The European Science Data Archive for the Hubble Space Telescope

R. ALBRECHT[1,2], F. PASIAN[1,3], F. OCHSENBEIN[4], B. PIRENNE[1], A. RICHMOND[1,5], G. RUSSO[1,6], C. VUERLI[1,3]

[1] *Space Telescope European Coordinating Facility, ESO*
[2] *Affiliated with Astronomy Division, Space Science Department, ESA*
[3] *On leave from Osservatorio Astronomico, Trieste, Italy*
[4] *European Southern Observatory*
[5] *Now at Space Telescope Science Institute, Baltimore, USA*
[6] *Now at Dipartimento Scienze Fisiche, University Naples, Italy*

Introduction

In the framework of the cooperation between NASA and ESA on the Hubble Space Telescope (HST) project the archiving of the science data have always been ranked high in priority. Not only will it be very difficult to apply for and obtain observing time on HST (as those who participated in the first round of observing proposal submission can confirm), the utilization and optimum exploitation of these very expensive data sets can only be achieved through archival research. This has been demonstrated through the very successful de-archival programmes operated by the IUE data archive.

Initially HST data will be archived at the ST Science Institute (STScI) in the Data Management Facility (DMF) on optical disks. As these disks are being produced, a second copy is made and is shipped to the ST European Coordinating Facility (ST-ECF) immediately. This means that the ST-ECF will get a full copy of the HST archive immediately after the observations are carried out. The implication of this is that the ST-ECF had to build a near identical archive with equivalent data retrieval capabilities and security measures: as is well known, HST data will nominally remain proprietary for a period of one year after the end of the respective observing programme, with some exceptions.

The HST Science Data Archive of the ST-ECF was designed and implemented in collaboration with ESO and with the STScI. Realizing that during the projected life time of more than 15 years it would not be possible to maintain hardware compatibility, emphasis was placed on a design which would allow software and data compatibility even on different hardware configurations.

The logical configuration of the two archive facilities (DMF at STScI and Science Data Archive at the ST-ECF) was designed during 1985, and mapped into available hardware. Initially the two hardware architectures were similar, but this changed as time passed. The software elements were identified, and STScI and ST-ECF agreed to assume responsibility for the development of the various parts of the system; in addition, site-specific utilities were developed in both places. Around 1987 the Canadian Astronomy Data Centre (CADC) at Dominion Astrophysical Observatory decided to build a similar archive and participated in the development through discussions and collaboration.

It should be noted that the DMF at the STScI is considered to be an interim facility. Its functions will be taken over by the Data Archive and Distribution System (DADS), which is currently in the early planning stages. This changeover is expected to occur not earlier than 1993. As far as the ST-ECF is concerned, the Science Data Archive is the European HST archive during the lifetime of the mission. Interface control documents have been negotiated to make sure that continued data transfer will be possible.

Archive Requirements

The HST will produce formidable amounts of data, with more to be ex-

pected when the second-generation scientific instruments go into operation in the mid-1990's.

Currently we expect 1-2 Gbytes of data per day, which roughly translates into about an optical disk platter per day. As mentioned above, these data sets will be shipped to the ST-ECF soon after the observations were taken.

Part of the information, the observation catalogue (i.e. the information contained in the header of the sets) will get sent to the ST-ECF in near real time via computer network. For this purpose we recently installed a dedicated link to ESOC Darmstadt in order to connect to the NASA Network using also TCP/IP. In addition, SPAN can be used as long as both archives are using VAX computers. This ensures that information on what has been observed can be made available to European astronomers as soon as possible.

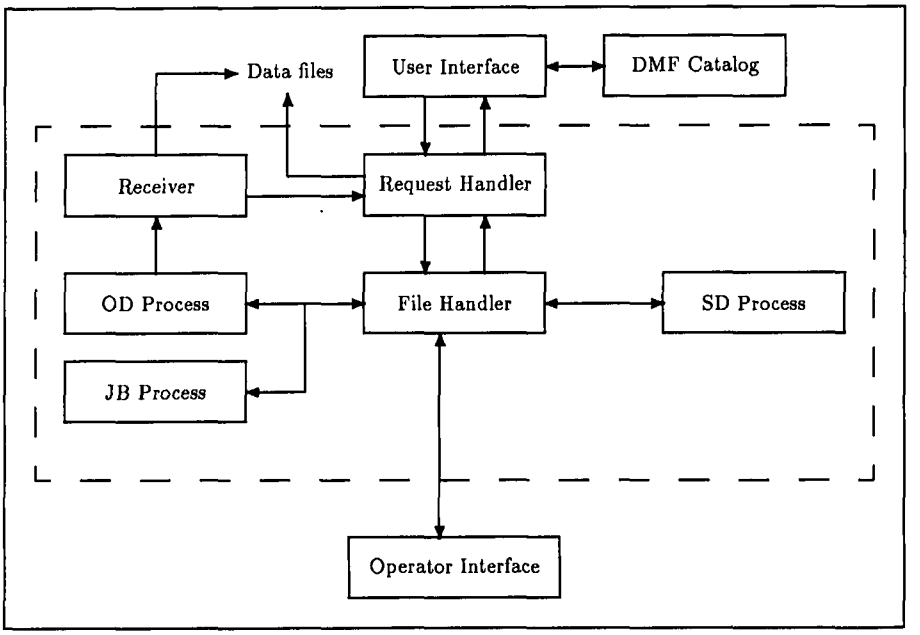


Figure 1: *Software structure showing system components and their interrelations.*

Archive Hardware Design

The archive hardware system is designed following the client-server model: it consists of an archive server (holding the mass storage devices and taking care of bulk data handling), and of a catalogue server, allowing efficient access to the HST catalogue of observations and to some astronomical catalogues, to be used for scientific support. No user accounts are maintained on the archive and the catalogue servers, both for decoupling system work from standard user work, and for archive security reasons. User access (local or remote) to the HST archive and catalogue is guaranteed on other computers in a LAN, acting as clients.

At the ST-ECF, the archive server is a microVAX II/VMS computer, connected to the ESO/ST-ECF LAN and equipped with standard magnetic media (magnetic tape, 1.8 Gbytes of staging disk area). Four optical disk units (an Alcatel-Thompson Gigadisk, two LMSI Laserdrive 1200 s and a Maxtor 800) can be used to physically store the data.

Currently acting as the catalogue server there is a Britton-Lee IDM 500 database machine, running the OMNIBASE/IDM software, and accessible from the archive microVAX and from one of the two clustered 8600 VAXes connected to the ESO/ST-ECF LAN. This device is currently being replaced by a more efficient system based on a general-purpose computer (a Sun) and a commercial data base management system (DBMS), selected after benchmarks and comparisons between different DBMS's have been made, in close collaboration with ESO/IPG.

An 8-mm cartridge tape device has recently been made available to allow a cost-effective eventual distribution to users of large quantities of data. The helical scan technology is currently developing fast, so it is likely that other devices will be purchased, after the community of users has been contacted. The upgrade of the microVAX itself with a more efficient system is being considered, although with low priority at this time.

Archive Software Architecture

As already mentioned, a software system for the archiving of HST data has been built in a joint effort between STScI and ST-ECF: the File Handler and related processes, database and network interfaces were developed at the STScI, the User and the Operator Interfaces and the Request Handler were written at the ST-ECF. The User and the Operator Interfaces are layered on the Proteus/TermWindows software (P/TW), developed as a collaboration between ST-ECF and ESO/IPG.

STARCAT (Space Telescope ARchive and CATalogue) is the user interface to the HST archive: it can be accessed from any terminal in ESO/ST-ECF, and through computer links from remote sites. STARCAT is the way of accessing the HST catalogue and other catalogues available for reference which are stored on the catalogue server. It is the only part of the system actually visible to the user, therefore it has been built in such a way that it is easy to use: there is no command syntax to learn, and catalogue queries can be issued through a form-filling mechanism. It can be considered as a stable software product, since it has had lots of feedback from the astronomers' community. De-archiving requests can be issued from STARCAT, and are managed by the Request Handler. Actual manipulation of files, including OD handling, is taken care of by the File Handler and its processes, which rely on the Operator Interface for exchanging messages with the system operator(s). All of the components of this rather complex system are different processes, some of them running on different computers, and connected to each other by NET, the network interface.

All of the code has been written to be portable through different operating systems, DBMS's, network protocols. VMS and Unix implementations are available for STARCAT, catalogues have been integrated in the IDM and Sybase systems, and network connections have been using DECnet and TCP/IP protocols.

Software developments at the ST-ECF will consist in improvements to STARCAT (in collaboration with ESO/IPG), and in software developments for the rest of the system (in collaboration with STScI and CADC). In particular, only the STARCAT side of this exercise will be of relevance to the user, since it involves the interface the archival researcher will find when dealing with the system; the other upgrades of the software will handle the kernel of the system (being mainly relevant to operator-related operations) and therefore will be transparent to users.

Archive Operations

As mentioned above, the observation catalogue will be accessible to on-site and off-site users through computer networks. This will normally not be true

for the data, except maybe in the case of spectral data (several 10**3 bytes), which could be transmitted at least locally.

However, the main reason (in addition to bulk) why the data will normally not be available on-line is the fact that it is indeed difficult to keep them on-line. To make a year's worth of data available on-line would require a robotic arrangement able to handle in the order of 500 12-inch platters. Not only would this require a substantial amount of money, it is also very difficult to plan in the absence of operational experience. Thus it is foreseen to initially perform the actual de-archiving of data through operator intervention; this also provides an additional measure of data security.

Before accessing the HST catalogue, an identification phase will be needed: browsing the catalogue, a free-of-charge activity, will require just selfregistration, while data retrieval will require privileged ST-ECF staff intervention, if involving costs for the user.

Scientists will use STARCAT to browse through the HST observations catalogue and, through its form-filling mechanism, to identify and select data of interest. From within STARCAT, users will then issue a de-archive request which will be verified for validity, existence and proprietary status of data, and available user credit, and finally queued for operator action. STARCAT can then be left, and the system will notify completion of operations via an e-mail message.

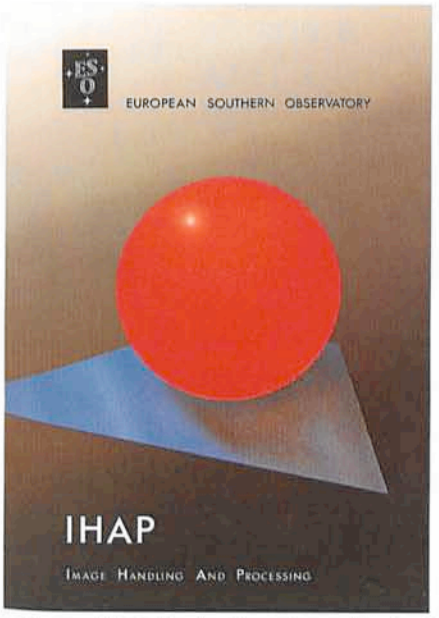
When the files are retrieved, they will be directly delivered on the user's data analysis work area on disk, if at ESO/ST-ECF; otherwise, a hard medium will be produced. Magnetic tapes are currently supported; upon special request, data can be shipped also on 8-mm cartridge tapes or Maxtor 5" 1/4 WORM OD's. As the market develops, other storage media are likely to become available.

Current Status

At this time the hardware elements of the archive are in place and the software has been developed. An end-to-end test of a data transfer from the HST (at that point it was located in a clean room at Lockheed) through the NASA system, through the DMF at the STScI, into the ST-ECF Science Data Archive was carried out earlier in order to verify that the planned transfer will indeed work. An archive readiness review was held at the STScI in mid-November. At the present time we are going through the final preparations for launch. The overall archive system is being exercised at the ST-ECF, before the first HST data are available to users (roughly 7 months after launch time), on the IUE LBL archive, to get a feeling of what day-to-day de-archival operations on the HST archive will look like.

Acknowledgements

The HST archive system has been designed and implemented in close collaboration among STScI, ST-ECF, ESO and, more recently, CADC; all of the staff at these sites participating in this effort are herewith acknowledged. In particular, we would like to mention Leslie Hunt and Sergio Restaino, who have contributed to this project at the ST-ECF.



A new IHAP User's Manual, version April 1990, is available:

- the command description is alphabetically ordered
- it contains all updates since March 1985
- more introductions are given
- reduction methods for 2-D spectra are described
- the manual was revised by M. Véron (OHP), D. Baade, P. Grosbol (ESO) and others.

The technical details of IHAP are described in a new IHAP "Engineering Manual". Please contact Ch. Euler for these new manuals.

IHAP is mostly used at La Silla where HP/1000 computers control telescopes (including the NTT), instruments and detectors.

P. Biereichel

MIDAS Memo

ESO Image Processing Group

1. Application Developments

Since December 1989 the Portable MIDAS has finally become the default version within ESO on the VAX/VMS machines. This led to a heavy workload of fixing bugs and ironing out problems all over the system. We now have a much more stable system and many minor improvements and additions have been included. No major new applications have been added in the meantime, except some new algorithms in the LONGSLIT package.

2. System Developments

Most UNIX workstations running MIDAS do not have their own tape unit but must share a common network tape station. A remote tape server task is now available for UNIX systems using TCP/IP protocols. When installed, it enables the tape commands in MIDAS to access a remote drive as if it were a local device. The interprocess communication interfaces have been rewritten and now use sockets instead of pipes. Therefore, they should be more portable and work also on pure BSD machines, e.g. Alliant.

3. Better Support of DECwindows Under VMS

The portable MIDAS is supported on both VAX/VMS and UNIX systems. To ensure full compatibility of MIDAS with the latest release of VAX/VMS, a VAX station 3100 running VAX/VMS was purchased by the Image Processing Group. This system will be kept updated with respect to the VMS operating system and the DECwindow X11 based display manager. New releases of MIDAS will be verified on this VAX station and thereby certify them for the full range of VAX/VMS systems. The IDI server (the programme which manages the MIDAS windows) used to be run as a batch job. That created problems for distributed systems with a single generic batch queue. Now the IDI server is spawned as a separate process (which also speeds up execution).

4. Change of MIDAS Release Cycle

The release cycle of MIDAS has been modified to ensure greater stability and reliability in the future. The internal development version of MIDAS (i.e. new) is frozen every second month. The last frozen version before a release will be tested both internally and at a number of external beta-test sites. The actual