GMU. All the data site information is transparent to SIESIP users. Users do not know the location of the data sites before they issue the request. For demo purpose, SIESIP provides a monitoring window to notify users what is going on in the system for fulfilling users' queries. Only through this window would users notice the system fetching data from different sites to answer their queries.

Figure 4 shows another functionality of SIESIP on-line analysis. The figure is scatter plots among three time series data of geophysical parameters over certain areas. In the SIESIP result interface, users could also see a list of parameters with temporal ranges and spatial coverage information. Furthermore, values of the associated correlation coefficients are also displayed on the same web page.

Independent of the Java applet interface, SIESIP also supports a CGI based dynamical analysis tool called *Orbit Viewer* (Kelley, Kwiatkowski & Kafatos, 1998). *Orbit Viewer* was developed by GMU technical staff at the TRMM Science Data and

Information System (TSDIS) based on IDL. The tool can read standard HDF files and output images in GIF and postscript formats. The on-line analysis system on SIESIP is a

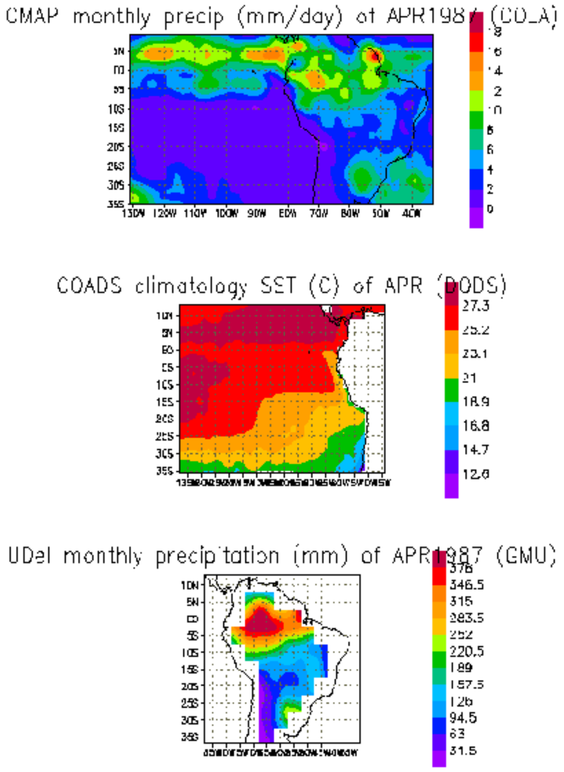

Figure 3

tool for analyzing hurricane or typhoon events based on TRMM data. User can access the system through the form-based interface and input his/her selections on parameters. The system creates an image based on user's choice. Figure 5 is an example of the image created by this system. The image shows the rainfall surface of on April 29, 1999 for Tropical Cyclone Leo. Interested users can use this system by visiting http://esip.gmu.edu/siesip/orbit3D/html/orbit3D.html.

4. Extensions of Siesip

Currently, SIESIP is used to assemble Earth observing data physically residing at different locations and in heterogeneous formats. The system does include some limited capabilities for exploratory type of analysis. Clearly, after browsing the data on the web, S-I scientists can decide to download the data for further analysis at local site. However, SIESIP can be extended to certain domains which are rather specific at GMU.

4.1. Geographic Information Systems (GIS)

Earth observing data are of interest to not only S-I scientists, but also of broad-base interests to many scientists conducting research related to the

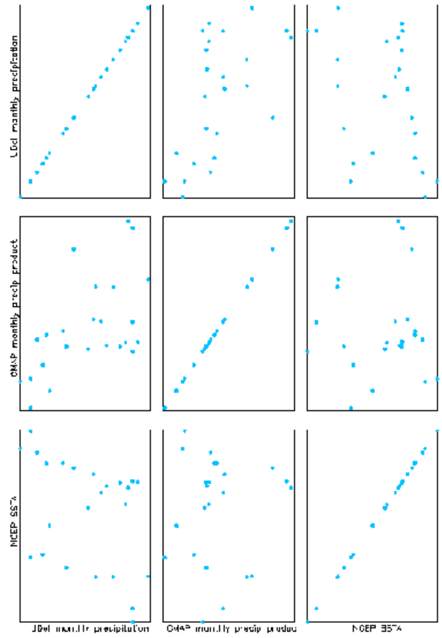

Figure 4

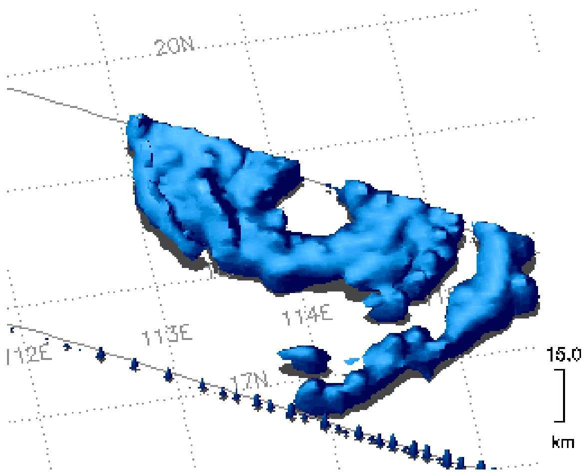

Figure 5

environment. Environmental scientists in general, but more specifically ecologists, climatologists, oceanographers, and even emergence response personnel (such as the Federal Emergence Management Agency – FEMA) are interested in data and system such as SIESIP. These scientists and practitioners, however, include many GIS users. They study various types of environmental phenomena and of various levels of geographic scale (global, regional, and local). Space observing or remote sensing data can be combined with other types of spatial data, such as population or Earth science data, to provide more comprehensive analyses. To enhance SIESIP along this direction, web-based GIS functions can easily be incorporated into the system to allow users to conduct more sophisticated and more GIS-analysis in real time on the web. On the other hand, after browsing the data and completing preliminary analysis on the web, GIS users can elect to download the data for local usage. But most space observing data are stored in particular formats that cannot be directly imported to GIS. Conversion procedures to disseminate remote sensing in popular GIS formats can be built into SIESIP to broaden the usage of the system to GIS users. The work in converting TRMM data from HDF format into GIS formats was an experiment along this direction (Wong, 1999).

4.2. Scientific Visualization, Statistical Graphics, and Virtual Reality

As shown in previous section, one major function of SIESIP is to visualize remote sensing data but very much in a screening mode. Most scientific data, however, are massive in size and effective visualization of such large databases is critical to scientific inquiry. Statisticians and computer scientists at GMU have made tremendous progress at this front and the development of SIESIP can potentially benefit from these advancements in visualization. For instance, Carr (1995) suggested new methods to visualize gigantic databases. New statistical graphics techniques have been introduced. Several powerful statistical graphics techniques for visualization of spatial data, including the micromaps (Carr and Pierson, 1996), have been developed. Some of these tools have already been implemented in the web environment to be used by any web users (Symanzik, et al 1998, 1999). In addition, new models to represent and visualize global data effectively have been suggested (Carr et al 1997). Virtual reality technology has also been employed to explore and analyze high-dimension data (Wegman et al 1999). With our experience in these areas of scientific visualization, SIESIP can possibly be extended to incorporate at least some of these components.

4.3. Data Mining, Machine Learning and Discovery

Data mining activities are taking place at CEOSR (see above) as part of its remote sensing and Earth science data activities; data mining activities are also taking place at the Center for Computational Statistics, utilizing statistical mining techniques involving, among others, multi-dimensional visualization tools, statistical exploratory data analysis such as the "grand tours" (Wegman and Carr, 1992); and at the Machine Learning and Inference Laboratory.

A variety of inductive learning tools exist at GMU's Machine Learning and Inference Laboratory. To support scientific analyses, data mining tools such as AQ15c should possess the following properties: The discovered pattern (or the form of knowledge representation) should be relatively easy to interpret by humans, and; the data mining system should be able to benefit from the domain knowledge (which is easy for scientific work since the domain knowledge is well-documented). Inductive learning systems that generate decision rules (Michalski et al. 1986; Wnek et al., 1995) have been utilized for scientific queries related to El Nino events (Li, Kafatos and Michalski, 1997). Rules are represented in the VL1 (Variable-valued Logic system 1) notation. When building a decision rule, AQ15c performs a heuristic search through the space of logical expressions to determine those that account for all positive examples. Such tools are also able to handle inconsistent data and noisy data. Interesting hypotheses can be generated by applying inductive machine learning tools which can then be tested by the scientists.

4.4. Remote Sensing for Regional Applications

Hyperspectral remote sensing is an emerging technology for accurate remote sensing of the Earth for a variety of applications. Typical future satellite missions such as ORBIMAGE's *OrbView-4* will provide more than 200 spectral channels with four-meter resolution imagery. As such, emerging capabilities in satellite technology, remote sensing technology and large database management afford the remote sensing industry and many stakeholders including agricultural concerns, environmental agencies, policy makers, oil industries as well as State of Virginia agencies, and the federal government with opportunities that did not exist even a few years ago. Efforts at GMU are being carried out to work with industry which will be providing users and stakeholders with remote sensing products that do not currently exist for a variety of applications, with new and unexplored information content, with image clarity, increased accuracy, ease of use, timeliness and at a competitive cost.

Hyperspectral imaging instruments, mounted on low Earth-orbit satellites, and/or coupled with similar instruments aboard aircraft, can provide data and information products with sensing information to locate mineral and petroleum deposits. Other products can assess the productivity of large bodies of water such as the Chesapeake Bay, monitor soil before planting and analyze and monitor crops as they grow, assess road traffic flow, map routes for roads and pipelines, and help educate the next generation of research and applied scientists through modeling, simulations and algorithm developments.

Dissemination of hyperspectral imagery and products for a variety of usages is an integral aspect of these remote sensing activities. These products can be marketed to customers including companies involved in, among many areas including petroleum and mineral exploration, environmental assessment and remediation, agriculture and agribusiness. It is expected that State of Virginia and federal agencies will also benefit and make usage of these products.

5. Future Developments

Data interoperability is achieved by SIESIP through ftp and DODS protocols. Metadata interoperability is planed for future developments. The usual way to manage metadata is to use database management system (DBMS). However, for most small data providers and Earth scientists, DBMS is not their favorite tool since they consider DBMS systems to be complex and are not suitable for scientific data and metadata. Another major difficulty of metadata is their diversity. Even with commercial DBMS, different groups will result in different database designs which are difficult to access dynamically as a distributed system.

In CEOSR, we plan to use Extensible Markup Language (XML) to handle metadata. With XML it is easy to handle diversity of metadata by providing a common and easy to use framework. Most scientists who would not use DBMS may learn to use XML very easily. Full text search is much easier to implement with XML than with DBMS. Since XML is well structured document and web browsers support (or will support) XML display, it is possible to allow SIESIP users to browse metadata in XML and to form their own queries instead of using the predefined queries. Moreover, XML is very flexible in structure. Therefore, it is possible to store users' input, such as comments on a data set, into the system metadata. The new information will be used as part of metadata to support users' queries. Most importantly, XML was developed for Internet applications and the features in XML provide an easy way to achieve on-line diverse metadata interoperability.

Another plan in CEOSR is to extend our efforts on content-based search. The content-based search is a specific type of data mining. The content-based queries cannot be answered by metadata only, data values must be involved. In SIESIP, content-based search is supported by statistical summary data of parameters (Li et al. 1998). For example, minimum, maximum and mean values of Sea Surface Temperature (SST) over regions covering global oceans are stored inside the system. A query for searching regions over which mean SST values is in a given value range can easily be answered. Many more complex queries such as queries with logical operators and involving more than one parameters can also be fulfilled by the system.

In collaboration with the departments of Geography & Earth Science and Biology, CEOSR has proposed a feasibility study to the State of Virginia leading to the establishment of a center of excellence for hyperspectral remote sensing in Northern Virginia. The center will be involved in applied and fundamental research in remote sensing for regional applications. The center will provide an environment for remote sensing research to thrive, concentrating initially on a small number of pilot projects carried out by two of Virginia's premier institutions, George Mason University and Virginia Tech. Furthermore, the hyperspectral and remote sensing research and applications at the center will be coupled to other center activities. Innovative products such as combining hyperspectral imagery with synthetic aperture radar (SAR) will be developed for greater information content.

The center's overall research activities will, among other things, emphasize:

- the detection, tracking, monitoring, assessing & leveraging spectral signatures
- the development of appropriate hyperspectral remote sensing algorithms
- the correlation of contaminants, tracers and spectral signatures

Initial research support areas selected to be studied, include: Chesapeake Bay and coastal wetlands, (e.g. water quality analysis, phytoplankton abundance and wetland inventory); and forestry (e.g. foliar nitrogen characterization, health of trees, water stress, etc).

Finally, remote sensing can provide support for other regional applications such as mineral/petroleum exploration; environmental monitoring and assessment; fresh water and watershed runoffs; agriculture and other renewable resources; environmental and coastal mapping, etc.

The GMU projects and capabilities described in the present work provide numerous opportunities to contribute to the realization of the Digital Earth. Our view is that the Digital Earth will require diverse remote sensing and other data at different spatial & spectral resolutions, distributed data systems linked via federated solutions and innovative information technology tools. Digital Earth projects can also provide opportunities for interdisciplinary education and training of future generations of Earth scientists and applications users.

Acknowledgement: We acknowledge partial funding support from the Earth Science Enterprise ESIP CAN Program as well as from George Mason University. Other members of the SIESIP team include C. Lynnes & G. Serafino (GDAAC); B. Doty, J. Kinter & C. Steinmatz (COLA); P. Chalermwat, L. Chiu, X. Deng, J. McManus, P. Schopf, J. Shukla, J. Vongsaard, H. Weir & K-S Yang (GMU); C. Willmott & K. Matsuura (Udel); and H. Wolf (IRMA).

References

Board on Sustainable Development, 1995. *A Review of the U.S. Global Change Research Program and NASA's Mission to Planet Ear/Earth Observing System*. National Academy Press, 96pp.

Carr, D. B. 1995. "Perspective on the Analysis of Massive Data Sets". Computing Science and Statistics, Proceeding of the 27th Symposium on the Interface. Vol. 27. p410-419.

Carr, D. B. and S. Pierson. 1996. Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps. Statistical Computing & Graphics Newsletter, Vol. 7 No. 3 pp. 16-23.

Carr, D. B., R. Kahn, K. Sahr, and A. R. Olsen. 1997. "ISEA Discrete Global Grids," Statistical Computing & Graphics Newsletter, Vol. 8 No. 2/3 pp. 31-39.

CLIVAR Scientific Steering Group, 1995. *CLIVAR: A Study of Climate Variability and Predictability: Science Plan*. World Climate Research Programme, WCRP-89, (WMO/TD No. 690), 157 pp.

Deser, C. and Wallace, J.M., 1987. El Niño Events and Their Relation to the Southern Oscillation: 1925-1986, *J. Geophys. Res.*, 92:14189-14196.

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

Gallagher, J. and G. Milkowski, (1995), Data Transport within the Distributed Oceanographic Data System, World Wide Web Journal: Fourth International World Wide Web Conference Proceedings, pp691--702.

Kafatos, M., et al., 1994 *The GMU ECS Federated Client-Server Architecture*, report to NASA/Hughes.

Kafatos, M., Li, Z., Yang, R. et. al., 1997. The Virtual Domain Application Data Center: Serving Interdisciplinary Earth Scientists, *Proceedings of the Ninth International Scientific and Statistical Database*, 264-276. IEEE.

Kafatos, M., Wang, X., Li, Z., Yang, R., Ziskin, D., 1998. "Information Technology Implementation for a Distributed Data System Serving Earth Scientists: Seasonal to Interannual ESIP," *Proceedings of the Tenth International Conference on Scientific and Statistical Database Management*, 210-215. IEEE.

Kelley, O.A., Kwiatkowski, J.M., Kafatos, M., 1998, "A Tool for Dynamically Generating Images for TRMM Validation," Poster Paper, AGU spring meeting, Boston, May, 1998.

Kirtman, B.P., Huang, B., Shukla, J., Zhu, Z., 1997. Tropical Pacific SST Prediction with a Coupled GCM, *Experimental Long Lead Forecast Bulletin*, Vol. 6, No. 1, 14-15.

Kramer, H.J., 1996. *Observation of the Earth and Its Environment: Survey of Mission and Sensors.* Springer, Germany

Kyle, H.L., McManus, J.M., Ahmad S. et. al., 1998. "Climatology Interdisciplinary Data Collection, Volumes 1-4, Monthly Means for Climate Studies," NASA Goddard DAAC Science Series,

Earth Science Enterprise, National Aeronautics & Space Administration, NP-1998(06)-029-GSFC.

Li, Z., Wang, X..S., Kafatos, M., and Yang, R., 1998. A Pyramid Data Model for SupportingContent-based Browsing and Knowledge Discovery, *Proceedings of the Tenth International Scientific and Statistical Database*, 170--179. IEEE.

Li, Z., Kafatos, and M., Michalski, R. 1997. Data Mining Application for El Niño Teleconnection Research, GMU for Machine Learning Laboratory Report.

Li, Z., and Kafatos, M. 1999, Interannual Variability of Vegetation in the US and its Relation to ENSO, *Remote Sensing of the Environment*, (in press).

Mikalski, R.S., Mozetic, I., Hong, J., and Lavrac, N., 1986. The AQ15 Inductive Learning System: An overview and Experiments, ISG86-20, UIUCDCS-R-86-1260, Department of Computer Science, Univ. of Illinois, Urbana.

NASA Press release, Dec. 2, 1997. NASA Selects Earth Science Information Partners,http://www.nasa..gov/releases/1997/.

The National Science and Technology Council, 1998. A report by the Subcommittee on Global Change Research,: *Our Changing Planet: The FY 1998 U.S. Global Change Research Program.* Office of Science and Technology Policy.

TOGA Panel, 1996. *Learning to Predict Climate Variations Associated with El Niño and the Southern Oscillation: Accomplishments and Legacies of the TOGA Program*, National Academy Press, 171 pp.(Washington)

Symanzik, J., Wong, D., Wang, J., Carr, D. B., Axelrad, D.

A., Woodruff, T. J. (1999): Web-based Access and Visualization of Hazardous Air Pollutants, Proceedings GIS in Public Health 1998.

Symanzik, J., Axelrad, D. A., Carr, D. B., Wang, J., Wong, D., Woodruff, T. J. (1999): HAPs, Micromaps and GPL - Visualization of Geographically Referenced Statistical Summaries on the World Wide Web, © 1999 American Congress on Surveying and Mapping, Annual Procee-dings (ACSM-WFPS-PLSO-LSAW 1999 Conference CD).

Udel Climate Station data products, see URL: www.siesip.gmu.edu/presentations/climfields.html.

Wegman, E.J., and Carr, D.B., 1992. Statistical Graphics and Visualization, Center for Computational Statistics, George Mason University.

Wegman, E. J., Symanzik, J., Vandersluis, J. P., Luo, Q., Camelli, F., Dzubay, A., Fu, X., Khumbah, N.-A., Moustafa, R. E. A., Wall, R. L., Zhu, Y. (1999): The MiniCAVE - A Voice-Controlled IPT Environment, In: Bullinger, H.-J., Riedel, O. (Eds.) 3. International Immersive Projection Technology Workshop, 10./11. May 1999, Center of the Fraunhofer Society Stuttgart IZS, Springer, Heidelberg, 179-190.

Wnek, J., Kaufman, K., Bloedorn, E., and Michalski, R.S., 1995. Inductive Learning System AQ15c: The Method and User's Guild, Reports of the Machine Learning and Inference Laboratory, MLI 95-4, George Mason University, Fairfax, VA.

Wong, D.W.S., 1999, Developing a Prototype Procedure to Disseminate Remotely Sensed Data from NASA Recent Missions in GIS Formats. Technical Report submitted to NASA/GSFC Digital Active Archive Center (DAAC).

Yang, R., Li, Z. & Kafatos, M., 1998, Remote Sensing Data Access via WWW-enabled Search and Analysis Engines, Proceedings of The 4th Pacific Ocean Remote Sensing Conference, Qingdao, China, July 28-31, 1998.