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 Data Archive: Opening the Treasures of the HST to the Community

J. A. Pollizzi, III

Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218

Abstract. This paper briefly summarizes the history and present state of the archive systems for the *Hubble Space Telescope*.

1. DMF as the Hubble Data Archive - Launch through Present

The data to be collected by the *Hubble Space Telescope* (*HST*) promised to be a significant new resource to the astronomy community, even before *HST*'s launch. In plans made to capture data from *HST*, the capacity to accommodate the expected data volume and the astronomers' accessibility to the data were both considered important.

Prior to launch, the Space Telescope Science Institute (ST ScI) initiated a prototype program to explore technologies that could be used in archiving *HST* data. The prototype, called the Interim Data Management Facility (hereafter, DMF), was developed, collaboratively with the Space Telescope European Coordinating Facility (ST-ECF). DMF experimented with optical disk media, jukeboxes, database indexing of the data, and user interfaces. The ST-ECF took the lead in developing the user interface, and the well-known STARCAT program was the result of their efforts. It was expected that DMF itself would only be short-lived, with plans for a major software development contract to deliver a permanent archive system underway. However, when the contract on the permanent archive was delayed, DMF was pressed into operational use.

The DMF system, with a subsequent and substantive augmentation, operated as the primary archive from *HST* launch until early fall 1994. DMF was configured as a VAX-8650 with a Cygnet jukebox and LMSI optical drives. Each LMSI platter stored about 1 GB/platter; the jukebox held 76 platters, for a total on-line capacity of 76 GB. With the data rate from the *HST* at 1.5 GB day^{-1}, the DMF system could hold approximately two months of data on-line. As part of its ingest process, DMF would create a second "safe-store" platter that was shipped to the ST-ECF. An auxiliary process, developed by the Canadian Astronomy Data Center (CADC), copied selected data from DMF for CADC platters.

Within DMF, platters were in one of two states: they were either writable, or they were closed for writing and were only available for retrievals. Since platters generally filled-up with data within a day, there was only a slight (roughly 24 hour) delay in being able to access new data through DMF. Had a larger capacity disk been available for DMF, this delay might not have been as acceptable.

DMF was restrictive in that it only took data in the native ground-system format (GEIS). This would typically have to be converted to some standard format (i.e., FITS) for use by astronomers. DMF could only deliver data to internal Institute systems, which required remote users to log-in to an ST ScI system (called a "host" system) to retrieve their data and then manually copy the data back to their home site. Nonetheless, DMF well supported users from around world in accessing *HST* data.

2. STDADS, the New Hubble Data Archive

The Space Telescope Data Archive and Distribution Service (STDADS) was developed as the permanent archive by Loral Aerosys, under contract to NASA. Its initial version was released to the ST ScI in the spring of 1993. Since then, a joint ST ScI–Loral team, under the supervision of ST ScI, has been evolving and deploying the system. STDADS has been routinely archiving *HST* data since the first *HST* servicing mission in 1993 December. From 1994 February to 1994 November, all of the DMF data were installed into STDADS. STDADS was enabled as the operational archive in 1994 October.

STDADS is based on a cluster of Digital Equipment Corporation VAXes and Alphas. It utilizes four Cygnet 1803 Jukeboxes and nine SONY 930-series large-format optical disk drives. The SONY platters can each store about 6 GB of data; the new jukeboxes each hold 130 platters. This gives STDADS 3.1 TB of on-line capacity or about 5 years' worth of *HST* data at the current data rates.

Unlike DMF, STDADS converts all the data (as practical) to FITS format, prior to writing it to optical disk. This eliminates the need for astronomers to convert the data when they retrieve it. As with DMF, STDADS writes to two platters (a primary and a secondary) as it ingests data. Unlike DMF, STDADS can retrieve data from a platter while it is still open for writing. Once closed, the secondary platter is kept at the Institute at an on-site safe-store. An auxillary subsystem with STDADS, called "bulk distribution," prepares copies of the data for the ST-ECF and CADC in an off-line fashion from the secondary platters. This subsystem can be configured to recreate platters for any of the three sites (ST ScI, ST-ECF or CADC). It can also be modified to support future additional distribution sites, should that be required.

STDADS will also have the ability to deliver data, using standard ftp, to any system that is accessible via the Internet. This will eliminate the need for remote users to have to manually copy data from ST ScI "host" systems, although these systems will remain to support those users that require them. We have adopted the use of e-mail as the mechanism for requesting data from STDADS. A properly formatted mail message is sent to a defined account name and read by the STDADS software. Using this approach, STDADS is open to many different forms of interfaces as well as to our own; a paper by Richon (1995) more completely describes this interface.

3. StarView - the HDA User Interface

As mentioned above, ST-ECF developed STARCAT as a user interface to DMF. However, STARCAT was limited to principally being a CRT interface, and was layered on top of a large body of public code that was deemed unmaintainable. In the early 1990's, the ST ScI negotiated with NASA to develop a new user interface to the *HST* data archive (HDA). This interface, named StarView, was designed from its beginning to be a highly portable tool: specifically able to support both CRT and X-Window devices, and both VMS and UNIX systems.

StarView was developed using object-oriented techniques and coded in the C++ language. The use of these advanced tools has given StarView the flexibility it needs to meet the current and future environments of *HST* researchers. StarView has been supporting access to the HDA for about two years.

StarView employs many notable features, but perhaps the two most significant are the use of ASCII text definitions for all database and display definitions, and the use of a universal-relation SQL generator. By keeping all the definitions of databases and displays in text files, StarView is adaptable to any SQL database that accepts a callable interface. Currently, StarView can interface to any Sybase database. Changes to displays or to the database (or even to new databases) can be made without modifying the code. Moreover, the definitions can be made and maintained by non-programmers.

The universal relation SQL generator within StarView uses information in the database definition files to automatically determine the appropriate joins to be used in making a query. This eliminates the need for either the general user, or the form designer, to have knowledge of the database relations (or even of SQL) to construct working queries. Using these features, the displays, forms, menus, and even help have been carefully constructed by HDA Astronomers to best aid the *HST* GO or researcher in locating and using data within the archive.

4. HDA and the World Wide Web

The use of the Internet, and specifically the World Wide Web, has grown explosively over the past year. It is now expected that most of the information resources available to the public are reachable through some of the more popular WWW tools (e.g., Mosaic, Netscape, etc.). The ST ScI has long been a proponent of using the Internet for connectivity to the astronomy community (in particular) and those interested in *HST* in general. For some time, the Institute has supported WWW, Gopher, WAIS and basic ftp access to its resources. The HDA is also accessible through these resources. For example, pointers on the Institute's home page[1] provide links to the *HST* Data Archives. From there, the user is lead to information allowing them to start-up StarView on one of the host systems. They can initiate either a CRT or an X session.

Beyond using WWW as a reference pointer, we have experimented in providing basic archive services directly through a WWW supported interface. We developed a prototype WWW interface (Travisano 1995) that could initiate

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

queries of the Science or Proposal tables (queries of the Science table can be based on an area about a sky position), and display data in tabular format for the science table (in portrait format for proposal data). Where the science data is public, it can be previewed if the appropriate supporting browser is installed on the user's system (e.g., SAOimage for images, and "xmgr" for spectra). Users with archive accounts can mark selected public science data for retrieval from the archive.

This prototype was developed using the new forms features of HTML, with a custom "common gateway interface" program which was developed to accept the form's parameters. While the prototype was successful, with the easy availability of StarView, we are currently not pursuing an extension of the prototype into actual use.

5. HDA Futures

Certainly STDADS will continue to evolve in meeting the needs of *HST* researchers. In the near term, the ability to deliver data directly to users' home sites, and to respond to requests for data via magnetic tape are the next features to become available. Beyond that, we also have plans to provide access to the Guide Star catalog and other *HST*-related datasets as they become available to ST ScI.

StarView's features will also evolve. With the generic nature of StarView's interface to databases and forms, the Institute is seeking to connect StarView to other catalogs (and possibly other archives) that further enhance a researcher's ability to work with *HST* data. New means of specifying and viewing queries, especially with an eye to using the advanced graphical features of X, are planned. While the current focus will be on evolving StarView, continual attention will also be paid to the evolving nature of the WWW and its browsers, and its potential use with the HDA.

The collection of *HST* observations is a treasure trove for astronomers of today and a generation to come. The Hubble Data Archive is vested with the responsibility for gathering and disseminating this collection to the community. As an integrated set of systems, the HDA will grow and change to meet this responsibility and to participate in its own way, in mapping the future of astronomy.

References

Richon, J. 1995, this volume, p. **??**

Travisano, J. 1995, this volume, p. **??**