The HIRAAS INFOBASE

SH51A-0426

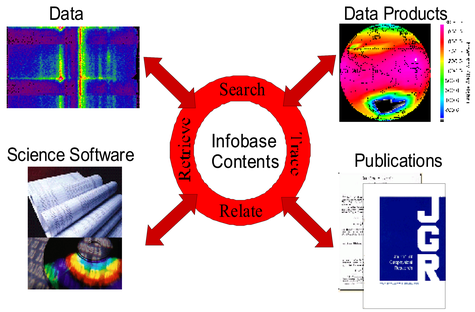
C.B. Fortna, SFA Inc, Largo, MD,
K.F. Dymond, S.A. Budzien, E. O. Hulburt Center for Space Research, Naval Research Lab, Washington, DC,
J.B. Gara Raytheon Technical Services Company (ITSS), Lanham, MD

SH51A-0426

The High Resolution Airglow and Aurora Spectroscopy (HIRAAS) experiment was successfully launched on the USAF Advanced Research and Global Observation Satellite (ARGOS) and remotely sensed the thermospheric and ionospheric airglow from May 1999 to April 2002 using three ultraviolet spectrographs.

W

The HIRAAS "Infobase" is a dynamic scientific data management system that integrates data products, processing software, and parent data sets with full heritage tracking. The Infobase breaks from the traditional catalog/flat-file approach, instead storing all data products internally in a relational system, customized for limited object-oriented functionality. Custom code provides for direct delivery of data products to science users for reduction, modeling and analysis. The Infobase paradigm has wide applicability for current and future remote-sensing and ground-based datasets being managed at NRL.



The Infobase is a dynamic, integrated, scientific, database management system:

DYNAMIC -- On-going incorporation of new and imporved data products.

INTEGRATED -- Data Storage, searching, product generation, and heritage tracking are all handled within the Infobase.

SCIENTIFIC -- Focused scientists' and the demands of scientific data analysis (especially aeronomy).

DATABASE MANAGEMENT SYSTEM -- The database standard requirements for data integrity, accessibility, security, robustness.

IMPROVED PRODUCTIVITY

The Infobase is a system designed by scientists for scientists, based upon experience using previous data sets and the requirements for atmospheric modeling and analysis. The Infobase stores all HIRAAS measurements and data required for analysis. The Infobase is required to be reliable, robust and ensure data integrity. It must organize the information in a searchable format. Best effort is made to comply with industry standards for cross-platform compatibility. The Infobase automatically processes telemetry files and acquires environmental parameters. Documentation and software are incorporated so that data reduction and calibration techniques are standardized amoung various users. Configuration management and heritage tracking of data products and software is provided. Measurements, data, and products can be directly exported into analysis software. The Infobase is scalable to grow with the experiments requirements. In this way the scientists productivity is improved.

GUI or SQL STANDARD RETRIEVAL:

-- Data product and information retrieval may be performed through Custom GUI interface software
-- or any of a number of SQL compatible commercial applications.

REPRODUCIBLE RESULTS:

-- The processing software is never lost because it is always retrieved from the Infobase.
-- Results can always be replicated because both the parent data sets and parent processing software are stored and linked to the child data result set produced.
-- One may find all datasets and software that went into a product.
-- Results of one software run may be directly compared to results of a different software run.

KNOWN VERSIONING:

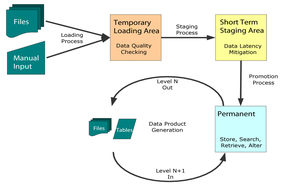
-- Software versioning and configuration management is handled by the Infobase.
-- Documentation is contained along with the software and products.
-- Data Products are tagged with the software version that produced them.

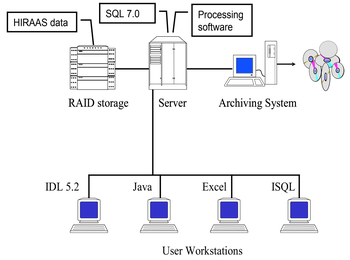
PUBLICATIONS STORED & TRACKED:

-- One may retrieve all publications and products made from a dataset.
-- One may find all datasets and software that went into a product or publication.

MAIN FEATURES

On line all the time. (NO Files)
Contains Data, Software and Data Products.
Heritage Tracking of Data Products.
Organically grows to accommodate new products.
Applicable to numerous experimental data sets.



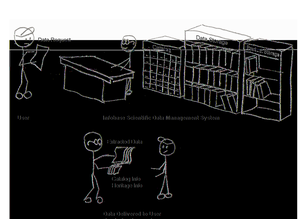


This is a diagram of the Infobase computer system. A Server with RAID storage system runs the Relational Database and additional Infobase software processes. Data is retrieved and processed on user work stations through standard application SQL interfaces.

COMPARISON OF INFOBASE TO CONVENTIONAL SYSTEM

Unlike traditional scientific database catalog with flat file data, the Infobase remains dynamic throughout the entire period of scientific analysis, even after observations have ceased. The full set of measurements, data products, software, and documentation remains accessible on-line at all times and may be accessed through a wide variety of off-the-shelf interfaces. Consequently, the Infobase system delivers an unequalled combination of ease of use, efficient data storage, extremely flixible search capability, numberous data retrieval options, and tracking of data.

FEATURE	INFOBASE	TRADITIONAL
Data delivery	RESULT SET	FIXED LENGTH FILES
Data access	100% CONTINUOUS	FILE-BY-FILE
Searchable on relationships and conditions	ANY IMAGINABLE	PROGRAMMED
Retrieve cross joined data sets	YES	NO
Retrieve sparse data scattered over entire mission	YES	PAINFUL
Full product tracking from raw data to final publication	IMPLEMENTED	POSSIBLE
Cross listed data source, products, software, version and author	ALL	ONLY PROGRAMMED
Standard access method to any user application	ANSI SQL-92	CUSTOM FILE PARSER
Routine automated data processing and product generation	MANAGED	PROGRAMMED
Manual data entry, processing and product generation	INCLUDED	POSSIBLE
Standard description of data product definitions	"SCHEMA"	MAYBE
Documentation	INCLUDED	SEPARATE FILES



HOW DOES IT WORK?

The full Infobase system incorporates hardware, software, and information. The hardware components include a Server with RAID storage device networked to user workstation computers. The software elements include Database Management Software (DBMS), the Infobase schema itself, telementry stripping codes, data processing codes, automated data ingestion routines, customized data delivery routines, and configuration management code. Finally the information resident in the Infobase includes HIRAAS experiment measurements, HIRAAS engineering data, ARGOS satellite engineering data, environmental parameters, along with advanced data products like viewing geometry, retrieved atmospheric densities and temperature profiles, movies, and journal publications.

A Relational Database (RDB) system was chosen as the core of the Infobase. Although Object-Oriented Database (OODB) systems are available supporting inheritance of data properties and methods, they are not efficient at ad hoc queries due to their entity tree-like storage structure. RDBs are very effiecient at storing, searching, and delivering large amounts of data. Additionally, RDBs have built in mechanisms for ensuring data integrity and providing authentication/security. Object-Oriented features for heritage tracking and configuration management are incorporated into the RDB system by carefully designed table relationships and externally interfaced data processing software.

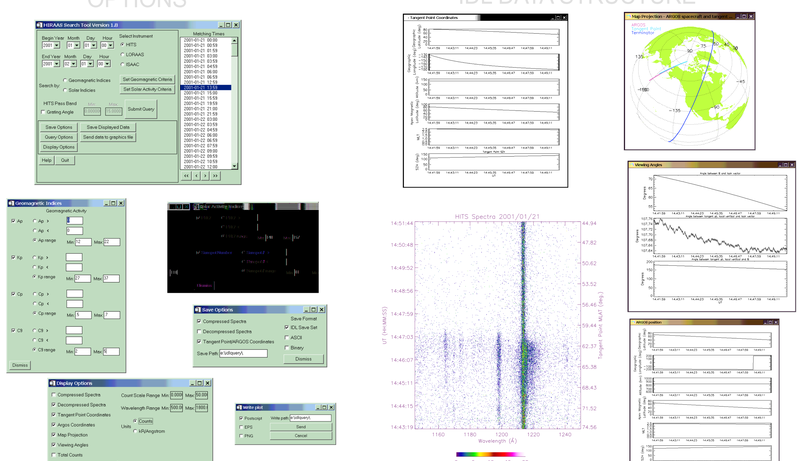
The RDB system is based on the ANSI Sequential Query Language (SQL-92) standard which includes a Data Manipulation Language (DML) and Data Control language (DCL) to define the user and application interface methods. Data is stored as rows in tables. Each table column defines a data type. The Binary Large Object (BLOB) data type is utilized to store files representing executable software, documentation, and publications into the Infobase. The SQL languague is used to establish relationships amoung the columns and tables. The SQL operations are performed on sets of the entire database and are not limited by file structures.

A thorough understanding of the data types, data formats, data set sizes, and data quality considerations are required to design the Infobase schema of individual tables, the relationships among tables, and the processes used to populate the tables. Satellite telemetry files are stripped out and moved through loading and staging area tables to quality check the data, remove redundancy and latency inherent in receiving files. Data is then moved into the Infobase Level 0 Data Product tables with full Heritage Tracking. Hand entered data and calibration files may also be incorporated into Level 0 Data Product tables.

Each table in the Infobase containing raw measurements or processed data is treated as a "Data Product." The information in Data Product tables is controlled by strong "Heritage Tracking". Every Data Product table has a "Data Product Descriptor" with identification key index. All automated processing software is contained in Infobase tables under a configuration management schema. When a lower Level Data Product is processed into an higher Level Data Product, the processesing software version is retrieved from the Infobase for execution. The parenting Data Product blocks and Processing Software are key indexed to the resulting Data Product blocks, thus carrying the full Heritage Tracking information.

Information is distributed to user applications through the SQL standard internet access methods. SQL interfaces are available to most computer languages including: C++, JAVA, EXCEL, IDL. Requested "Queries" for data sets are returned directly into the requesting language data types. There is no need for intermediary files. Custom GUI interfaces are written to give a "know nothing" User easy access. The more knowledgeable users can generate their own Queries to obtain result sets unbounded by Custom GUIs or file sets.

HOW DO YOU USE AN INFOBASE?

SELECT QUERY OPTIONS

DATA RETRIEVED INTO IDL DATA STRUCTURE