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

Implementing a New Data Archive Paradigm to Face HST Increased Data Flow

B. Pirenne,[1] P. Benvenuti,[2] and R. Albrecht[2]

Space Telescope – European Coordinating Facility

Abstract. The Hubble Space Telescope Archive at the Space Telescope – European Coordinating Facility[3] (ST-ECF) has undergone several refurbishments in order to cope with user requirements and with the advances in storage technology. In preparation for the installation of the Near Infrared Camera and Multi-Object Spectrometer (NICMOS) and Space Telescope Imaging Spectrograph (STIS) instruments during the 1997 Servicing Mission, it is again necessary to upgrade the archive. This paper describes the adopted strategy and its rationale.

1. Introduction

The science instruments to be installed on *HST* during the 1997 Servicing Mission will substantially increase the current data volume. In order to cope with this increase some of the ECF archive concepts need to be adapted. A second reason for modifying the archive is the high cost of the bulk storage media and the associated hardware. As CD-ROMs have firmly established themselves in the computer market in recent years, it is attractive to look for a solution based on CD-ROM technology. A third reason is that, because of the availability of fast world wide networking, the initial requirement of having in Europe (i.e., at the ECF) the exact copy of the STScI archive can be relaxed. The ECF effort in the archive area should therefore focus on a more dynamic archive environment which complements the work of the STScI. This new approach has been endorsed in the "Report of the ST-ECF Independent Review Group" (May 1996).

2. The Current Situation

Following the transition of the STScI archive to the "DADS" system in 1993, the ECF started to archive *HST* data on the bulk data devices used in DADS, i.e., 12″ Sony optical disks (6.5 GB per platter). The disks are generated at STScI and shipped to the ST-ECF. Currently the ST-ECF receives a full copy of the

[1] European Southern Observatory, Garching bei München, Germany

[2] European Space Agency, Space Science Department, Astrophysics Division

[3] http://archive.eso.org/

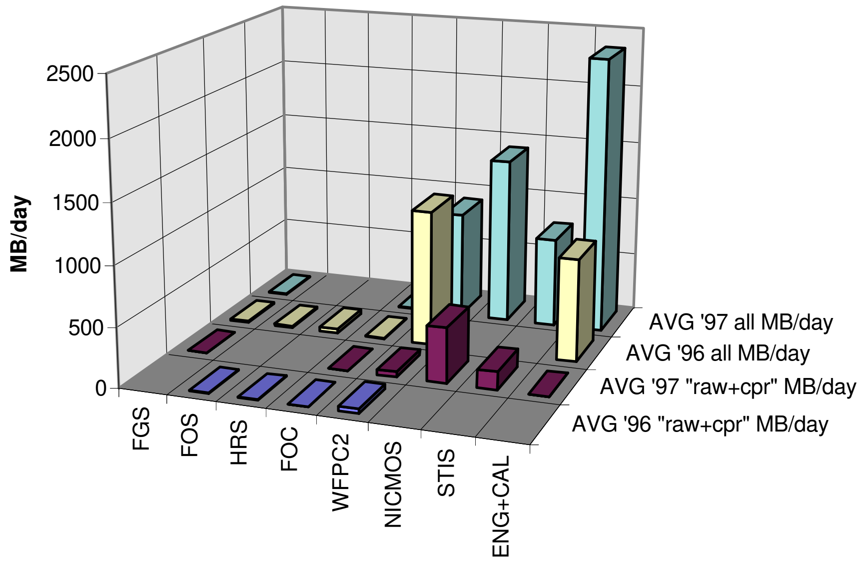

Figure 1. Share of data volume by data type.

archive, including engineering ancillary data. The current average data rate is ~2 GB/day or ~112 Sony disks per year.

Since we receive an exact copy of the archive, software protection mechanisms ensure that proprietary data can only be delivered to authorized users.

3. The Data Rate After the Second Servicing Mission

With the installation of STIS and NICMOS on *HST* in February 97, the data rate is expected to increase up to ~5.4 GB/day or 303 Sony disks per year (270% increase, see Figure 1). It should be noted that a large fraction of the data volume is represented by engineering data (~0.8 GB/day current, ~2 GB/day after 97). The access to engineering data by the (external) European community has been nil.

4. CD-ROM Technology

While CD-ROMs have considerably less capacity than the current Sony optical disks (650 MB/volume vs. 6.4 GB per platter) they are more cost effective owing to their very low unit price (8 USD vs. 300 USD unit cost). In addition, all CD-ROM related hardware (readers and juke-boxes) is cheap, while Sony optical disk hardware—in particular juke boxes—is quite expensive. It is envisaged that the CD-ROMs will eventually be replaced by Digital Versatile Disks (DVD, 3.95 GB per platter; see Scientific American, July 96), which use similar technology and are expected to constitute the next generation CD-ROM standard. DVD readers are expected to be backward compatible with CD-ROMs.

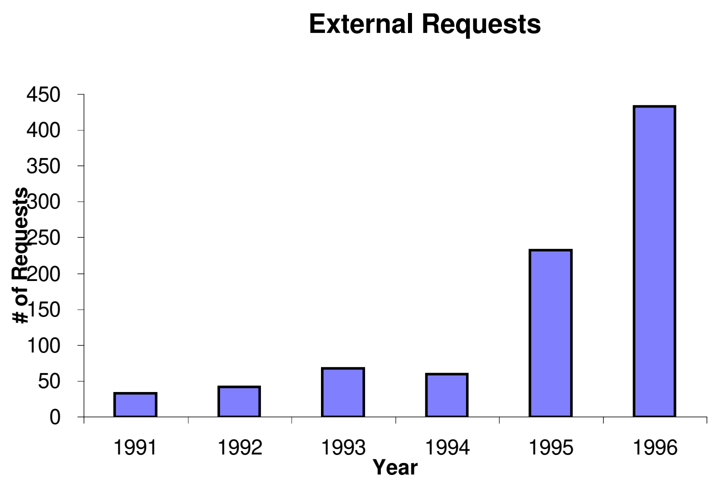

Figure 2. Number of external requests for archived data.

5. On-the-fly Calibration

The Canadian Astronomy Data Centre (CADC) and the ECF developed the capability to calibrate data "on-the-fly," i.e., during the process of retrieving them from the archive (Crabtree et al. 1995). This method has the advantage that the data can be re-calibrated using the best available calibration files, as opposed to the standard calibration done right after taking the observations, when the data have to be calibrated using calibration files taken before the observations. The process has been tested operationally on representative data sets and found to work in a satisfactory manner. Since the capability was announced in May 1996, we noticed an increase of archive requests as users try to improve the quality of the calibration of their data (Figure 2).

Beside offering a valuable user service, on-the-fly re-calibration alleviates the need to transfer and archive calibrated data, considerably reducing the total data volume. However, it also implies that all data distributed out of the archive will have to be calibrated prior to distribution, including multiple re-calibrations for repeated requests of the same data products. On the other hand it makes it possible to have all the raw data on line by copying them to CD-ROMs mounted on jukeboxes (currently, the entire mission *raw science data* is stored on about 60 CD-ROMs). Additional juke-boxes can subsequently be added to the system, making the retrieval and re-calibration process entirely automatic.

Future astronomical projects (e.g., VLT, NGST, ...) are already planning their archives on the on-the-fly calibration concept.

Normal requests and mass retrievals will be handled by spawning off calibration tasks to various archive and ECF computers. This is possible because the ST-ECF Archive Request Handler and the calibration pipeline (OPUS, see Rose et al. 1995) were designed to share the calibration tasks among many machines. The long term solution is to ship compressed raw data and calibration files through the network and perform the decompression and the calibration via client software (e.g., a Java applet) at the user site. But this approach is still beyond our reach.

6. Observation vs. Exposure

So far, the *HST* catalogue only included the notion of individual exposures. The new NICMOS and STIS instrument will introduce the concept of "associations" of exposures (sometimes also called product). Associations group exposures into logical observations, opening up windows of opportunity for further automatic processing of entire observations (e.g., mosaics of neighboring areas of the sky). This concept was not present with previous instruments. The ST-ECF archive is presently retrofitting the existing WFPC2 exposures archive into groups of observations. These groups will enable users to look at logical groups when browsing the catalogue. It will also enable us to provide extra services as part of the on-the-fly recalibration (e.g., cosmic ray rejection).

7. PreView

One of the first major enhancement brought to the Hubble Space Telescope Archive by the CADC and ECF was the concept of PreView: "Imagettes" processed with the current best calibration system are available on-line in the form of PreView (Quick-look) images or spectra that can be viewed instantaneously, thereby helping users assess exposure data quality in a convenient way. PreView will be re-generated on a regular basis so as to use the best calibration available. For the WFPC2, we also plan to use, whenever possible, the cosmic ray-cleaned image.

8. Conclusions

The plan we are describing here involves moving away from expensive 12″ optical storage technology in favor of more economical and stable CD-ROMs. In doing so, we benefit from the availability of the data on-line at a fraction of the current costs. We also open up a window of opportunity to execute large research projects requiring access to a substantial fraction of the archive.

References

Report of the ST-ECF Independent Review Group, ST–ECF Document, 1996

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

Rose, J., Choo, T. H., & Rose, M. A. 1996, in ASP Conf. Ser., Vol. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes (San Francisco: ASP), 311