Reprint Seventeenth Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology, 81st AMS, Jan. 14-19, 2001, Albuquerque, NM, American Meteorological Society, 2001, pp. 399-403.

10.13 Network-based Visualization Using the Distributed Image SpreadSheet (DISS)

K. Palaniappan, A. F. Hasler[†], J. Fraser, and M. Manyin[†]
Department of Computer Engineering and Computer Science
University of Missouri-Columbia, MO 65211
[†] Laboratory for Atmospheres
NASA Goddard Space Flight Center, Greenbelt MD 20771

1. INTRODUCTION

Geophysical digital libraries are now approaching petabytes in size with data accumulating from the new generation of Earth Observing System (EOS) high data rate satellite sensors [1]. Visualization tools that can directly access data from digital libraries or other network storage devices without necessarily trying to create a "DAAC (Distributed Active Archive Center) on a desktop" will facilitate more thorough and timely use of data in remote sensing archives. The growth of large databases, hyperimaging satellites, multisource observations, and complex coupled earth-ocean-atomsphere models combined with the limited number of scientists that analyze this large volume of data, provides a strong need to develop new tools that combine visualization and analysis to improve and leverage scientific productivity. These tools must be capable of manipulating gigabyte-sized scientific datasets now and terabyte-sized datasets in the near term.

The DISS extends the Interactive Image SpreadSheet (IISS) software tool for accessing large distributed remote sensing digital libraries [2][3]. The DISS provides high speed network access using http and ftp methods directly, combined with novel data compression schemes, and network-based data caching for low latency and efficient bandwidth utilization. Both the DISS and IISS extend the popular business spreadsheet paradigm from scalar quantities to multidimensional image and data arrays. The DISS software tool uses a multidimensional spreadsheet arrangement of cells and frames for manipulating gigabytes of multichannel image and model data, and supports image and grid analysis algorithms. An example of 3D assimilated (HDF) data visualization is shown in Figure 1. A unique DISS capability is to rapidly and smoothly zoom, roam, animate and execute functions in synchrony for a large set of 2-D or 3-D data cells (e.g. Fig. 1). Large dataset visualization is facilitated using multiresolution tiling methods [4]. Other software systems are beginning to adopt the multi-cell organization of displays [5] that was pioneered in the IISS environment.

The DISS extends the IISS capability for handling extremely large Earth science hyperdatasets on local file systems, to take advantage of the new ultrahigh performance Internet network systems. Combined with Bulk Data Service (BDS), a software extension to the commonly used NFS (Network File System) protocol, the DISS allows for fast transfers of large sequentially read data files. SGI developed BDS to allow NFS devices to fully exploit high performance networking technologies, such as Fibre Channel, HIPPI, and ATM. Native NFS, developed in an era when (10BaseT) Ethernet was state-of-the-art, is well suited for small or randomly accessed files but not optimized for large data transfers. For sequential data transfers, BDS delivers two to four times the bandwidth available with native NFS across channels.

Due to the wide range of performance capabilities and requirements, many visualization and analysis tools have been developed [6]. The DISS and IISS tools enhance human perceptual abilities for the visualization and analysis (*visanalysis*) of extremely large multidimensional geophysical datasets – image, observation or model "hyper-datasets". We emphasize the fact that both visualization and analysis are equally important components, in the scientific study of complex phenomena using multiple data sources and large data volumes by synthesizing the term *visanalysis* in 1992. The term hyper-dataset is used to refer to the fact that modern remote sensing datasets have increased resolution in several dimensions such as temporal, spatial, spectral or radiometric quantization. Visanalysis tools offer a flexible approach to the exploratory analysis, understanding and qualitative assimilation of complex datasets.

The DISS uses high performance networks to enable scientists to study and compare the large volumes of data often in a near realtime mode. Highly interactive visual browsing tools combined with 2-D and 3-D graphics rendering in each frame of the DISS have been demonstrated to be highly successful for quickly inspecting thousands of separate datasets. The DISS has been successfully tested in a geographically distributed heterogeneous environment using high performance networks such as the NASA NREN, NSF vBNS, Internet2. The rest of the paper describes the DISS methods, protocols, performance and tradeoffs for remote data access.

2. DISS ENVIRONMENT FOR REMOTE ACCESS

The DISS Environment is currently being used to analyze remote sensing multispectral data sets from NOAA/AVHRR, GOES/VISSR, Meteosat, GMS, FY-2,

Reprint Seventeenth Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology, 81st AMS, Jan. 14-19, 2001, Albuquerque, NM, American Meteorological Society, 2001, pp. 399-403.

ERS-1/ATSR, Nimbus-7/TOMS, Landsat TM, DMSP/SSMI, TRMM/TMI, TRMM/PR Earth remote sensing satellite instruments [2][3]. Products that have been derived from NOAA/AVHRR and Landsat TM data, using the DISS, include color composites, vegetation indices, perspective and stereo views [3]. The products can be combined, frames selected within a cell, frames animated interactively and image formulas evaluated. The spreadsheet-based visual interface (see Figure 1) has been found to be extremely intuitive and highly productive since it reduces the need for a user to explicitly deal with input/output based programming. The resulting DISS Environment provides the scientist with an effective and powerful visualization tool for concentrating on algorithm development and data exploration. The DISS has been used to successfully prototype and evaluate the performance and utility of using high speed wide area networks such as NREN and vBNS for visualization of large remote sensing datasets (Figure 4).

The DISS Environment uses a 3-D arrangement of elements to accommodate organizing large datasets in the time or channel dimension. The object oriented view is shown in Figure 2. Note that in the current version of the DISS the collaborative or remote access component is only handled at the Frame level and there is no Book object (several separate sheets comprise a Book); these are new features that are being added. Cells contain a stack of frames so each page or layer of the spreadsheet can manipulate an independent group of data with the added flexibility that relationships between spreadsheet cells and layers can be directly specified. Each frame of the image spreadsheet contains one or more multidimensional data sets to be visualized as shown in Figure 3. Visualization data sets include raw and processed satellite imagery, graphical (vector) data, surface and terrain models, and three-dimensional volumes. The DISS supports distributed network access to datasets and computation capabilities.

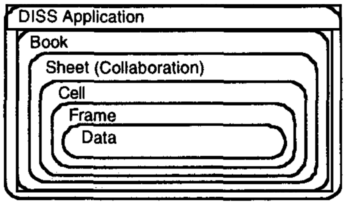


Fig. 2 Hierarchy of spreadsheet objects and terminology.

2.1 Remote and Compressed File Access

DISS uses the common Internet protocols FTP (file transfer protocol) and HTTP (hypertext-transfer protocol) for remote data access. Virtually any web-accessible machine running an FTP or HTTP server can act as a data server for DISS since these are widely supported Internet protocols. Remote data is accessed within DISS using the familiar universal resource locator (URL) syntax within a formula or read panel in place of a file name. For example,

```
Read[filename->
"ftp://rsd.gsfc.nasa.gov/pub/Weather/
GOES-8/jpg/vis/4km/0latest.jpg"]
```

is a generic read formula where `filename` is a URL. The Read[] operator installs the data into a frame exactly as though the data were local, but instead the data is retrieved transparently via FTP from a remote machine. HTTP can be used identical to FTP as in the example. FTP is currently restricted to retrieving data from servers allowing anonymous FTP. Enabling anonymous access avoids the inconveniences associated with user authentication. For instance, if a user were simultaneously accessing data from 10 distinct FTP sites, and each site required authentication, the user would have to supply 10 username/password pairs. Security can still be tightly maintained using TCP wrappers to monitor RPC requests and restricting `portmap` services using the `/etc/hosts.allow` file to control client access.

Remote access to data within the DISS framework benefits the user by enabling convenient sharing of results and visualizations for collaboration, and by allowing transparent access to the latest near-realtime data. Users can generate a DISS header file in which all data references are remote and specified by URLs (FTP or HTTP). Sharing these results with other users, particularly users not local to the data, is easily done by sharing the text-only DISS header file. When all data references in the header are remote, then data is exchanged only between the user and the remote sensing archive since the retrieval uses the embedded URLs. Similarly, a DISS header can be constructed to use the latest near-realtime data provided the site serving the data has made provisions for providing the latest data through a single named source. For instance, in the Read[] example the file named 0latest.jpg would probably exist as a named pointer to the most recent data. Thus, any formulas referencing this named pointer would always use the latest data available at that Web site.

2.2 Data Compression, File or Data Caching, and Formulas

In addition to providing network data access via HTTP and anonymous FTP, DISS provides automatic decoding of gzip or Unix compressed file formats [7]. The decoding is transparent to the user allowing direct access to compressed data within any read formula operator. Here a digital orthophoto quarter quadrangle (DOQQ) is read from the Missouri ICREST site:

```
Read_band[filename->
"http://davis.geog.missouri.edu/icrest/
data/doqq/o3709354nws.bip.gz",
header_size->24752, src_xdim->6188,
src_ydim-> 7563, flip->TRUE]
```

Both gzip and Unix compress formats are useful because they provide lossless compression of any data type. Gzip support is provided through the zlib library [7]. Both remote access and compressed file reading can be used in conjunction with remote access for efficient network transport. Data compression decreases data size thus decreasing access time, while maintaining data fidelity. Scalability to extremely large datasets such as the US Landsat mosaic require data reorganization, progressive transmission, error resilience in addition to compression.

In a typical DISS session, a single file will usually be accessed many times. For instance, a multispectral dataset such as a multiband Landsat product may be accessed several times by different operators. To prevent retrieving or decompressing the same file for each frame access, a per-session cache is kept of all remote and compressed file accesses. Internally DISS maintains a mapping of URLs and compressed filenames to local files in temporary storage that may be on a network attached storage device as shown in Figure 4. After a file has been retrieved or decompressed once, each subsequent access is made from the cache. Pseudo code for the caching of remote or compressed files is shown below:

```
cache_file( filename ) {
if( is_remote(filename) OR
    is_compressed(filename) ) {
  if( in_cache(filename) )
    return( cached_filename )
  else {
    if( is_remote(filename) )
      retrieve file
      put file incache(cached_filename)
    if( is_compressed(filename) )
      decompress file
      put file in cache
    return( cached_filename ) } }
else
  return( filename ) }
```

Currently, remote and compressed file reading is supported for the following operators: Read[], Read_HDF[], Read_HDFL2[], Read_multiband[], Read_band[], Read_MPEG[], Read_VSLCCA[], Volume[]. The cached file may reside on a local filesystem or on a network attached storage device close to the end user for faster network access (Fig. 4). All of these operators implement the cached remote and compressed access using the method described above. Using this single point of entrance for caching provides an extensible framework for adding other data types that require caching. Should another form of data ingest require caching support, merely changing the above caching function would add that functionality to all the operators currently supporting caching. Through these operators any (compressed) file can be read remotely including raw format, MPEG, VSLCCA, Vis5D, PNM, SGI, JPEG, GIF, and TIFF.

2.3 Disk Cache vs. Memory Pool

While all remote and compressed accesses are cached to disk, two operations cache to memory: Read_VSLCCA[] and Read_MPEG[]. Whenever a VSLCCA or MPEG movie sequence is read, it is placed into an in-memory pool (similar to the colormap pool). Any references to a frame in that sequence actually points into the pool. Any additional references to the movie or video file resolves to the frames already in memory. Only when no references to a video sequence remain, is the memory freed and removed from the pool.

When a movie sequence is retrieved remotely or compressed, it is both cached to disk and cached in memory. First the sequence file is retrieved or decompressed and saved locally to disk. Then the sequence is read into memory and it is placed in the movie pool. At some point when there are no longer any references to the sequence, it is removed from the memory pool, but the local disk copy still remains. Any further references to the sequence will cause it to be read from the local disk cache and placed into the memory pool. This supports better interactive access to movie sequences than video streaming (frames are discarded after viewing) but at the expense of additional storage requirement close to user.

3. EXPERIMENTS

The DISS was selected, as a testbed application to demonstrate the potential of the US Next Generation Internet (NGI), the NASA Research and Education Network (NREN), the National Science Foundation Very high speed Backbone Network Service (vBNS), and Internet 2. Early NGI prototyped applications exploiting the capability of a 1000 times faster Internet than the one in place today. The DISS environment was developed to participate in these prototyping efforts [8], (http://apps.internet2.edu/demos98/diss.htm).

Early demonstrations of the DISS were completed in June 1997 at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, as part of the Global Observation Information Network (GOIN) workshop, a US-Japan bilateral demonstration of information networks. The user client workstation was a remote visualization terminal at NCAR to access data remotely at GSFC, about 1500 miles away, from a high performance storage system via a high performance WAN. The WAN connectivity included a NASA NREN connection at OC-3 (155 mbs) rates with connectivity to NCAR via a vBNS connection at San Diego Supercomputing Center at OC-12 (622 mbs) rates and Fiber Distributed Data Interface (FDDI) network interfaces for the server and client. The slowest link was a DS-3 line through ESNet. In September 1997 at NASA Ames Research Center in Moffett Field, California, DISS was demonstrated using Asynchronous Transfer Mode (ATM) end-to-end for the first time. An OC-3 connection using NREN's network via the Sprint ATM cloud was established between GSFC and Ames [8] Application performance of the DISS dramatically improved by a factor of ten using ATM end-to-end but only a fraction of the total available OC-3 bandwidth was harnessed. Significant increase in bandwidth utilization (64 to 100 mbs) was achieved in August 2000 and August 1999 at the latest NREN workshops and included using SGI's Vizserver for remote framebuffer display, Bulk Data Service (efficient NFS) for remote file access, multithreaded image spreadsheet for improved i/o performance, and tuning TCP/UDP window sizes [8].

For realtime direct broadcast, the data rate for GOES-8 and GOES-9 receivers is 2 GB/hour or 4 Mb/s. The Terra EOS direct broadcast (X-band, 8.2125 GHz) data rate for MODIS will be 5.9 GB/hour or 13.125 Mb/s. The processed data rate will typically be 16 times higher or 384 Mb/s for GOES (4 new floating point 32-bit parameters calculated for each 8-bit pixel). Both the raw and processed data streams may need to be distributed to several sites for processing and visualization and be continuously available. Jitter and bulk data transfer rates are the key issues and are relevant only if the available bandwidth is sufficient. Using network storage devices the data streams could be buffered for several hours to reduce the bandwidth requirement. The DISS is used to visualize direct broadcast data in near realtime.

The collaborative component of the DISS for access to geospatial datasets several modes are feasible: (i) the user, the data, or both are remote from the server, (ii) the visualization will be precomputed ("canned") or observed in real time ("interactive"), (iii) the portions of the total visualization workload are partitioned between the client and server systems. In each mode the primary data exchange will be a combination of video, distributed graphics language (DGL), or distributed files. This affects network QoS requirements (both the total data volume transferred and the transfer rate required) and client system visualization and processing requirements. In the video mode, the high performance server does the rendering processing, generating the video display, and sending it to the client over the network as video. A rate of 5 frames/sec for is needed for interactivity. Using a 1920 x 1035 wide 24" monitor display with 48 bits per pixel requires a visualization video bandwidth of 480 Mb/s. Image compression reduces bandwidth requirements (i.e. the SGI Vizserver). For mission critical applications such as Air Force weather simulations the acceptable jitter would be less than 5 %. Latency needs to be less than 0.5 seconds for interactivity and update of user mouse and keyboard inputs. In the DGL or vector mode, the client sends lower bandwidth geometry or polygon information along with user interactions to the server, and the server interprets these user commands and processes the datasets into geometry or vector data which are sent over the network to the client using geometry compression for efficient bandwidth use. Performance evaluation is continuing.

4. SUMMARY

The utility of the DISS as an interactive scientific visualization tool has been demonstrated for efficient quality control of large datasets, organizing a large volume of multidimensional time varying multisource data, understanding interrelationships between complex datasets, rapid prototyping of algorithms constructed by algebraic operations, or comparing model data with observed data. The DISS has been successfully tested for large data set visualization over high performance networks using FTP and HTTP servers and incorporating lossless data compression. The DISS has also been used for collaborative visualization of realtime data from several geographically dispersed sites. Network-based disk and memory caching significantly improves performance and interactivity for visualization.

ACKNOWLEDGEMENTS This work was supported in part by NASA/GSFC Award NAG-5-3900, NASA MTPE Award NAG-5-6968, NAG-5-6283, NAG-13-99014, and the NSF vBNS Award 9720668. Andy Germain (NASA GSFC) and Dick DesJardins (NASA Ames) were instrumental in the NREN demonstrations. John Harwell assisted with remote access, and Shelley Zhuang added MPEG and VSLCCA video support.

REFERENCES

1. Dodge, J., "The Earth Observing System Direct broadcast and receiving stations", *Earth Observation Magazine*, Apr 1999.
2. K. Palaniappan, A.F. Hasler, M. Manyin, "Exploratory analysis of satellite data using the Interactive Image Spreadsheet (IISS) environment", *Ninth Intl. AMS Conf. Interactive Information and Processing Systems*, American Meteorological Society, 1993, pp. 145-152.
3. Hasler, A. F., K. Palaniappan, M. Manyin, and J. Dodge, 1994: A High Performance Interactive Image SpreadSheet (IISS), *Computers in Physics*, Vol. 8, No. 3, 1994. pp. 325-342
4. Palaniappan, K., J. Fraser, 2001: Multiresolution tiling for interactive viewing of large datasets, *17th Int. Conf. on Interactive Information and Processing Systems (IIPS)*, Albuquerque, NM, Amer. Meteor. Soc.
5. Chi, E. H., J. Riedl, P. Barry, J. Konstan, 1998: "Principles for information visualization spreadsheets", *IEEE Computer Graphics & Applications*, pp. 30-38.
6. Szuszczewicz, E.P. and J.H. Bredekamp, *Visualization Techniques in Space and Atmospheric Sciences*, NASA SP-519, Washington, DC., 1995.
7. Gzip and zip libraries. ftp://ftp.cdrom.com/pub/infozip/zlib/zlib.tar.gz, http://artpacks.acid.org/pub/infozip/zlib/
8. "Distributed Image SpreadSheet for Earth and Space Science Data", *NREN HPCC Workshop V: Gigabit Networking*, Aug. 14-16, 2000, *Workshop IV: Bridging the Gap*, Aug. 10 -11, 1999, and *Workshop II: Tomorrow's Networking Applications Today*, Sept. 15-17, 1997, NASA Ames Research, Mountain View, CA. http://www.nren.nasa.gov/eos_distribution.html, http://www.nren.nasa.gov/workshop4.html

Reprint Seventeenth Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology, 81st AMS, Jan. 14-19, 2001, Albuquerque, NM, American Meteorological Society, 2001, pp. 399-403.

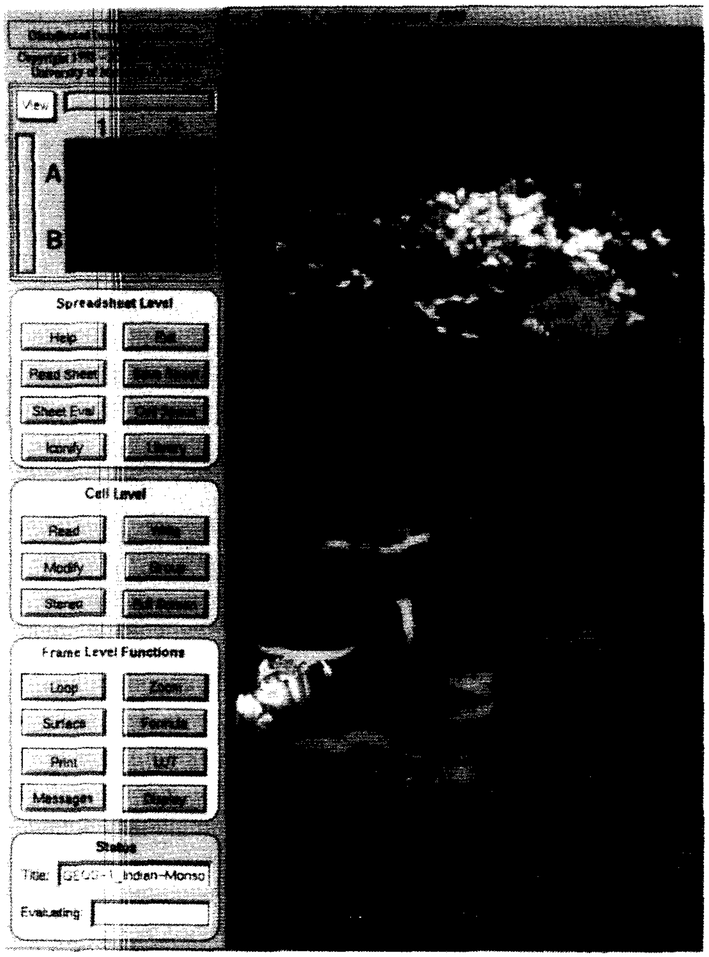


Figure 1 Visualization of DAO GEOS-1 assimilated data using the extended DISS environment with Vis5D capabilities for studying the onset and extent of the 1988 Indian Monsoon. Cell A1 and B1 show selected variables using isosurface and trajectory visualization.

Figure 3 Organization of cell data includes a list of frames contained in the cell (circular doubly-linked list for framestack), pointer to current frame and each frame containing original and display data information.

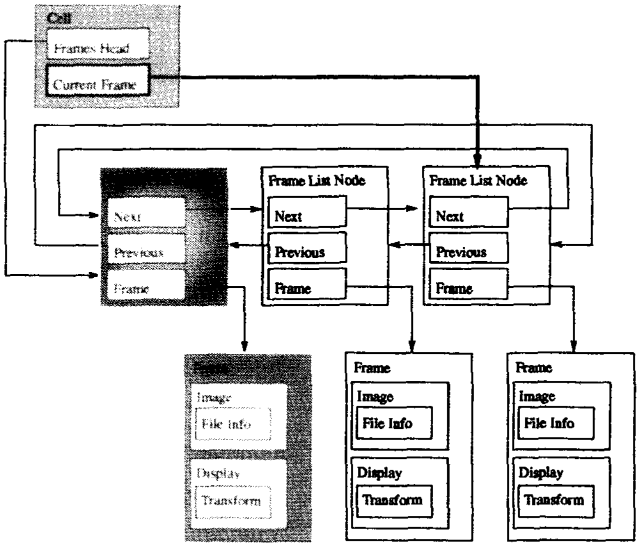
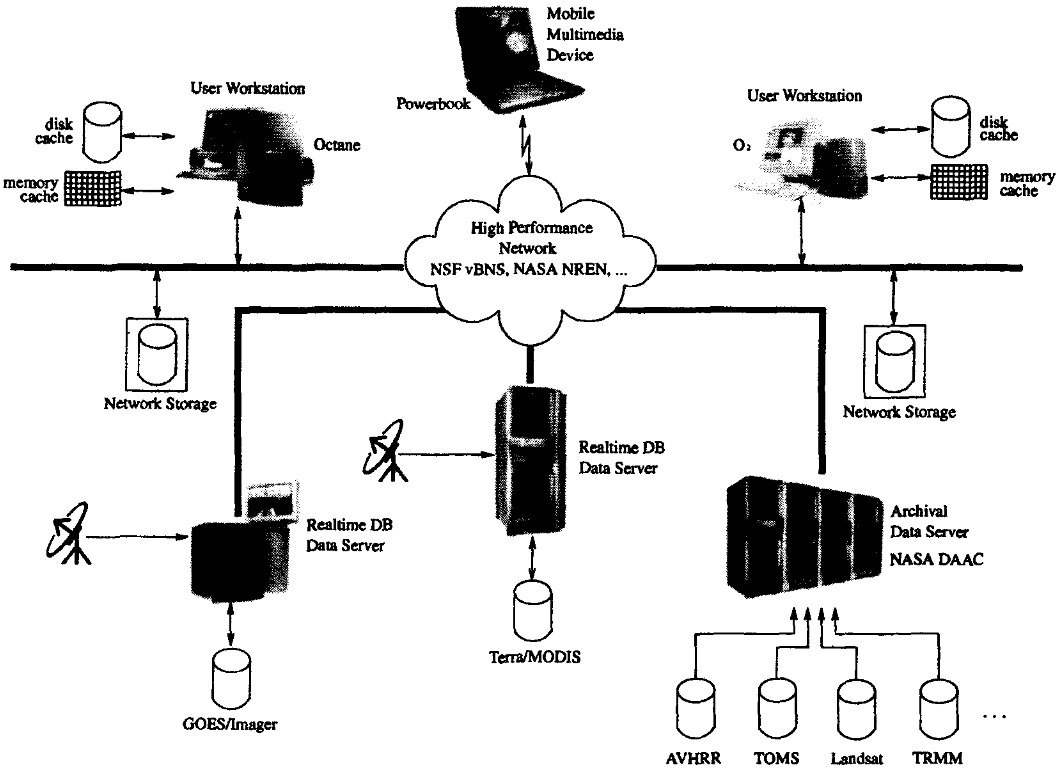


Figure 4 Distributed archives of large remote sensing imagery accessed via the Internet using wireline and wireless computers. Storage nodes may be attached to servers or directly on network.



Reprint Seventeenth Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology, 81st AMS, Jan. 14-19, 2001, Albuquerque, NM, American Meteorological Society, 2001, pp. 399-403.

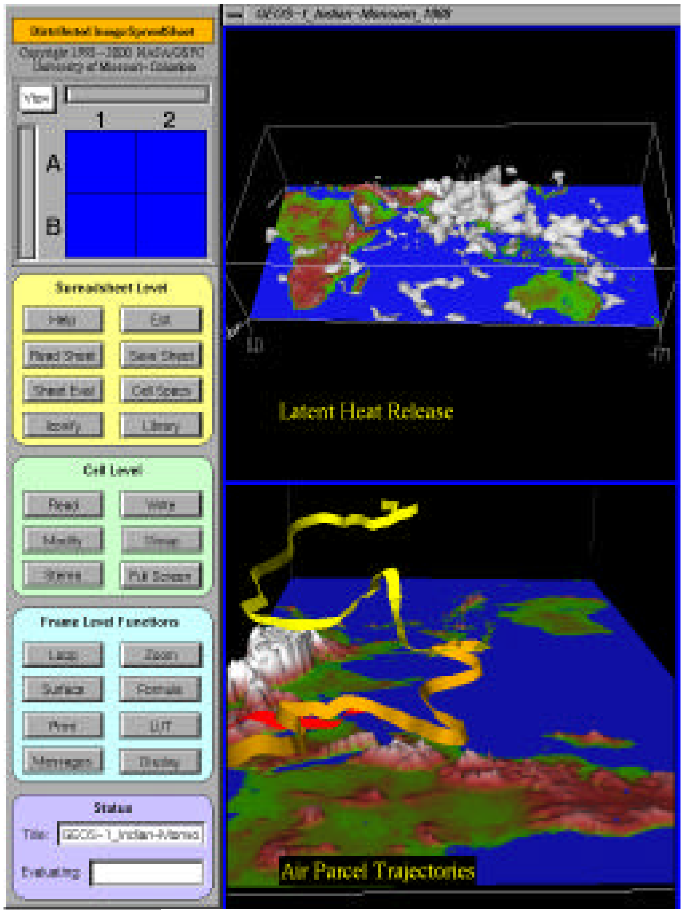


Figure 1 Visualization of DAO GEOS-1 assimilated data using the extended DISS environment with Vis5D capabilities for studying the onset and extent of the 1988 Indian Monsoon. Cell A1 and B1 show selected variables using isosurface and trajectory visualization.

Figure 3 Organization of cell data includes a list of frames contained in the cell (circular doubly-linked list for framestack), pointer to current frame and each frame containing original and display data information.

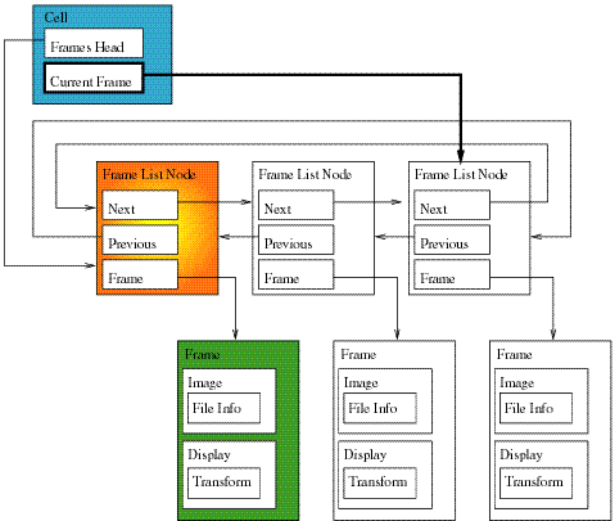
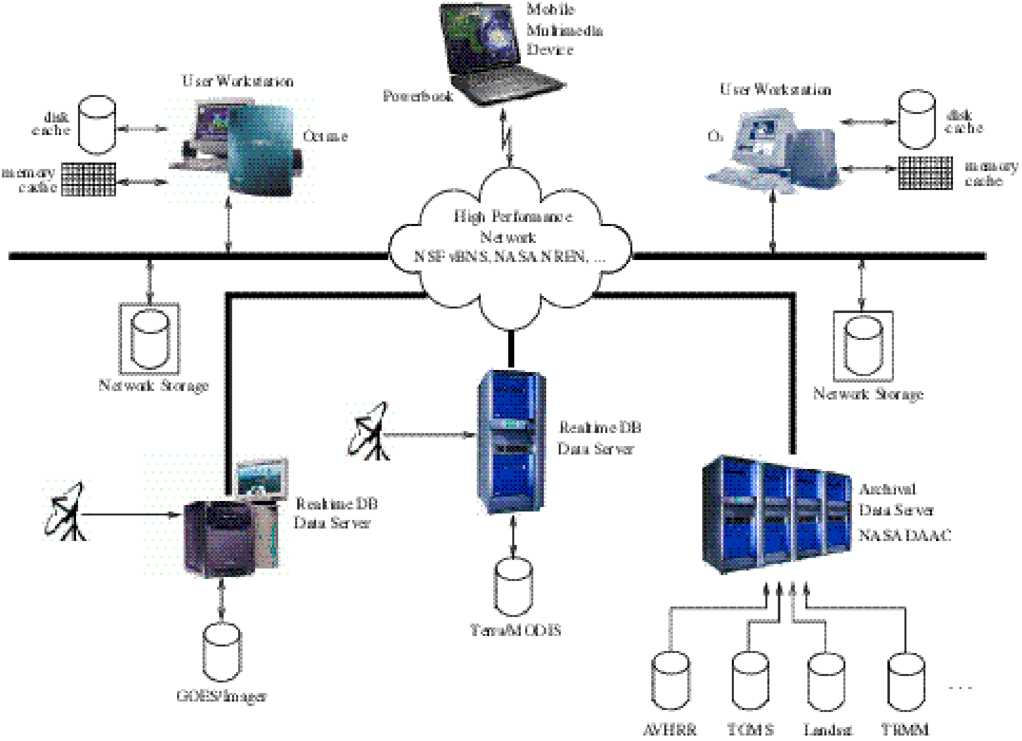


Figure 4 Distributed archives of large remote sensing imagery accessed via the Internet using wireline and wireless computers. Storage nodes may be attached to servers or directly on network.


Mobile Multimedia Device
Powerbook
User Workstation
disk cache
memory cache
Octane
User Workstation
O2
disk cache
memory cache
High Performance Network
NSF vBNS, NASA NREN, ...
Network Storage
Network Storage
Realtime DB Data Server
Realtime DB Data Server
Archival Data Server
NASA DAAC
Terra/MODIS
GOES/Imager
AVHRR
TOMS
Landsat
TRMM
. . .