Astronomical Data Analysis Software and Systems IX
ASP Conference Series, Vol. 216, 2000
N. Manset, C. Veillet, and D. Crabtree, eds.

Magneto Optical as a Viable Storage Solution for the HST Archive

F. Abney, J. Richon

Space Telescope Science Institute

Abstract. In these days of CD, DVD, HD-ROM, NFR, and myriad other acronyms, a tried and true storage technology came out a winner for HST data. Some of the issues we considered were: current data investment, growth of the HST archive, the depth and inimitability of the data, the heavy duty cycles on the hardware due to unrelenting retrieval patterns, and initial and operating costs. We will present the methods used to define requirements, evaluate technology solutions, why we selected Magneto Optical as our current storage solution, and why we continue to watch the storage industry for new products with an eye towards support of NGST. We also explain why there are no feasible long-term storage solutions, and that data migration will always be part of any active, valuable archive.

1. Current Status

During the month of October 1999, the HST data holdings exceeded 7 TB. This data is stored on over one thousand twelve inch Sony optical platters. The current jukebox capacity is only 524 disks, thus about half of the platters are managed using manual intervention and jukebox external drives. This is both time and labor intensive.

The archive is currently growing at a rate of approximately 1.8 TB per year. After ACS is installed during servicing mission 3B, we expect our daily data ingest rate to increase considerably. In addition, we experience retrieval patterns about 4.4 times that of our ingest rate. Both ingest and retrievals employ the same platters, jukeboxes and drives. This almost constant duty has caused our hardware to age and is showing up by creating maintenance and reliability nightmares. The eight internal jukebox drives are approaching read and write limits, and hardware down time is increasing due to heavy usage and age.

2. Requirements and Evaluation

The archive operations and development teams met together to determine requirements for the new archive system and to evaluate the hardware and software systems available in the market. The first requirement was that hardware and software purchases had to be within our budget and would cost much less then our current archive hardware. This was found to be readily available since the cost of storage has dramatically decreased in the past few years. Another requirement was for a proven technology that could be trusted with our data over the life of the media. Also, the hardware needed to be robust enough to withstand the sizeable ingest and retrieval rates that we experience now and the increase we expect to incur over the next few years. Lastly, the requirement was to purchase as much of-the-shelf software as possible, in order to eliminate costly, specialty software either written for or by us.

To evaluate what was available, we first searched out other archive sites to find out where and how they stored their holdings. In doing so, we discovered that many places considered archives as "one-way" data holdings, they copied the data to some storage media to preserve it, yet rarely had retrievals from it. We had to distinguish the difference in the word "archive" that is prevalent in the industry today. We did find many other sites with similar ingest and retrievals patterns and similarly sized data sets, so either visited or spoke with personnel from these places. Some of these were astronomical holdings at NASA Goddard. We found a significant number of archives using MO technology, and in almost every case, they were very happy with their choice of media.

DVD also held much interest for us. The promise of a high density, inexpensive media has been in the market for quite some time and we were hoping that it would be the solution for our archive. After reviewing the market status of DVD, we were disappointed that it has still not matured. There are currently two different (and most likely incompatible) DVD writers on the market, and one more soon to be announced. It is unknown which format will "win" and be around in the next few years. We were concerned about investing in hardware for a version of the technology that may not be around in a year or two. There were also concerns about the DVD media itself. Although based on CD technology, DVD is a relatively new product and we didn't know the reliability factors of constant handling of the DVD disks either in or out of a jukebox. It was decided not to go with a DVD product for our archive at this time, but that we should keep a watch on the technology to see what the next few years hold for it, and hope for some format convergence and stability.

After we narrowed our choice down to magneto optical technology, we evaluated the hardware and software vendors on the market. We compared track records, hardware costs, available software packages, and data throughput. Once we had a potential hardware vendor identified, we did a site visit to test out the jukeboxes and software packages to make sure they could perform to our ingest and retrieval requirements.

3. Why MO?

Magneto Optical is a very stable storage niche. The development path for MO is well defined and understood by the industry providing a growth path that we can plan for. MO is widely used by groups and agencies that place a high value on their data. MO is used by: the FBI to store fingerprint images; NASA for several missions; and many hospitals for patient records and images. While MO cannot be considered a commodity product, it is widely used and supported through out the world.

4. Other Storage Technologies

While investigating optical storage solutions we came across many emerging storage technologies that promise higher densities at lower costs. These solutions range from one-terabyte tapes to 180 TB jukeboxes using a broad spectrum of material. Below is a partial list of storage solutions:

- Terastor Near Field Recording (NFR) 20 GB/drive (www.terastor.com),
- Norsam HD-ROM 200GB/120mm Disk (www.norsam.com),
- Maxoptix Optical Super Density 40 MB/MO disk (www.maxoptix.com),
- Read-Rite 1.5GB/3.5 Disk (www.readrite.com),
- Keele High Density 86 GB/Centimeter^2,
- Colossal Storage Inc 500 GB/Inch^2,
- Nanochip Inc. 1.4TB drives in 3.5" form factor (www.nanochip.com).

One can visit many web sites that provide links to emerging storage technologies. As with all computer technologies, whatever solution one selects today will be out of date within five years, if not sooner.

5. Conclusion

As the permanent source archive for HST data, we decided that MO technology was the best solution since it is proven to be reliable, robust, safe and within budget. However, we know that a few years down the road many more storage technologies will meet the same requirements as MO does today, so constant reevaluation of the storage market is necessary. Without migration every three to five years, the technology chosen is bound to be left behind as support and media for it becomes scarce.