DATA STORAGE TECHNOLOGY: COPING WITH THE EVOLUTION

B. PIRENNE
Space Telescope — European Coordinating Facility
European Southern Observatory
Karl-Schwarzschild-Straße 2
D-85748 Garching, Germany

ABSTRACT. Modern astronomical data is now either produced directly in digital form or converted from a high resolution analog medium to digital form. These data are therefore best conserved and handled with computer techniques.

This paper will first define 'computer storage technology' as the set of software and hardware tools needed to perform the management of an archive and the control of the archiving and redistribution operations. Various implementation choices that one faces when setting up an archive will be reviewed. The chasm between the life expectancy of the stored data and that of the equipment and procedures needed to handle it will be one of the most important points considered.

Currently available technological solutions will be reviewed. Future trends of the storage technology will be discussed in comparison with the expected data production rate of the next generation astronomical instruments.

1. Why Archiving?

Nowadays, 'archive' is some kind of 'buzzword', a fashionable term in astronomy. A number of reasons explain it, but the more obvious ones are the cost and scarcity of large modern astronomical instruments.

1.1 THE CONSEQUENCES OF THE 'THINK BIG' APPROACH TO SCIENCE

The world's largest observatories are now in the process of building huge instruments at costs which will reach astronomical heights. For instance the future European Southern Observatory's Very Large Telescope (VLT) has a planned cost of roughly 500 million ECUs. If one expects the instrument to be operational for 30 years and to be used 300 nights per year, the nightly cost comes to over 50,000 ECU per night, to which one has to add the running costs. This is a tremendous gift to any successful program. Wise men have decided that this is too much; the output of any particular scientific night must be shared. Hence the data must be archived and made available to the largest community of scientists.

Another argument in favour of data archives is the high demand of those instruments. Due to their rare characteristics, many scientists will try to get access to them; and an over-subscription rate of 5 is probably going to be common. Another way to make the instrument

339

performance accessible to more users is to build an easily accessible archive.

Also, a scientist having successfully applied for time on an instrument is going to look at the targets chosen with a certain purpose in mind. This purpose might overlook some other potential uses of the data. Therefore, the data must be preserved for other scientists who might have other ideas as to the use of it.

1.2 THEY DID IT FIRST

The first groups to make use of digital archive in astronomy were to be found in the space community. Because of the uniqueness of the instruments and their costs, the reasoning described above took place. Two successful current examples are IUE (Wamsteker 1991) and HST (Pirenne et al. 1993; Pasian et al. 1993). Ground-based observatories started some archival ventures in the early eighties, but these do not seem to be well known (Raimond 1991). Radio astronomers were also early players in this game (Wells 1993).

1.3 THE REAL PROBLEM

Since many observatories are now considering digital archives as a way to further the lifetime and multiply the production rate of their most valuable instruments, the time is ripe both for a review of the technology available in this field and also, and more importantly, to confront the major problems of archives: the chasm between the expected lifetime of any storage technology and that of the data. Solutions to overcome this problem will be described as accessible only through a sound and evolution-orientated archive system design.

1.4 THE ARCHIVE CONTRACT WITH THE COMMUNITY

An archive should always provide the following set of services, reflected in a modular design:
1) a catalogue describing the data stored, together with an index pointing to it;
2) a mechanism to store the bulk of the data and to produce the catalogue;
3) a mechanism to retrieve the data efficiently;
4) a user interface to allow easy access to the various data-browse and data-request services;
5) a 'help desk' composed of knowledgeable scientists, involved with the data reduction and the scientific applications.

For more information on these various points, which will not be analyzed in depth, see Pirenne (1993b).

1.5 COSTS

The experience from well-known space missions (Taylor 1992) is that the cost of running and maintaining an archive is of the order of 5 to 10% of the operational budget. It should be of the order of 1 to 2% of the overall observatory budget if one includes design and construction phase of the instruments, the infrastructure and the archive. In this paper, we will explain why the evolution of both the observatory and the archive technology will force one to maintain the level of the budgetary commitment to the proper level.

The rest of this paper will deal with — often contradictory — design requirements to be satisfied in order to guarantee both lifetime and usefulness to the archive.

2. Definitions

Before proceeding with technicalities, it is important to define three important terms: *Storage Technology* — the subject of this paper — but also the *Lifetime of storage technologies* and the *Dynamic and Static capacities* of the archive system.

2.1 STORAGE TECHNOLOGY

The storage techniques used in modern computer systems can be decomposed into a set of several different hierarchical levels. These levels (see Table 1) are all necessary for an archive to work. Level 4 is explained in some detail in Pasian & Pirenne (1992) for the case of the HST/ESO/CFHT archive. Level 3 should be covered by industry or international standards. However, this is not always true; for instance optical disks do not enjoy an internationally recognized standard. Therefore those who want to use this technology must usually define their own (Russo et al. 1993) or buy a commercial proprietary system. Levels 2 and 1 usually are a question of choice; we will cover these in the next sections.

Table 1. The various hierarchical levels of storage technology

Level	Description	Example	Category
1	storage *medium*	CD-ROM, tape reel,...	Hardware
2	*reading device*	disk drive, tape drive	Hardware
	Data communication	Computer data bus	Hardware
3	data structure on medium	file system	Software
4	control & management procedures	archive storage and re-distribution	Software

2.2 LIFETIME OF STORAGE TECHNOLOGIES

Here we touch on one of the crucial problems of archives. Repositories of information are supposed to maintain it safe and legible over long periods of time (Pirenne et al. 1993). That is, both hardware and software should live long. However, if the software can usually be adapted with some effort to cope with the general computer evolution, the hardware situation is more of a problem; both the storage medium and its reading device decay. In some cases, the medium lifetime can be longer than that of its reading device. For example, one of the reasons why the HST archive is now moving from one storage system to a more modern one is that the vendor of the optical disk readers and platters used so far would not support the two systems together and has become increasingly reluctant to repair the drives or to manufacture media. The existing media already written will probably still be readable in 20 years, but from now on it will become increasingly expensive to maintain a reading capability!

2.3 THE ARCHIVES FROM A *STATIC* AND *DYNAMIC* STANDPOINT

The archive systems must be designed with sufficient capacity to store all the data produced by the observatory throughout its lifetime; that is, bulk data and calibration data. The archive must be designed with growth in the plan. It is not sufficient to plan the system with just enough capacity for a current observatory set up, for instruments evolve and produce more and more data.

Moreover the archive system must be designed to sustain the data input/output rate; in order to avoid backlog in archiving and long wait queues in the data distribution service, it is necessary to design data channels and recording/retrieving performance that will manage to cope with the throughput and its rate evolution.

3. Technological Choices

Designing an archive system implies making a number of technological choices which will only be valid for a certain time. These technological choices can be 'modelled' as a function of 3 parameters, themselves function of time. One could write the following 'formula':

$$Techno\_scheme = F(avail\_technology(t), input(t), output(t))$$

Input, the input data volume is normally a constant in the case of a Space-borne mission and evolves as a staircase function for lively ground-based observatories to follow the trend in detectors (see again Pirenne 1993b).

The *output* parameter (re-distributed data volume per unit of time) has a slightly different shape, which the experience shows as modelable using a logarithmic function. For very successful observatory/space mission, the data output rate can sometimes outgrow the data ingest rate. The example of IUE is very useful in this respect (Wamsteker 1991).

The *available_technology* parameter will be discussed below as far as its present state is concerned. In the past years, fortunately, its evolution has followed the trend in detectors; and it has been possible for most archives to be set up as off-line data centres at reasonable costs for most lively ground-based observatories or space missions. Completed mission archives can, usually within a few years time, be migrated from off-line archives to on-line or easily reproducible high-density media when the technology catches up.

3.1 THE HARDWARE

Within the limits of the available technology, and with some reasonable model of the data volume and input - output rates, the choice of tools for running an archive will also be guided by the following — again sometimes contradictory — constraints:
1) storage is best done in a sequential (chronological) manner and is best served by cheap sequential access devices (e.g. tapes);
2) storage must be done at the end of an analysis and possibly also of a calibration pipeline. Data must undergo a minimum of checking and standardization before being permanently stored;
3) storage must not slow down retrieval operations;
4) storage should be done in a software format that will survive as long as possible and is a standard: e.g. data in a computer/software/medium independent format (FITS);
5) distribution must serve requests in a predictable and timely fashion;

6) distribution must provide the choice of various standard output media (for the data not on-line) and should use the standard FITS format;

7) distribution must respond to requests non-normally ordered in a chronological way. Therefore, best data redistribution system will use random access storage systems.

The current choice in techniques suitable for archiving can be structured in the way presented in Table 2. The characteristics of each of those types are analyzed and compared in Table 3.

Table 2. Structure of the various types of storage devices commonly available today

Type	Description	Examples	Recommended usage
RND-SIL	Random access silicon-based 'disk'	RAM-disk	Very short term storage
RND-MAG	Random access magnetic disk	Standard mag disk	Data processing, staging space
RND-OPT-CD	Random access optical CD-ROM	CD-ROM device	Wide distribution of small archives
RND-OPT-WRM	Random access write-once/read-many optical disk	5¼" or 12" optical disk	Larger archives
RND-MOP	Random access magneto-optical disk	5¼" rewritable optical disk	Short term archiving
SEQ-MAG-HSC	Sequential access tape, helical scan	4-mm DAT/DDS or 8 mm Exabyte	Backup, data distribution
SEQ-MAG-STK	Sequential access tape, Serpentine track	QIC-xxx tapes	Backups. software distribution
SEQ-MAG-PTK	Sequential access tape, parallel tracks	9-track reel tape	Old-fashioned?
SEQ-OPT	Sequential access optical tape	1-TByte optical tape	Large archives with limited file retrieval

Table 3. Review of the individual data storage devices available today and their physical and economical characteristics

Type	Drive cost (ECU)	Media cost (ECU/ GB)	Rnd File Access sec.	Transfer rate MB/s	Vol Cap. GB	Medium Lifetime year	Description
RND-SIL	N/A	45000	0.0005	100	1.0	N/A	RAM disk
RND-MAG	1500	1500	0.01	5	4.0	10	Standard mag disk
RND-OPT-CDW	8000	40	N/A	0.3	0.6	>30?	Writable CD-ROM
RND-OPT-CDR	950	20	0.33	0.15	0.6	>30?	Standard CD-ROM
RND-OPT-WRM	25000	65	0.6	0.3	10.0	100	12" WORM optical disk
RND-MOP	4000	225	0.35	0.2	0.5	>10?	Rewritable OD
SEQ-MAG-HSC-DAT	1500	18	< 60	0.17	2.0	3?	4 mm DAT/DDS
SEQ-MAG-HSC-EXA	3000	15	< 60	0.18	5.0	2?	8 mm Exabyte
SEQ-MAG-STK	1000?	18	?	0.5	1.4	5?	¼" QIC
SEQ-MAG-PTK	7500	100	60.00	0.6	0.2	5+	9 track reel tapes
SEQ-OPT	150000	5	60.00	1.5?	1000	>20?	Digital Optical tape

Notes on Table 3:

1) the prices given in ECU (= US $1.25) and are based on upper limits found on the European market in the summer of 1993;

2) SEQ-MAG-HSC devices (DATs and Exabytes) currently enjoy a great popularity due to their high density and the very low price per GB. The standard file seek time is however very bad on these devices and the figures presented in Table 2 can only be reached if a special software driver is installed in the computer operating system. Standard file random access times are normally of the order of an hour!

3) SEQ-MAG-HSC devices (DATs and Exabytes) can be purchased with internal hardware data compression. However, for archival purposes, the driving computer can also compress software-wise. For archive products distribution, compression is *not* recommended. See also the section on data compression;

4) SEQ-MAG-HSC devices (DATs and Exabytes) when used for data distribution should only use the basic common denominator model that most users can be expected to read easily i.e. 2.5 GB first generation Exabyte and standard DAT/DDS cassettes;

5) the real capacity of SEQ-MAG-HSC devices (DATs and Exabytes) can be very different from what is announced by vendors, for the inter-file gaps on the tape occupies a non-negligible space. A simple test could write 700 4.25 MB FITS files on an Exabyte (roughly 3 GB/5 GB). The DAT did obviously better with 1.9 GB/2.0 GB;

6) the lifetime of the various devices are values provided by vendors for most devices except for the SEQ-MAG category, where there is sufficiently long experience for a personal assessment. The question marks on the longer lifetimes insist on the fact that no practical experience can possibly prove or disprove those figures.

3.2 DATA COMPRESSION

Data compression has become fashionable in the past two years. Two types of compressions are available: destructive (lossy) and non-destructive (non-lossy). Each has its own application field. Lossy compression will usually be applied when quick data transfer is required and the resolution of the original image can be damaged without harming the application. Destructive compression can reduce the size of an astronomical image by a factor of 200 (Hill et al. 1993) without losing too much visual quality. Non-destructive compression is used for data transfer and obviously compresses less than the previous method; a factor of 3 in size is usually the best one can get nowadays for astronomical data. It has however to be borne in mind that this field is still under development and better methods may still appear.

Archives can not usually forecast the future use of their data. That means lossy compression is to be rejected from the design. Non-lossy compression or no compression at all is now a trade-off between the time spent in compressing/decompressing and the amount of space gained on the medium.

Data compression is available hardware- and software-wise. For astronomical purposes, where images are usually noisy, the existing hardware compression system are usually not well suited. They are normally designed for TV broadcasts (in a lossy form) or for backup purposes (non-lossy compression).

3.3 JUKE BOXES (AUTO-CHANGERS)

The role of juke boxes is to place a large collection of on-line media in a special enclosure in order to allow for a rapid insertion in the reading device using a robotic arm. The data made available in a juke box or auto-changer is often qualified as 'semi on-line', the file access time being limited only by the volume change operation, which takes a few seconds. This technique is very useful for unattended random retrieval of files from large archives. However, the technique will really pay off if the amount of data to be made available for rapid retrieval does not exceed the capacity of 2 to 3 such juke boxes, otherwise the costs become unbearable. Also, one should keep in mind that juke boxes will not remove entirely the manual operator work, for new media must be entered in or removed from the juke box from time to time. If retrieval time is not an overwhelming constraint, and if the archive cannot be held semi on-line with a

reasonable number of juke box units, human operators should be preferred. The example of the European copy of the HST archive is such a case (Pirenne 1993a).

The cost of this equipment can augment the price of the stand-alone reading device by a factor between 4 and 400. The lower factor concerns the random access devices (e.g. optical disks), the higher factor concerns the large SEQ-MAG-HSC auto-changers. As far as the capacities are concerned, the largest optical disks juke boxes can hold 1000 GBytes, and the larger Exabyte or DAT auto-changers will hold up to 580 GBytes. (All figures with uncompressed data and standard drives.)

4. Conclusions

When setting up archives, a number of questions arise. Besides those dealing with the presently available technology for solving current problems, these questions usually deal with the long term prospects of the chosen technology. We have seen the current possibilities in the previous section; let us now say a few words about the future.

The astronomical data production rate is nowadays increasing at a tremendous rate. the average size of CCDs images has doubled every other year and the prospects for the near and medium term future call for a continuation of the current trend. To demonstrate the archival challenges for the years to come, we note that there are many examples of CCD mosaic being planned, built, or even in use. Will the storage devices support this exponential data growth? The answer is probably yes, since the last ten years of computer history have shown a similar — if not more sustained — trend towards larger volume capacities.

Another related problem is the data transfer from the CCD camera to the archive centre and from the archive centre to the final users. The development in computer networks have shown large increases in both bandwidth and reliability. The growth of the data production rate of future astronomical detectors following a parallel trend is also a forecast that one can venture without too many risks.

The processing of larger and larger amounts of data are of course also a concern for archive centres, which sometimes are also given the responsibility of calibration and homogenisation pipelines. Again one need not be overly optimistic to confidently admit that the processing capabilities of tomorrow's computers will continue to overpower the data amount growth. To illustrate this assertion: it is today far less expensive to purchase a computer for astronomical data handling of say, 4K by 4K CCDs than it was 6 years ago for 512 by 512 pixels.

There nowadays is a fair number of astronomical archive centres with considerable experience to share with would-be archives. Today, it is no longer necessary to invest years of effort in re-inventing software and hardware solutions when setting up a project. Looking around and adopting proven solutions will allow one to concentrate on the really important aspect of archives: the data, rather than on the tools to manipulate them.

References

Hill, N., Crabtree, D., Gaudet, S., Durand, D., Pirenne, B. and Irwin, A., 1993. 'Generation and Display of Online Preview Data for Astronomy Data Archives', The Third Annual Conference on Astronomical Data Analysis Software and Systems, October 13-15, 1993, Victoria B.C., Canada. Proceedings in press.

Pasian, F. and Pirenne, B., 1992. 'Migration to Unix of the ST-ECF Archive', ST-ECF Newsletter **17**, 20.

Pasian, R., Pirenne, B., Albrecht, R. and Russo, G., 1993. 'The European HST Science Data Archive', *Experimental Astronomy*, **2**, 377.

Pirenne, B., Benvenuti, P., Albrecht, R. and Rasmussen, B.F., 1993. 'Lessons learned in setting up and running the European copy of the HST archive', SPIE International Symposium on Aerospace and Remote Sensing, Space Telescopes, Orlando, Florida, April 1993. Proceedings in press.

Pirenne, B., 1993a. 'Two years of HST Archive operation — first conclusions', ST-ECF Newsletter **19**, 18.

Pirenne, B., 1993b. 'Choosing the right tools for an Archive', Workshop on Handling and Archiving data from ground-based observatories, Trieste, Italy, May 1993. Proceedings in press.

Raimond, E., 1991. 'Two Observatories and their Archives: the Isaac Newton Group of telescopes, La Palma and Westerbork Synthesis Radio Telescope', ed. D. Egret, Kluwer Academic Publishers, Dordrecht, p. 115.

Raimond, E., 1993. Workshop on Handling and Archiving data from ground-based observatories, Trieste, Italy, May 1993. Proceedings in press.

Russo, G., Russo, S. and Pirenne, B., 1993. 'An Operating System Independent WORM File System', submitted to *Software — Practice and Experience*.

Taylor, B.G., 1992. 'Space Astronomy from Archives and Data-Bases: the European Context', Astronomy from Large Databases II, eds. A. Heck and F. Murtagh, Haguenau, France, September 1992.

Wamsteker, W., 1991. 'The many faces of the Archive of the International Ultraviolet Explorer satellite', Databases and On-Line Data in Astronomy, ed. M. Albrecht and D. Egret, Kluwer Academic Publishers, p. 35.

Wells, D., 1993. Workshop on Handling and Archiving data from Ground-based Observatories, Trieste, Italy, May 1993. Proceedings in press.