Astronomical Data Analysis Software and Systems VI
ASP Conference Series, Vol. 125, 1997
Gareth Hunt and H. E. Payne, eds.

The Evolution of the HST Archive

J. J. Travisano,[1] J. G. Richon

Space Telescope Science Institute, Baltimore, MD 21218

Abstract. The Hubble Space Telescope Archive has been in operation since the launch of HST in April 1990. There have been two generations of archive systems (DMF and DADS) and user interfaces (STARCAT and StarView). In this paper, we describe recent projects and future directions in the continued evolution of the HST Archive.

1. Components of the HST Archive at STScI

The Data Archive and Distribution System—DADS—comprises the core of the Archive (Pollizzi 1995). It stores all HST data onto optical disks. All datasets are cataloged into a Sybase database from the science header keywords and dataset/file information. DADS runs on Digital Alpha and VAX platforms.

StarView is the primary user interface to the HST Archive (Williams 1993). It is used to query the DADS databases, preview selected images and spectra, and submit retrieval requests for data. StarView is supported in CRT and X/Motif modes on Sun and Digital platforms.

World Wide Web (WWW) access is now available. Additional information on this and the HST Archive in general is available through the STScI WWW pages.[2]

2. DADS

2.1. Recent Enhancements

Blackboard Ingest and Distribution The Blackboard Ingest and Distribution projects replaced internal memory lists and mailbox messages with a blackboard trigger mechanism. The OpenVMS cluster-wide filesystem acts as a blackboard, upon which files are created with names indicating work to be done. Individual processes look to this global area for work, claim it, and perform it.

The idea of using blackboards was adopted from the STScI OPUS system (Rose 1995). This architecture gives DADS increased flexibility in that processes can run on different hosts in the cluster. Multiple copies of some programs can be instantiated, improving scalability in dealing with peak loading conditions.

---

[1] Computer Sciences Corporation

[2] http://www.stsci.edu/

Aptec Replacement & Port to Alpha The original DADS system delivered to STScI consisted of a cluster of VAX systems with an Aptec I/O Processor. The Aptec was used for performing I/O to the Sony optical disk drives and for doing data format conversion from STSDAS GEIS format to FITS. This specialized device was not well suited for this application, was unreliable, put a high kernel interrupt load on the host VAX, and was difficult to maintain (both hardware and software).

A project was started in 1995 to replace the Aptec. This project included porting most of DADS to OpenVMS Alpha, replacing the optical disk I/O routines to use native OpenVMS and SCSI services, and using an IRAF/STSDAS task (stwfits) for FITS conversion. The Aptec staging disks were replaced with a RAID array that is local to the Alpha host and cluster-accessible. The resulting system is faster and more reliable, with reduced maintenance costs.

2.2. Work In Progress

HST Servicing Mission 2—February 1997 Significant efforts are underway to support the second HST Servicing Mission. Two new instruments will be installed in the telescope: the Space Telescope Imaging Spectrograph (STIS) and the Near Infrared Camera and Multi-Object Spectrometer (NICMOS).

DADS is being modified to ingest the new FITS keywords in the STIS and NICMOS data files. FITS extensions are used, combining images and tables into single files. There is the new concept of *associations* of datasets, where individual exposures, products from processing these exposures, and ancillary files are grouped together into a collected set.

STIS and NICMOS will produce significantly more data than the current instruments. This and the fact that more observations can be done in parallel will result in three to six times as much data going into the Archive.

Compression One of the biggest problems facing DADS is that the optical disk jukeboxes will soon be filled to capacity, sometime in the spring of 1997. Part of the solution is to compress the files on optical disk. This increases the overall near-line capacity of the Archive, as discussed in the Hubble Archive Re-Engineering Project (HARP) (Hanisch 1997).

2.3. Planned/Future

Archive Engine Redesign The DADS Archive Engine controls the I/O to and from the optical disks and manages the jukeboxes. We are planning to redesign this subsystem to improve overall efficiency, give more control to the operations staff for scheduling/directing requests, allow better segregation of data for near-line vs. off-line storage, and support a magnetic disk cache (Comeau 1997) to reduce contention as well as wear and tear on the jukebox and drive hardware.

Multiple Media DADS uses Sony 12-inch WORM optical disks (6 GB per platter) with Cygnet jukeboxes (four of them holding 131 platters each). We are investigating other archival media and plan to modify DADS to support multiple media. In particular, DVD-ROM looks promising for its storage capacity, lower costs, and the availability of jukeboxes of 500 or more disks. For distribution media, we are planning to support 4mm DAT and CD-ROM in 1997.

System Migration There are more applications to port to the Alpha to complete the migration from VAX. We are considering alternate operating systems for all or parts of DADS. The Alpha hardware can run OpenVMS, Digital UNIX, and Windows NT. Linux is also available for some Alpha systems.

On-The-Fly Recalibration The Space Telescope European Coordinating Facility (ST-ECF) and Canadian Astronomy Data Centre (CADC) have copies of HST data. Both sites support on-the-fly recalibration (Crabtree 1996). We discussed this idea in HARP as a way to reduce the amount of archived data. Due to user support needs, it is not feasible to store only raw science data. However, on-the-fly recalibration will be a useful service for the HST Archive at STScI.

Non-HST Data The Archive contains non-HST data, such as the VLA FIRST Survey (Faint Images of the Radio Sky at Twenty-cm). More than 5000 VLA FIRST datasets have been archived, each catalogued by over 20 parameters (RA, Dec, frequency, etc.). We continue to consider the potential needs of non-HST data, positioning the Archive for new data from HST and other observatories.

HST Servicing Mission 3—1999 The third HST Servicing Mission will add the Advanced Camera for Surveys (ACS). There will be an increase in data volume, estimated to be double the volume after Servicing Mission 2.

3. StarView & The World Wide Web

3.1. Recent Enhancements

World Wide Web Interface After a few prototypes, including one presented at ADASS IV (Travisano 1995), a Web interface to the HST Archive is now available. Users can search for observations and proposal information, preview images and spectra of public data, and retrieve datasets to an FTP area on archive.stsci.edu. The Web interface uses HTML and CGI scripts written mostly in Perl, along with some back-end StarView utilities.

HST Field of View Overlay on The Digitized Sky Survey From StarView or the Web interface, a user can ask for any section of the sky from the Digitized Sky Survey (DSS). The user can now request the HST field of view, containing the primary instrument apertures, as a graphics overlay on the DSS image.

Automatic Retrieval of Calibration Reference Files Retrieving the reference files necessary to recalibrate science data formerly required multiple StarView screens. Now the user can select whether the original and/or best reference files are to be retrieved along with the selected science datasets. The StarView retrieve utility (also used by the Web interface) will look up the appropriate files and add them to the retrieve request sent to DADS.

Improved String/Text Searching Until recently, the only string searching capabilities supported were an explicit string, a wildcarded string, and an **or** list of these two. StarView Release 4.5 includes support for the **and** operator and the **not** qualifier. This improves the ability to find what you are looking for, and exclude what you are not, especially when searching on proposal abstracts.

3.2. Work In Progress

HST Servicing Mission 2—February 1997 StarView and the Web interface are being modified to support STIS and NICMOS. This work includes data dictionary updates, new instrument screens, the HST field of view overlays, screens for associations, additional retrieval options, etc.

Improved Coordinate System Support StarView Release 4.5 includes improvements to display the coordinate system and equinox on query results screens. These displays are based on user selections of how to display coordinate values from the catalog, which are stored in Equatorial J2000.

3.3. Planned/Future

Continued Improvements to Web Interface The Web interface was released in September 1996. We are monitoring usage and feedback to plan new features. Supporting Internet delivery of data will require safeguarding the account information sent from Web browsers to the server.

HST Servicing Mission 3—1999 For the third servicing mission, the StarView and Web interfaces will again be updated. Instrument screens for ACS, the field of view overlays, etc., will be added in support of the mission.

3.4. Concluding Remarks

We continue to improve the HST Archive systems—DADS, StarView, and the new Web interface. The HST Servicing Missions in 1997 and 1999 require much work, as do the continued efficiency improvements to DADS necessary to support operations on a daily basis for the 1000 or so registered HST Archive users.

References

Comeau, T., & Park, V. 1997, this volume, 318

Crabtree, D., Durand, D., Gaudet, S., & Hill, N. 1996, in ASP Conf. Ser., Vol. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes (San Francisco: ASP), 505

Hanisch, R., et al. 1997, this volume, 294

Pollizzi, J. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 162

Rose, J., et al. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 429

Travisano, J. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 80

Williams, J. 1993, in ASP Conf. Ser., Vol. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes (San Francisco: ASP), 100