S23-82

p. 11

CIA9 October, 1993

N95-34241

DATAHUB: SCIENCE DATA MANAGEMENT IN SUPPORT OF INTERACTIVE EXPLORATORY ANALYSIS

Thomas H. Handley, Jr. *
Mark R. Rubin**
Jet Propulsion Laboratory
Pasadena, California 91109

Abstract

The DataHub addresses four areas of significant need: scientific visualization and analysis; science data management; interactions in a distributed, heterogeneous environment; and knowledge-based assistance for these functions. The fundamental innovation embedded within the DataHub is the integration of three technologies, viz. knowledge-based expert systems, science visualization, and science data management. This integration is based on a concept called the DataHub. With the DataHub concept, science investigators are able to apply a more complete solution to all nodes of a distributed system. Both computational nodes and interactive nodes are able to effectively and efficiently use the data services (access, retrieval, update, etc.) in a distributed, interdisciplinary information system in a uniform and standard way. This allows the science investigators to concentrate on their scientific endeavors, rather than to involve themselves in the intricate technical details of the systems and tools required to accomplish their work. Thus, science investigators need not be programmers. The emphasis is on the definition and prototyping of system elements with sufficient detail to enable data analysis and interpretation leading to information. The DataHub includes all the required end-to-end components and interfaces to demonstrate the complete concept.

Setting the Stage - The Issues

It is difficult, if not impossible, to apply existing tools for visualization and analysis to archived science instrument data [2]. This difficulty is generally the result of (1) incompatible data formats and the lack of available data filters; (2) the lack of true integration between the visualization and analysis tools and the data archive system(s); (3) incompatible and/or non-existent metadata; and (4) the exposure of the scientist to the complexities of networking. These problems will be multiplied by the avalanche of data from future NASA missions [8, 32]. New modes of research and new tools are required to handle the massive amount of diverse data that are to be stored, organized, accessed, distributed, visualized, and analyzed in this decade [4, 26].

The areas of most immediate need are: (1) science data management; (2) scientific visualization and analysis; (3) interactions in a distributed, heterogeneous environment; and (4) knowledge-based assistance for these functions. The fundamental innovation required is the integration of three automation technologies: viz. knowledge-based expert systems, science visualization, and science data management. This integration is based on a concept called the DataHub.

With the DataHub, investigators are able to apply a complete solution to all nodes of a distributed system. Both computational nodes and interactive nodes are able to effectively and efficiently use the data services (access, retrieval, update, etc.) in a distributed, inter-disciplinary information system in a uniform and standard way. This enables the investigators to concentrate on their scientific endeavors, rather than to involve themselves in the intricate technical details of the systems and tools required to accomplish their work; thus, investigators need not be programmers.

DataHub addresses data-driven analysis, data transformations among formats, data semantics preservation and derivation, and capture of analysis-related knowledge about the data. Expert systems will provide intelligent assistant system(s) with some knowledge of data management and analysis built

---

* Technical Group Supervisor
Member of AIAA

** Member of Technical Staff

in. Eventually DataHub will incorporate mature expert system technology to aid exploratory data analysis, i.e., neural nets or classification systems. Additionally, as a long term goal, DataHub will be capable of capturing and encoding of knowledge about the data and their associated processes. The DataHub provides data management services to exploratory data analysis applications, i.e., LinkWinds [23], PolyPaint+ [15], exploratory data analysis environments.

In developing DataHub we utilize the problems as posed by the science co-investigators to aid in directing capability and development decisions. DataHub's general problem-solving structure will be applied in the general science problems, as described by the science co-investigators.

Goals and Objectives

Our goal is to integrate the results from science data management, visualization, and knowledge-based assistants into a scientific environment; to demonstrate this environment using real-world NASA scientific problems; and to transfer the results to science investigators in the appropriate disciplines.

The specific objectives of the DataHub work are to:

1. Define and develop an integrated system that is responsive to the science co-investigator's needs.
2. Demonstrate the interim capabilities to the participating science users of the system in order to receive their suggestions.
3. Transfer the results of this effort to a broad base of science investigators as appropriate.
4. Provide a system that will enable the science investigator to obtain publishable scientific information.

Emerging Relationships

As illustrated in Figure 1, LinkWinds is providing two functions: (1) a visual data exploration or analysis environment; and (2) visual browsing and subsetting services. In the first function, LinkWinds will be notified via a message of the presence of data. The existence of this data will be incorporated into the LinkWinds database menu and, hence, be made available to the user immediately. The second function will be used when it is more convenient to graphically select the subsetting attributes. After selection of the attributes, a message will be sent to DataHub, the filtering accomplished, and the results re-submitted to LinkWinds for analysis.

A new link is being established with PolyPaint+. PolyPaint+ will provide a interactive visualization of complex data structures within three-dimensional data fields, in addition to visual subsetting services. Interactions with PolyPaint+ will require DataHub to expand its understanding of formats and data, and to provide different filtering capabilities.

The application of machine learning techniques to feature recognition in datasets of interest at JPL. The specific problem is to detect and categorize small volcanoes on Venus using the Magellan SAR data. The techniques is user interaction for feature selection and machine learning will be directly applied to the pre-processing tools used in the DataHub environment.

The Navigation Ancillary Information Facility provides a capability called SPICE (Spacecraft, Planet, Instrument, C-matrix, and Events)[19]. SPICE contains all the ancillary data associated with a mission. The data along with an extensive library are available concerning an expanding set of missions. The SPICE capability, initially developed to support science analysis, is now available as a toolkit. It is our intention to investigate the use of the SPICE toolkit in association with other applications to provide needed ancillary data and processing.

Approach

We have analyzed the management of distributed data across different computing and display resources. Subsequent to this analysis and design, we implemented the specific components required to provide needed science functions. Several prototypes have been provided to illustrate the capabilities. Additionally, we have attempted to apply knowledge-based expert system and machine learning technologies to provide "assistants" for the science investigator in data discovery and selection, tools selection and science processing. Today's solution, DataHub, takes the first steps toward the integrated solution needed to provide the means to satisfy the technology and science requirements in the 1990s by providing a high performance, interactive science workstation with the capabilities to handle both exploratory data analysis and science data management.

The Basic Concept

Figure 1 depicts the current functional architecture for the DataHub. The major functions of the DataHub include providing (1) an interactive user interface; (2) a command-based query interface; (3) a set of data manipulation methods; (4) a metadata manager; and (5) an underlying science data model. The interactive user interface, basic data operators and a data interchange interface with LinkWinds have been implemented in the initial prototypes.

The command-based query interface, such as with LinkWinds illustrated by the double-headed arrow in Figure 1, is designed for the data visualization system to issue data management commands to the DataHub. The data manipulation methods provide the selection, subsetting, conversion, transformation, and updates for science data. The metadata manager captures the necessary knowledge about science data. Finally, the science data model supports the underlying object-oriented representation and access methods.

Figure 2 depicts the current software architecture. A layered architecture has been adopted for the implementation, which implies that any layer can be changed and/or replaced without affecting other layers. The top layer is the external interface that links to the human users via an interactive interface provided by DataHub or the visualization system via a connection interface. The data model is implemented in the intelligent data management layer. The data interface layer provides the physical data access functions.

Current Capabilities

DataHub Version 0.5 has been implemented and tested in the Sun SPARCstation and the Silicon Graphics environments. The implementation uses the software structures illustrated in Figure 2.

From a user's stand point, DataHub recognizes/understands several common datasets either by name or format, plus several other popular formats. The datasets include MCSST, CZCS, Voyager, Magellan, AVIRIS, Viking, and AirSar; the formats include VICAR [17], DSP, HDF [24], netCDF [20, 27], and CDF [3]. Present preprocessing capabilities are data filters, e.g., temporal or band selections, subsampling and averaging options, and spatial subsetting. With the data link with LinkWinds, the user may select and process a dataset of interest then proceed to the LinkWinds environment for exploratory data analysis.

The current DataHub user interface and a typical user session including interactions with LinkWinds are illustrated in Figures 3 and 4 respectively. A description of the interface design update and development my be found in [12].

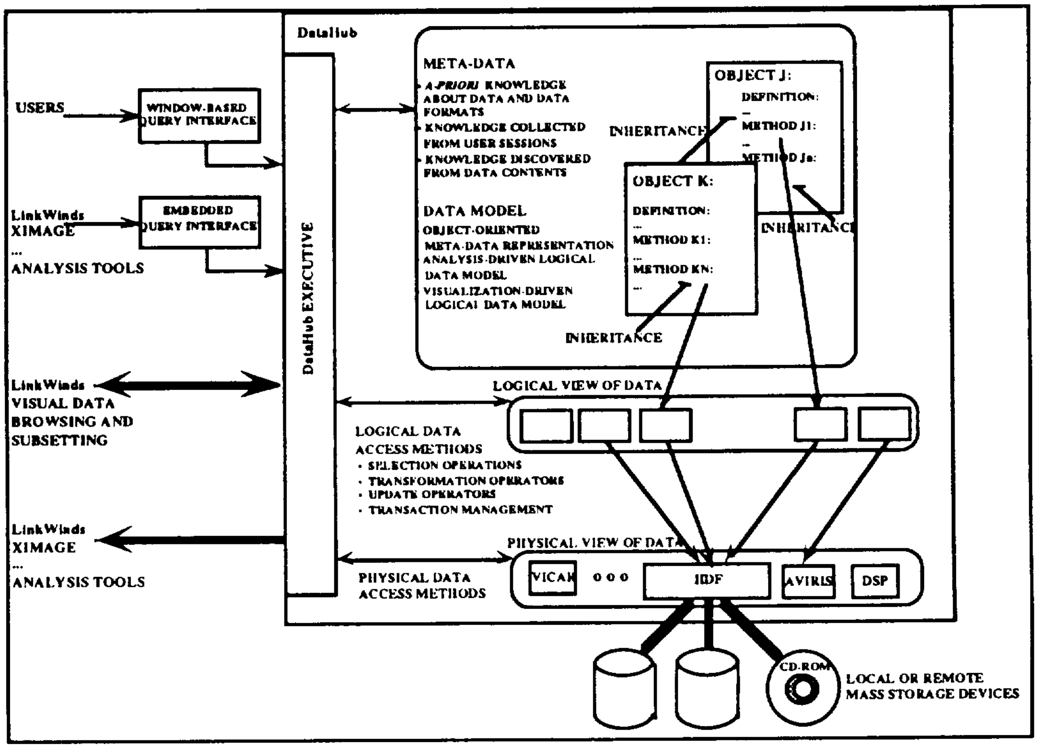


Figure 1 -- Functional Architecture

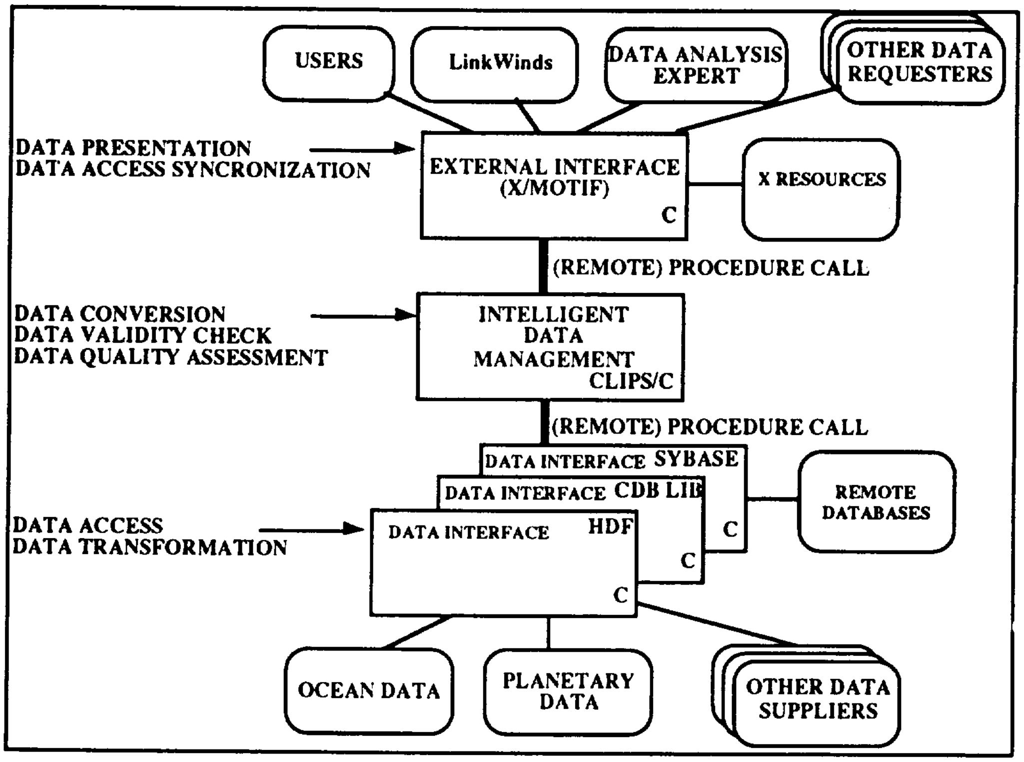


Figure 2 -- Software Architecture

Our initial experience with knowledge-based or machine learning technology was based on work accomplished using artificial neural nets. This work was spurred by our science co-investigators' needs to model regions of the ocean for which the visible and infrared imagery is obscured by clouds, and thus extrapolating biological and physical variables from cloud-free regions in space and time to the cloud-obscured regions. This produced acceptable science products but required too much technical expertise to translate into a generic tool. As described above, new machine learning techniques are being investigated to provide feature recognition capabilities with a more user-friendly interface.

A Recent Developments

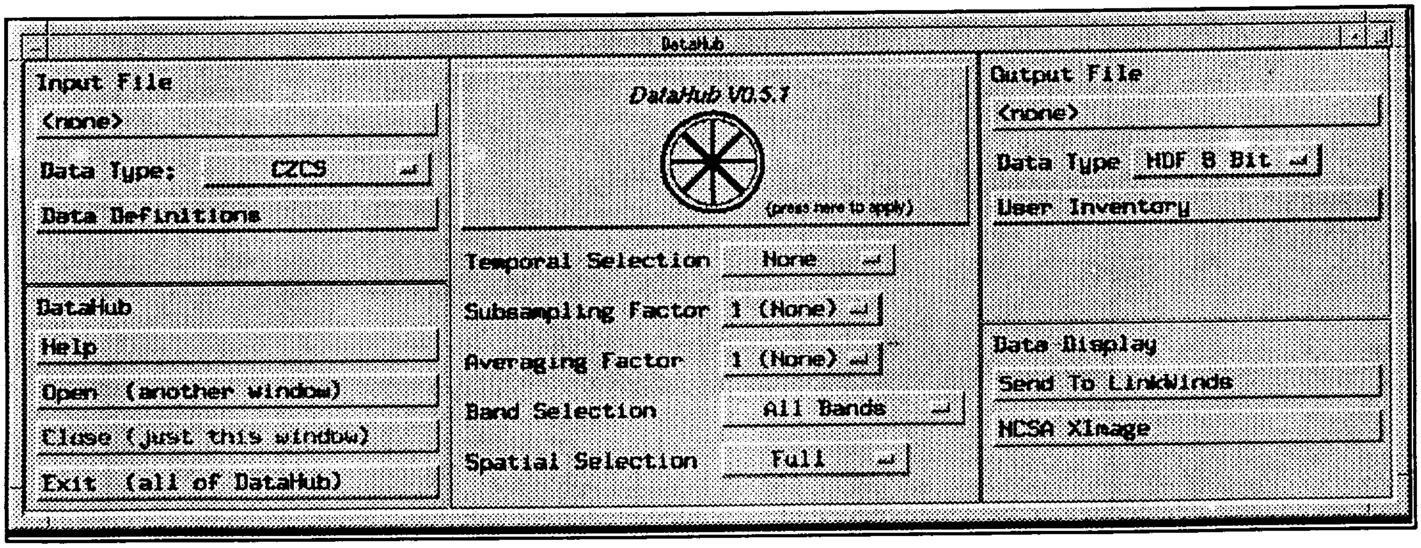
Context Sensitive Help

The DataHub user interface is intended to be self-explanatory and intuitively usable with little or no instruction. In the area of user interfaces, however, intent and reality often diverge.

In packaging DataHub for distribution to a user site outside the development environment, it was obvious that a traditional "README" file was needed to detail installation instructions. It was also clear that although the DataHub user interface had largely succeeded in achieving its goal of intuitive usability, there remained a need for a small amount of instruction to get the first-time user started. While writing a short (< 10 paragraphs) explanatory document, it became obvious that this text could be integrated into the main help system that had been designed into DataHub.

A benefit of using the X Windows resource manager to control an application's user interface is the ease with which all aspects of the interface can be customized. Textual material can be modified as simply as more traditional customizable user interface elements such as colors and layout. Because of this, any instructional text that might otherwise be included in a separate help document (either hard-copy or on-line) can be easily integrated as a dynamic part of the application itself, and eliminate the problems of help being unavailable or not findable when needed.

At the same time, a full-blown hypertext system is neither needed or appropriate for DataHub. Help for DataHub falls into two categories: Initial, new user help, and context-based help for particular DataHub capabilities. The former can be satisfied by a fairly large (as dynamic, on-screen help texts go) set of instructions, and the latter by small explanations easily accessible while the user is performing, or contemplating performing, a DataHub operation. In particular, the navigation of a help system is replaced by the navigation



Input File <none>

Data Type: CZCS

Temporal Selection: None

Subsampling Factor: 1 (None)

Averaging Factor: 1 (None)

Band Selection: All Bands

Spatial Selection: Full

Output File <none>

Data Type: HDF 8 Bit

Figure 3 -- Current DataHub Interface

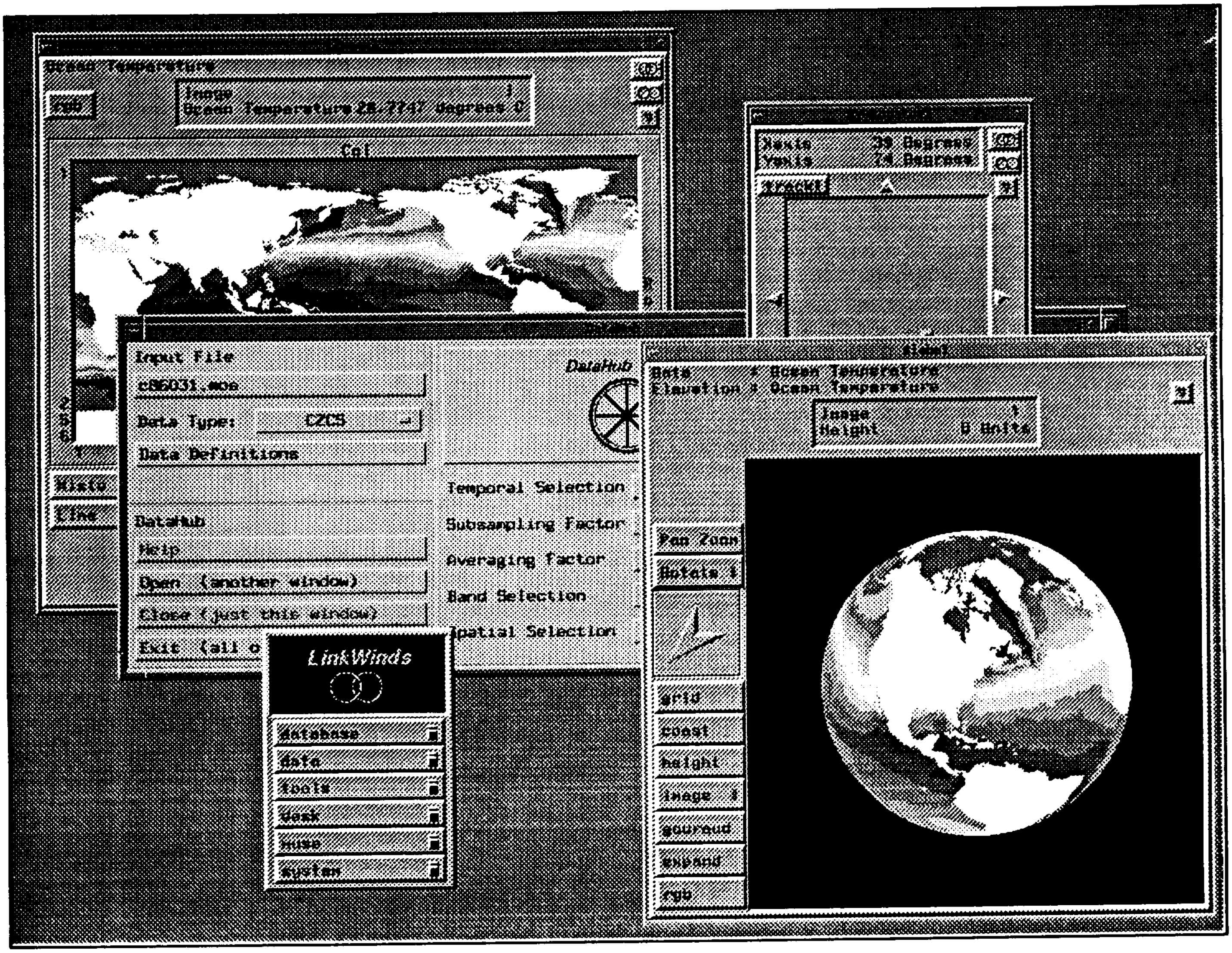


Image: Ocean Temperature 28.7747 Degrees C

Input File: c86031.moe

Data Type: CZCS

Data: Ocean Temperature

Elevation: Ocean Temperature

Image Height: 0 Units

Figure 4 - Typical User Interaction

of the DataHub user interface itself, with single-level help available at each node of the interface.

Multiple, individual, help buttons fit naturally in many parts of the DataHub user interface. A help pulldown menu was added to the section of main DataHub window devoted to generic DataHub control issues. It is in this menu that an item for popping up the introductory text was placed. Additionally, all normal DataHub popup windows have help buttons that popup text dialogs containing help on their particular subject.

More difficult was deciding how to access help for graphical user interface elements (i.e. for the interface element's operations) in cases where the interface was a single button or menu with no place for a separate help button. Pulldown menus can have an additional help item added; simple pushbuttons cannot.

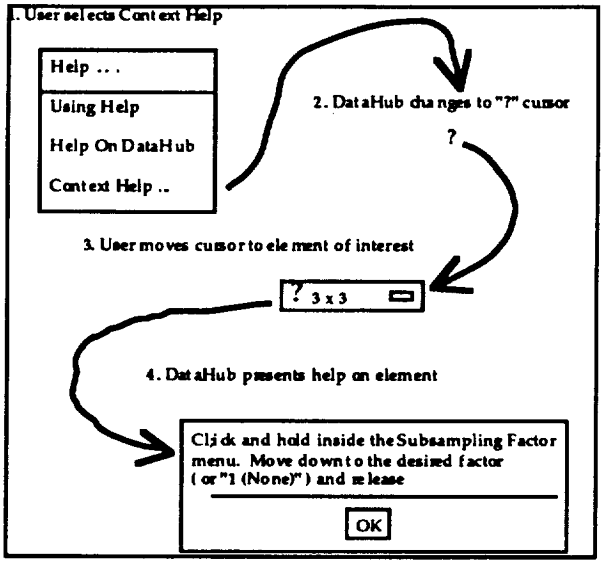
A context help mechanism was implemented for the case of pushbuttons, see Figure 5. The user selects "Context Help ..." from the main help pulldown menu. DataHub acknowledges this input by changing the mouse cursor to a question mark ("?") shape.

The user can then move the question mark cursor to any element of the DataHub interface, and release it to see a help dialog about that element. The underlying code sends a message requesting help to the object representing the graphical element, which in turn displays its textual help.

This method handles any and all kinds of graphic elements, regardless of their screen real-estate limitations. In fact, in the case where an element has a dedicated help button, the context-help method also works, invoking the same message and displaying the same help dialog.

Additionally, help hierarchies are a natural by-product of this implementation. Dropping the question mark cursor onto a graphical element gets help on that subject. Dropping it into the area surrounding the element gets more generic help on the type of interaction the element is a part of. For example, selecting "Subsampling Factor" or "Averaging Factor" displays help on their respective topics, but selecting physically between the two displays help on the subject of subsetting data in general.

The help system can grow and evolve using this framework. If the user drops the question mark cursor onto a graphical element that does not have a help message defined, the message automatically propagates to the ancestor of the element, repeating this process if necessary until it finds one that does have a defined help method. In this way, the user can get help (although more general) even when specific help is yet to be implemented.



Click and hold inside the Subsampling Factor menu. Move down to the desired factor (or "1 (None)") and release

Figure 5 Context Sensitive Help

Portability

Since the goal is to provide an extensible system capable of evolving to provide solutions to broader science and engineering domains, portability is a significant issues. Initially, we conceived using a combination of C, PROLOG, and Common Lisp for the implementation. Today, protablity and minimizing the cost to the user is being addressed by using common platforms (viz.. SUN SPARC stations, Silicon Graphics) and portable and public domain tools (viz. C/C++, FORTRAN, X/MOTIF, CLIPS, UNIX and SQL database interface).

netCDF Data Format

The data format Network Common Data Form (netCDF) was developed by the Unidata Program, sponsored by the Division of Atmospheric Sciences of the Nations Science Foundation. The emerging standard is distributed as an I/O library which stores and retrieves scientific data structures in self-describing, machine independent files. DataHub now recognizes this format.

The current implementation supports
- recognition of netCDF as a file type.
- a set of rules for conversion of netCDF to and from HDF format.

This new capability has been included to facilitate the use of netCDF data in LinkWinds and HDF data in the PolyPaint+ environments.

At this time, netCDF can be seen as providing richer structures. This is supported by the breadth of metadata annotations available as native functions. We found translation from HDF to netCDF more straight forward than the reverse.

What needs to be done?

From the design point-of-view, we have defined a general framework for science data management, and identified a critical subset of data operators for the science data applications. From an implementation perspective, we have developed prototypes that enable validation of basic concepts of data resource sharing between the data suppliers and data consumers (e.g. a data visualization system such as LinkWinds).

Based on the object-oriented design of DataHub, it is straight forward to extend the data model to capture the definitions of an existing relational data system. For example, the comprehensive data catalog built by the Planetary Data System (PDS) will become part of DataHub's data model with specialized data access methods defined to access the existing information in PDS via a standard SQL interface. This approach makes discipline-oriented knowledge readily available to DataHub. Additionally, expanded knowledge about data formats and data semantics in various science disciplines will be built into DataHub. It is a goal that the understanding of the visualization and analysis tools will also become part of DataHub such that special data operators will be built automatically using basic known operators. The data quality assessment issue of science data after data transformation will be a research area for DataHub, and will be addressed in the next steps.

We will enhance the existing prototype to provide access to additional data sets while expanding the capabilities for direct support to the science co-investigator. Particularly, the issues associated with processing multi-spectral data will be addressed. We will be enhancing the preprocessing capabilities by accessing and utilizing the NAIF SPICE ancillary data as it become available.

Besides continuing to evolve to a more object-oriented implementation, several issues will be addressed. When data transformation or conversions take place, we need to assure the preservation of data validity or quality measures. We need to treat the data quality assessment issues such as (1) treatment of missing data and (2) data quality associated with data interpolation, data transformation, etc.

Expanded knowledge about the data is of significant importance. This includes knowledge of data formats (e.g. usage of metadata embedded in the data set headers), data semantics (e.g. meaning of data values, relationships between data sets, discipline-dependent data access/analysis methods) and data semantics as represented by the users' context in the visualization regime (e.g. what are the links, dataflows, etc. as encapsulated in the LinkWinds environment). The ability to detect and understand this expanded knowledge will be incorporated into the label-understanding expert system.

Additional understanding of the analytical tools required for data selection, data transformation and data conversion in order to support the visualization requirements is needed. These may to thought of as filtering tools to select and prepare data for use in the visualization environment. These additional tools will be defined and implemented.

In those cases where selection criteria are so complex that they are most easily exercised visually, it is clear that a close integration of the database management system, and the data visualization system is advantageous. Such integration will be studied by closely tying DataHub with LinkWinds so that DataHub will be accessible from LinkWinds and LinkWinds will be accessible from DataHub, each being used to best advantage in the data management processes.

Finally, we will address the issues associated with data presentation. In particular, data exchange protocols that facilitate visualization are to be addressed first.

Major Components

DataHub will be enhanced to include these capabilities:

- interactions to support finding, selecting and processing multi-spectral datasets (initially AVIRIS).
- band aggregations
- band filters (e.g., removal of artifacts of the instrument)
- 3D subsetting/averaging
- journal and transaction management will playback capability.
- expanded data model that includes user-defined data conversions.
- canonical set of data objects and methods
- self-describing data objects and methods
- user defined defaults for spatial regions, temporal periods, etc.
- incorporate the metadata into the interfaces with LinkWinds and PolyPaint+.
- expanded rule-based capability to understand foreign datasets, leading to a capability for interpretative conversions and transformations.
- expanded data dictionary for use in label recognition, plus the ability to dynamically add

new object attributes once their semantics are clearly understood.

- initial usage of calibration and registration data
- quality measures, to include
- processing lineage
- null and missing value recognition and usage in processing
- incorporation of content-based applications such as the machine learning capability described above
- expanded interactions with LinkWinds and PolyPaint+
- distribution of DataHub processing and interactions and remote services.

Using the DataHub, scientists will request data for presentation and analysis in a specific way for use in the their applications, without being particularly concerned with the original location and format of data being utilized. Applications adhering to the DataHub protocols and interfaces may interoperate sharing results through the DataHub.

As described previously, LinkWinds will be enhanced to have two-way communications with DataHub. Besides receiving the user's selected data for analysis, LinkWinds will provide graphical subsetting and transformation parameters and send a processing request for DataHub to execute and return the desired data.

PolyPaint+ will have a similar interface as LinkWinds. After this communications and processing link has been implemented, DataHub will be enhanced to provide more specialized processing for the PolyPaint+ community (that is say, netCDF, super computing, and modeling).

Machine Learning and Feature Detection

It is difficult for a scientist to examine and understand data with a large number of dimensions. Scientific visualization tools are one means for performing necessary transformations and dimensionality reduction to allow a scientist to "see" meaningful patterns in the data. However, these require that the scientist specify the necessary steps. Faced with multi-spectral remote-sensing data arriving over more than 200 channels, expecting a scientist to study the entire data set becomes unreasonable. This often results in using only parts of the data channels or using the data in very limited ways. An automated tool for aiding the analysis of such high dimensional data sets would enable scientist to get at more of the information contained in the data.

We will use machine learning and pattern recognition techniques to aid in the analysis of multispectral data. Consider a scientist interested in characterizing certain regions in the data, for example, locating the areas on earth where certain minerals are present, or where some phenomenon of interest occurred. By selecting portions of the data of interest and others that do not contain phenomena of interest, a scientist is essentially pointing out examples (instances) of the desired target. These can be treated as training data, and used by learning algorithms to automatically formulate classifiers that can detect other occurrences of the target pattern in a large data set. Furthermore, since the learning algorithms are capable of examining a large number of dimensions at once, they may be able to find patterns that would be too difficult for a scientist to derive by manual analysis. In a sense, this offers the option for a "logical" versus a "visual" visualization of the patterns in the data. That is, the algorithms produce a characterization of subsets of interest in the data in terms of logical expressions involving multiple input variables (channels). Often, it is possible to express such patterns in terms of compact rules involving an unexpectedly small number of variables. For example, channels 104 and 202 being in certain ranges may be highly predictive of a phenomenon that the scientists could not easily characterize using the first six channels.

The use of learning algorithms thus provides flexibility in terms of adapting to a wide variety of detection problems. Our decision tree based learning algorithms produce rules that are easily examined and understood by humans. This contrasts with a statistical regression or neural network based approach, where the resulting forms are difficult to interpret.

Distributed Blackboard System

The blackboard model allows for a flexible architecture with diverse knowledge sources cooperating to formulate a solution opportunistically [16]. A distributed blackboard system running across multiple workstations can allow multiple scientists in different physical locations to work together on a single problem cooperatively.

The DataHub metadata manager has been ported to a distributed environment across multiple Sun SPARCstations [22] . This distributed environment is the underlying layer of an ongoing distributed blackboard implementation. It is expected that the DataHub system can sit on top of this blackboard system to function as a Groupware for multiple scientists from multiple science disciplines.

With this capability, DataHub can distribute the data access and data conversion load across multiple computers. At the same time, multiple users can access multiple data sources via this distributed scheme of DataHub.

With the distributed blackboard, DataHub can have multiple data servers with metadata (i.e., discipline knowledge) about multiple data sources sitting across the network. Each data server acts as an independent knowledge source in the blackboard system. The DataHub data servers use a consistent data access mechanism provided by DataHub. The scientists use a consistent user interface of DataHub event though they are running the DataHub data client on their own workstations geographically separated from one another.

The inter-disciplinary knowledge about data can be stored in higher level knowledge sources (i.e., agents) in the blackboard system. Whenever a scientist has a need of a dataset that is outside a single discipline, this inter-disciplinary knowledge source is utilized to provide intelligent data access capability to access the right data from the right source.

The distributed blackboard is implemented using a reliable distributed computing protocol provided by Cornell's ISIS [1, 7]. ISIS version 2.1 is in the public domain. The concept of having process groups in a distributed environment with guaranty on message arrival sequence for messages from multiple senders fits the need of the blackboard implementation.

Development and Deliverables

We have planned three steps in the next phase of DataHub prototyping:

Step 1.

- Design and develop DataHub processing of multi-spectral data sets for the science co-investigator.
- Initiate the distribution of DataHub processing and provide general remote services.
- Design and develop interfaces to PolyPaint+. Collect functional requirements from the user community.
- Design and develop the machine learning interface.
- Demonstrations will use the data sets as determined by the science co-investigator.

Step 2.

- Provide data abstraction and knowledge engineering to support applications in the LinkWinds and PolyPaint+ environments.
- Demonstrations will use the data sets as determined by the PolyPaint+ user community.

Step 3.

- Provide the knowledge engineering required to utilize the computing environment and its tools. Incorporate this knowledge into the DataHub.
- Provide support within the DataHub of all the required datasets (homogeneous/regular and heterogeneous). .
- Demonstrations will use the data sets as determined previously.

The development cycle used to solve the problems addressed above will be to: define/expand the science co-investigator's problem; design, implement, integrate test, demonstrate, evaluate, and transfer to the scientist co-investigator; and then iterate these steps. In each cycle these areas will be addressed: (1) The DataHub; (2) Knowledge-based assistance for the DataHub; (3) Machine learning for feature recognition; (4) A problem posed by a science co-investigator ; and (5) LinkWinds/PolyPaint+ interface and protocol.

An incremental development methodology will be utilized: "do-a-little, test-a-little".

Throughout the implementation effort, the science co-investigator and other scientists will participate in the design. This feedback and evaluation is important in providing a product that contributes to the scientists' ability to accomplish their science objectives. The success of the proposed work will be measured by the science utility of the work products.

Benefits and Expected Results

The principle product of the proposed work is the demonstration of an integrated environment in which a science co-investigator will be able to accomplish data analysis and interpretation leading to publishable scientific information. Thus, DataHub is addressing broad aspects of:

1. Providing innovative ways to facilitate the scientific endeavor or "mean-time to discovery" [33] when working with large volumes of data. The traditional computing data life-cycle is typically a sequential process. This traditional view provides sequential support to what is actually a highly-interactive, iterative process. DataHub will provide a data life-cycle as illustrated in Figure 3.
2. Providing access to remote data, local data filtering and management and interactive exploratory data analysis.
3. Applying knowledge-based expert systems and machine learning at the original data selection, in intermediate data filtering and in rule-based applications.

The DataHub will provide an end-to-end solution to problems of this generic type, thus enabling science investigators to produce higher-level products through an analysis environment which provides an integration of required functions. This environment consists of:

1. An interface between the scientific visualization and analysis environment and the data required to perform the analysis.
2. Expert system / knowledge engineering-based analysis assistants and machine learning techniques to do:
   - data discovery and data selection
   - feature and image understanding preprocessing
   - visualization and analysis tool selection
3. The LinkWinds and PolyPaint+ environments and their analysis tools as the visualization mechanism and user interface environment.

The benefits to NASA deriving from the DataHub include:

1. Ability to analyze massive volumes of data in a cost-effective manner.
2. Freedom for the NASA mission scientists to do the interpretative, creative aspects of science work.
3. An advanced prototype for science support.
4. Availability of common system modules and data formats for other developers.

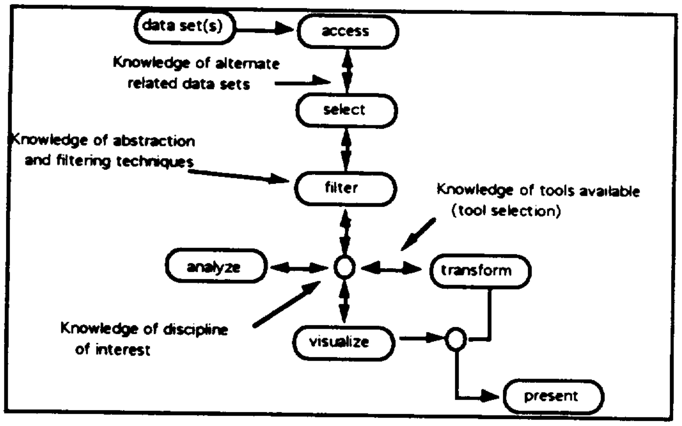


Figure 3 -- Knowledge-Based Visualization and Analysis

Acknowledgments

This work was supported by NASA through Joe Bredekamp's Applied Information Systems Research program. The authors wish to thank the scientist users of DataHub for their valuable insights into discipline-specific needs in data usage and semantics and user interactions.

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

If you have comments, questions, or would like to discuss the uses of DataHub, the authors may be reached at:

T. Handley
thandley@spacemouse.jpl.nasa.gov
M. Rubin
mark@phineas.jpl.nasa.gov

Bibliography

1. K. Birman, A. Schiper, P. Stephenson, "Lightweight Causal and Atomic Group Multicase", ACM Transactions on Computer Systems, Vol. 9, No. 3, August, 1991.
2. M. Botts, Informal Communication "Results of a NASA survey for EosDIS ", University of Alabama, June 2, 1992.
3. CDF User's Guide for UNIX Systems Version 2.1, January 9, 1992, National Space Science Data Center, NASA - Goddard Space Flight Center, Greenbelt, MA.
4. Joseph C. Ciarrarano, CLIPS User's Guide Version 5.1, September 10,1991, Volumes 1-5, Johnson Space Center, Information Systems Directorate, Software Technology Branch.
5. Peter J. Denning, "Information Technologies for Astrophysics Circa 2001- A Position Paper", NASA Workshop on Astrophysical Information Systems, May 23-25, 1990.
6. Stephanie Granger-Gallegos, et. al., "Automatic Data Distribution for a Large Geophysical Data Set", submitted to The Earth Observer, July 1992.
7. O. Hagsand, H. Herzog, K. Birman, R. Cooper, "Object-Oriented Reliable Distributed Programming", 2nd International Workshop on Object-Orientation in Operating Systems, 1992.
8. Thomas H. Handley, Jr., Larry E. Preheim, "High Rate Information Systems: Architectural Trends in Support of the Interdisciplinary Investigator", AIAA/NASA Second International Symposium on Space Information Systems, September 1990.
9. Thomas H. Handley, Jr., Larry E. Preheim, "NASA Information Systems Commonality and Convergence", AIAA Computing in Aerospace 8, October 1991.
10. Thomas H. Handley, Jr. P. Li, "DataHub: Intelligent Data Management for Data Discovery", International Space Year Conference on Earth and Space Science Information Systems, February 10-13, 1992.
11. Thomas H. Handley, Jr. Y. Philip Li, Allan S. Jacobson, An V. Tran, "DataHub: Intelligent Data Management", 1992 ASPRS/ACSM Convention, August 3 - 7, Washington, D. C.
12. Thomas H. Handley, Jr., Mark Rubin, Y. Philip Li, "DataHub - Knowledge-based Science Data

Management for Exploratory Data Analysis", SPIE Spring 1993, Orlando, Florida.

13. Thomas H. Handley, Jr., " DataHub: Intelligent Distributed Science Data Management to Support Interactive Exploratory Analysis", Computing in Aerospace 9, San Diego, California, October 19-21, 1993.
14. Thomas H. Handley, Jr., "DataHub: Science Data Management in Support of Interactive Exploratory Analysis", PECORA12, August 24-26, Sioux Falls, South Dakota.
15. Klemp, J. B., M. J. McIrvin, and W. S. Boyd. 1990: PolyPaint - A Three-Dimensional Rendering Package. In Proc. Sixth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, 5-9 February 1990, Anaheim, Calif. AMS, Boston, Mass., 286-293.
16. B. Hayes-Roth, "A Blackboard Architecture for Control", Artificial Intelligence, 26 (1985).
17. Susan LaVoie, et. al., VICAR User's Guide Version 2, June 1, 1989, JPL D-4186, Rev A.
18. Joel L. Morrison and Kathryn Wortman, Editors Special Issue "Implementing the Spatial Data Transfer Standard", Cartography and Geography Information Systems , Vol. 19, No. 5, December 1992.
19. NAIF AMMOS V18.1 MO Delivery Documentation Package, Navigation Ancillary Information Facility, 25 February 1993
20. NetCDF User's Guide Version 2.0, October 1991, Unidata Program Center, Boulder, CO.
21. T. Handley and L. Preheim, "High Rate Information Systems: Architectural Trends in Support of the Interdisciplinary Investigator", AIAA-90-5084, AIAA Second International Symposium on Space Information Systems, 17-19 September 1990, Pasadena, CA.
22. Y. Philip Li, "dCLIPS: A Distributed Clips Implementation", Computing in Aerospace 9, San Diego, CA, October 1993.
23. Allan S. Jacobson and Andrew L. Berkin, "LinkWinds: A System for Interactive Scientific Data Analysis and Visualization", Proceedings of High Performance Computing Conference, Society for Computer Simulations, March, 1993.
24. NCSA HDF Calling Interfaces and Utilities Version 3.1, July 1990, NCSA Software Tools Group, University of Illinois at Urbana-Champaign, IL.
25. L. Preheim, T. Handley, "Data System Interoperability Task, Requirements Commonality Study Report", JPL D-7187, February, 1990.
26 M. Stonebraker and J. Dozier, "Large Capacity Object Servers to Support Global Change Research", Report # 91/1, SEQUOIA 2000, University of California.
27. Russ Rew and Glenn Davis, "NetCDF: An Interface for Scientific Data Access", IEEE Computer Graphics & Applications, July 1990, pp. 76-82.
28. A. Silberschatz, Michael Stonebraker, Jeffery D. Ullman, "Database Systems: Achievements and Opportunities", TR-90-22 Department of Computer Sciences, The University of Texas at Austin, Report of the NSF Invitational Workshop on the Future of Database Systems Research held in Palo Alto, California, February 22023, 1990.
29. Spatial Data Needs: The Future of the National Mapping Program, National Academy Press, Washington, D.C., 1990
30. Lloyd A. Treinish, "SIGGRAPH '90 Workshop Report Data Structures and Access Software for Scientific Visualization", Computer Graphics, Vol. 25, No. 2, April 1991
31. Lloyd A. Treinish, "An interactive, discipline-independent data visualization system", Computers in Physics, Jul/Aug 1989, pp. 55-64.
32 James Warner, "Visual Data Analysis into the '90s", Pixel, January/February 1990, pp. 40-44.
33. M. Mitchell Waddrop, "Learning to Drink from a Fire Hose", Science, Vol. 248, pp. 674-675, May 11, 1990.