A System Design for Storing, Archiving, and Retrieving Hyperspectral Data

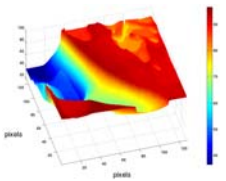
Ralph G. Dedecker, Tom Whittaker, Ray K. Garcia, Robert O. Knuteson*
University of Wisconsin-Madison, Space Science and Engineering Center

UW-Madison

Introduction

Hyperspectral data and products derived from instrumentation such as AIRS, CrIS, GIFTS and HES will impose storage and data retrieval requirements that far exceed the demands of earlier generation remote sensing instrumentation used for atmospheric science research. A new architecture designed to address projected real time and research needs is undergoing prototype design and development.

GIFTS Simulated Data Cube



1 Terabyte per day

The large volume of data collected and products produced from hyperspectral instrumentation will require large distributed storage devices employing several servers. The hardware infrastructure must be implemented in such a way that component augmentation, replacement, and maintenance can be undertaken without undue demands to modify user applications. User applications will need tools to simplify locating data files. User data selection facilities for retrieving specific information from storage devices for calibration, analysis, instrument inter-comparison, or reference purposes will also be necessary and standardized data formats and data delivery schemes will be important.

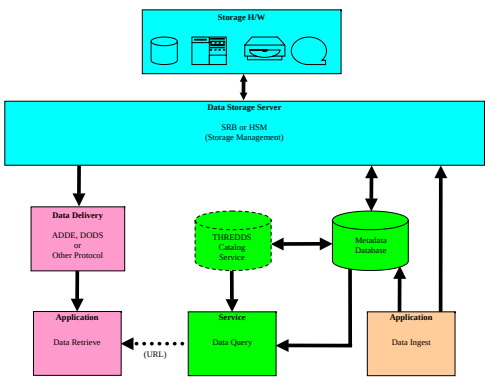
Conceptual Design

Design considerations -

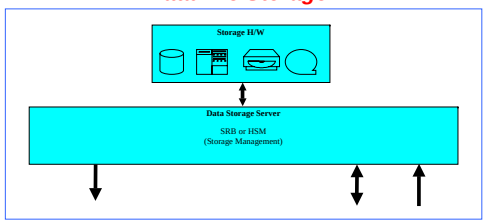
- Data file storage - distribution, interfacing, maintenance, upgrades, etc.
- Data cataloging
- Database query capabilities
- Data delivery methods

The design goal is to define a hardware infrastructure and software libraries to support the above while adhering to established research application standards where possible.

Conceptual Storage System Overview



Data File Storage

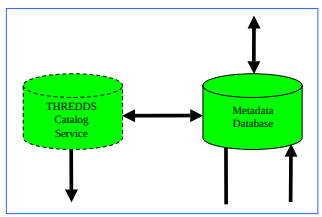


The data file storage subsystem simplifies user applications by decoupling the applications from the data storage devices and the system storage server.

Hardware interfacing foundation –
- Online devices, Off-line Archive
- Grid Interface
- Optional commercial HSM
- Easily replaced, repaired, etc.

Software infrastructure –
- Device independent user applications
- File links maintained by system
- Optional open source implementation (SRB)

Data Cataloging

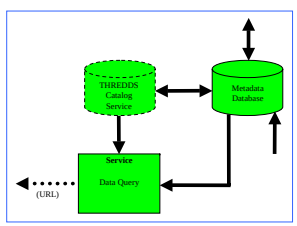


Data files include metadata –
- Describe measurements
- Describe products
- Quality assessment
- List associated data
- Links to associated data

Data catalog –
- Lists primary metadata
- Includes thumbnail images
- User comments
- Quality assessment
- Can be browsed
- THREDDS implementation

THREDDS provides a requesting user program with a URL for accessing the data represented by the catalog entry and a method for visualizing, overlaying, and defining a coordinate system for these data.

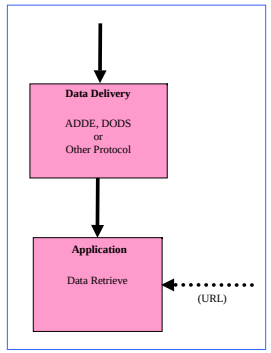
Database Query Capabilities



The total volume of stored data is anticipated to be very high and is expected to increase over time. File searches for results and case study data using traditional methodologies would be tedious and cumbersome. This architecture includes database features.

SQL searches –
- Speed data selection
- Multiple metadata search
- Linked to THREDDS

Data Delivery Methods



Data Delivery –
- Via Internet
- Directly to local workstations
- Near real time
- Via SCP, SFTP, etc.
- Scheduled with "ready" notification
- With data slicing, merging, translation

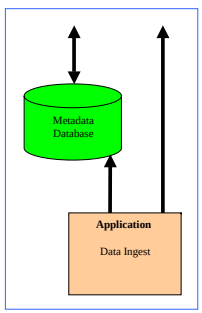
To facilitate remote delivery and differences in local data formats, standardized protocols and software tools will be employed for data access and transfer as well as for visualization of the data.

Protocols –
- ADDE
- DODS
- Others

ADDE -
- Retrieves data
- Delivers to client properly formatted
- Extracts subsets of image, grid, point data

DODS -
- Local format independent
- Serves in HDF, NetCDF, etc.
- Open Source
- Interfaces to Matlab, IDV, IDL, Ferret

Data Submission to the Storage Library



This design concept assumes user supplied data for storage in the library will meet quality screening requirements and will provide descriptive information and references to properly interface with the library infrastructure. This subsystem will provide tools to assist user data submissions.

Data library -
- Under curator control
- Maintenance functions
- Application tools for metadata generation and formatting

Implementation Status

This conceptual design is intended to serve as a guideline for efforts to explore operational system design facets through the development and evaluation of prototype systems. The intent is to refine the concepts and to implement an operational data storage system suitable for the next generation of instrumentation. Work at the UW-SSEC has already begun to prototype and apply some of these design concepts using data collected from existing hyperspectral and remote sensing instrumentation.

Information Sources

THREDDS: http://my.unidata.ucar.edu/content/projects/THREDDS/index.html
ADDE: http://www.ssec.wisc.edu/mug/prog_man/2003/servers.html
DODS: http://www.unidata.ucar.edu/packages/dods/index.html

* corresponding author - Ralph.Dedecker@ssec.wisc.edu

International TOVS Study Conference, 13th, TOVS-13, Sainte Adele, Quebec, Canada, 29 October-4 November 2003. Madison, WI, University of Wisconsin-Madison, Space Science and Engineering Center, Cooperative Institute for Meteorological Satellite Studies, 2003.