Astronomical Data Analysis Software and Systems IV
ASP Conference Series, Vol. 77, 1995
R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.

The Hubble Space Telescope Data Archive

K. D. Borne, S. A. Baum, A. Fruchter, and K. S. Long

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

Abstract. We describe recent progress at the Space Telescope Science Institute[1] in the development of the *HST* Data Archive[2] and of StarView[3], the user interface to the *HST* archive system. Access to this system is available to users through one of our *HST* archive host machines[4] or through the local use of the distributed version of StarView[5].

1. Introduction

The *Hubble Space Telescope* (*HST*) Data Archive has been open for archival research since early in 1993. It contains all of the observational data, calibration files, and related catalog information produced by the *HST* since its deployment. The archive currently comprises 1.2 TB of data, of which 60% are science data, more than 80% of that being publicly available. A new archive engine, ST-DADS (Space Telescope Data Archive and Distribution Service), is now in use. ST-DADS is designed to maintain 3 TB, or about 8 years' worth, of *HST* data on-line in four optical disk jukeboxes. ST-DADS stores all important science data files internally in FITS-compatible formats and will eventually be able to deliver data directly to remote workstations. The archive also includes an on-line catalog, which can be browsed by any user via a connection to one of two archive host machines (VMS and Unix). StarView is the user interface to the *HST* catalog at ST ScI, and it operates on Sun Unix and DEC VMS machines. There are two versions of the interface available: one is CRT-based (VT100-compatible) and the other is X-based. The latter includes a data-previewing capability and all of the point-and-click features typical of Motif-based graphical user interfaces. A distributed version of StarView allows users to run StarView locally on their Sun machines. It has the same functionality as the version of StarView running on the ST ScI archive host Unix machine and reduces network loading by creating X-windows locally and by accessing ST ScI machines only to query the database.

[1] http://marvel.stsci.edu/top.html

[2] http://www.stsci.edu/archive-html/archive.html

[3] http://www.stsci.edu/archive-html/starview.html

[4] http://www.stsci.edu/archive-html/host_mach.html

[5] http://www.stsci.edu/archive-html/distributed_sv.html

2. What was DMF? What is DADS?

The Data Management Facility (DMF) was the prototype archive for *HST* data from launch until 1994 September 21. The system was developed by ST ScI, ST-ECF, and CADC, and it had a 170 GB on-line data capacity (Long et al. 1993).

The Space Telescope Data Archive and Distribution Service (ST-DADS) is the permanent archive for *HST* data. It has been operational since 1993 December (the time of the First Servicing Mission to the *HST*). The system was developed by Loral Aerosys and ST ScI for NASA, and it has a 3.4 TB on-line data capacity, which corresponds roughly to 8 years' worth of *HST* data.

3. DMF-to-DADS Transition

The transition to full archive operations using ST-DADS has involved a number of steps. First, a new user interface (StarView) was developed at the ST ScI to replace the previous interface (Starcat). Next, all *HST* data obtained prior to 1993 December were transferred from DMF into ST-DADS (900 GB total). While the DMF-to-DADS data transfer was taking place, DMF and ST-DADS were operated in parallel for an extended period (from 1993 December through 1994 September). As part of the data transfer project, the completeness and integrity of the ST-DADS data and catalog were verified. We determined that the first pass through the data (in the DMF-to-DADS data transfer) was 99.5% complete and 99.99% accurate (i.e., roughly one file in 10,000 had some kind of error in ST-DADS). The cleanup of these data got underway in 1994 August and is continuing.

As part of the transition, a retrieval mechanism for the data in ST-DADS was provided to users through StarView, and a bulk copying mechanism was also developed for the generation of ST-DADS optical disk platters for use at other *HST* data archive sites: currently, the ST-ECF and CADC.

4. Contents of the HST Archive

The *HST* Data Archive consists of observational data (the *HST* Archive) and derived data (the *HST* Catalog). We plan in the near future to add user-derived *HST* data and supplemental non-*HST* data which will enhance the scientific use of the *HST* data (e.g., the Guide Star plate scans, *HUT* data, VLA FIRST survey data, and other astronomical catalogs). All of the data stored in ST-DADS are in either FITS or binary format (e.g., GEIS and ASCII files are converted to FITS before archiving).

The *HST* Catalog consists of $\sim$50 tables, including a science table, instrument tables, observation tables, target tables, engineering tables, and our internal archive bookkeeping tables. All of the scientifically interesting fields in these tables are accessible through the StarView user interface, either through fixed predefined forms or through the custom query mechanism in StarView (see below). As part of its extensive on-line help, StarView provides a description of the fields that it uses from the various database tables.

The current *HST* Archive data volume is 1.2 TB, comprising approximately 1.7 million individual files. The monthly ingest rate (for new data) into ST-

DADS is ~30–50 GB, corresponding to a yearly ingest rate of ~300–500 GB. The science data volume in ST-DADS is nearly 0.7 TB (60% of the total), comprising 65,000 science observation datasets. Approximately 83% of these science datasets are public. Since its launch, the *HST* has obtained data on ~5000 distinct astronomical targets. The number of science datasets in the *HST* Archive, by instrument, is: FOC: 4900; GHRS: 14,700; WFPC: 15,800; FOS: 14,000; HSP: 5200; and WFPC2: 10,500.

5. StarView—The HST Archive User Interface

StarView is the user interface to the *HST* Archive. It supports VT-100 compatible CRTs and X-windows, for both VMS and SunOS. A distributed version is also available for SunOS—this version uses network bandwidth only for database calls, not for any of the intensive X-window interactions.

User interactions with the StarView interface are through a variety of data screens. A "Quick Search" screen is provided (with direct access from the StarView "Welcome" screen), which can be used to initiate the most common type of informational searches of the *HST* Catalog—it is the most basic archival search screen, including only a few simple fields by which one can constrain a catalog search. After initiating a quick search of the catalog, the results are returned on a separate screen that contains numerous catalog entries related to each observation. In addition to the "Quick Search" screen, there are separate StarView screens for exposure parameters, target properties, proposal information, planned exposures, instrument parameters, calibration files, and engineering files. Each screen contains a wide selection of user-qualifiable fields connected to the archive database. For ease of maintenance and development of the user data screens, the screen definitions, formats, and contents are all kept "outside" of StarView software. StarView screens are used to search for specific datasets, to view the results of a query, to preview the data, to mark datasets for retrieval, and to submit the retrieval request. The results of a query may be viewed one record at a time (in portrait form) or many records at a time (in tabular form).

The principal science observation catalog information is collected into a single "science" table in the *HST* Catalog (to minimize database joins). StarView uses that table for most of its basic exposure search screens. Target searches are simplified in StarView by allowing the user to get target coordinates from either the NED or SIMBAD databases using a user-specified target name. An additional feature of StarView allows a user-provided target search list of RA and Dec pairs to be cross-correlated with the contents of the *HST* Catalog. This is particularly helpful when planning *HST* observing proposals or *HST* archive research proposals.

One of the more powerful features of StarView is a "custom query" interface, which allows the user to define (i.e., customize) their own query of the archive database, dynamically selecting any database fields of their choice to be included in the query. A corollary aspect of this feature is an on-line editor that allows the user to compose a personalized SQL query and to submit that query directly to the database server.

Of particular interest to researchers is a public data-previewing capability available through XStarView. Compressed images and spectra of all public *HST* data are available at the push of a button, allowing the user to see the stored image or spectrum prior to submitting a data retrieval request. (The compressed data are provided courtesy of the CADC.)

StarView provides users with a variety of on-line help, including "Strategy" help, which is available at the push of a button, for each StarView screen.

6. Access to the HST Archive

Two archive host machines are made available for external user logins (for VMS: *stdata.stsci.edu*, and for Unix: *stdatu.stsci.edu*). The username is `guest` and the password is `archive`. Both CRT-StarView and X-StarView are available on the host machines.

The number of logins to the ST ScI Archive hosts is currently 2500–3000 per month. There are ∼800 individual archive user accounts (∼600 of which are for non-ST ScI users), from which 15–20 GB of *HST* data are retrieved each month (∼30% by non-ST ScI users). Note: one must be a registered user of the ST ScI archive in order to retrieve data, but not to search the catalog or to preview public data.

Science data are retrieved for users in FITS format. Currently, retrievals are only supported on the two archive host machines: *stdata.stsci.edu* and *stdatu.stsci.edu*. In the future, ST-DADS will be set up to deliver data (via ftp) to a user-specified destination anywhere on the Internet.

To support users of the *HST* Archive, several forms of documentation are available: the *HST Archive Primer*, the *HST Archive Manual*, and the *HST Data Handbook*. In addition, archive hotseat support is available, either by sending e-mail to *archive@stsci.edu* or by phoning (410) 338–4547.

7. Current Status

As of this writing, StarView has been in successful operation for over a year. ST-DADS has archived all *HST* science, calibration, astrometry, and engineering data since 1993 December (i.e., since the First *HST* Servicing Mission). All of the science and calibration data stored in the DMF since launch have been transferred into ST-DADS. The transfer of all astrometry and engineering data from DMF into ST-DADS will be completed by 1994 December. Finally, archiving of new *HST* data to DMF was turned off in 1994 September. As a consequence of these developments, all access to the *HST* Data Archive has been through ST-DADS since 1994 October. The *HST* Data Archive is thereby operational and ready to support the *HST* data needs of the astronomical research community.

References

Long, K. S., Baum, S. A., Borne, K. D., & Swade, D. 1993, in Astronomical Data Analysis Software and Systems III, ASP Conf. Ser., Vol. 61, eds.

D. R. Crabtree, R. J. Hanisch, & J. Barnes (San Francisco, ASP), p. 151longk