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

Save the Bits - New Features for a New Millennium

R. L. Seaman

IRAF Group[1]*, NOAO*[2]*, PO Box 26732, Tucson, AZ 85726*

Abstract. The original *Save the Bits* (STB) installation has been running at Kitt Peak National Observatory for more that six years. NOAO has since commissioned four other STB installations to support the archiving of raw images at Cerro Tololo Interamerican Observatory, at the WIYN Consortium telescope on Kitt Peak, and for both the Mosaic-I camera at KPNO and the Mosaic-II camera at CTIO. STB is used outside of NOAO as well. The *Save the Bits* system supports untended queued archiving to both tape and CD-R media.

The NOAO/IRAF Mosaic Archive Pipeline is an initiative to automatically reduce all NOAO Mosaic images and to construct a searchable archive of all raw and reduced Mosaic data. To support the new Mosaic pipeline, as well as resolve the differences in the various older *Save the Bits* implementations, a new version of STB has been implemented. For use in the Mosaic pipeline some major new features have been added to STB. These include the ability to propagate non-FITS data into the archive using the new IRAF FITS foreign file extension mechanism, and messaging capabilities for the real time updating of archive index and catalog entries. By integrating STB with the Mosaic pipeline system the capability has been added to enter the index and header information extracted from the data by STB into an SQL database. General SQL queries can then be performed to determine what is stored on the data tapes produced by STB, and to produce selection sets to extract data from the STB archives.

1. Introduction

The NOAO/IRAF *Save the Bits* (STB) archive was originally commissioned at Kitt Peak National Observatory on 20 July 1993. Over 1.5 million images have been archived from this general purpose KPNO archive servicing the full range of IR and optical instrumentation of eight separate Kitt Peak telescopes, including

[1]Image Reduction and Analysis Facility, distributed by the National Optical Astronomy Observatories

[2]National Optical Astronomy Observatories, operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the National Science Foundation

the nighttime stellar program at the National Solar Observatory McMath-Pierce telescope.

A second NOAO STB system was installed at the Cerro Tololo Interamerican Observatory in March of 1996. Almost 700,000 optical and IR images have been archived from CTIO telescopes since then. The combined KPNO and CTIO holdings between these two original systems represent about 5 TB of imaging data. The holdings of these standard observatory instruments are growing at several hundred thousand images a year for about 1 TB per year between KPNO and CTIO.

The *Save the Bits* software has been upgraded to support CD-R media as described by the current author in the proceedings of ADASS VI, "WIYN Data Distribution and Archiving" The WIYN Data Archive and Distribution System (DADS) was commissioned in January of 1999. About 300 duplicate CD-R copies have been mastered at WIYN since then containing FITS data from the WIYN Imager and Hydra multi-object spectrograph instruments.

In addition to these three STB installations, NOAO has also dedicated two separate exabyte tape based systems to the NOAO Mosaic and Mosaic-II instruments at KPNO and CTIO, respectively. The Mosaic-II has just passed from its commissioning phase to the earliest shared risk observing runs.

The original Mosaic at KPNO was commissioned in 1997 with engineering grade CCDs and Mosaic STB operations were started in February of 1998. Since then the instrument has been upgraded to the thinned science grade CCDs and over 33000 Mosaic images have been archived to over 600 duplicate pairs of exabyte tapes. Each Mosaic frame is 8Kx8K or about 138 MB. The total northern Mosaic holdings are therefore somewhat over 4 TB.

We estimate that each Mosaic camera will be used to acquire about 20000 images each year when in full operation. This represents about 5 TB of data yearly just from these two instruments.

Save the Bits is freely available to outside institutions and is straightforward to install and manage. Hardware requirements are minimal. At least three STB installations are in current operation outside of NOAO.

2. New Features

The NOAO/IRAF Mosaic Archive Pipeline is an initiative to automatically reduce all NOAO Mosaic images and to construct a searchable archive of all raw and reduced Mosaic data. The IRAF group is actively working on various facets of this effort, including a variety of related improvements to the *Save the Bits* software that have already been installed on the online Mosaic archive servers. This STB update also provided an opportunity to merge the CD-R and original tape based versions of STB.

The Mosaic pipeline design required some major new features in STB. An important requirement of the MAP is to merge human generated metadata into the archive data stream. These metadata include information from the observing proposal as provided by the NOAO scheduling database, observing log files as automatically generated at the telescope, real time comments from observers, weather data, nightly reports from the observatory's telescope operators and potentially many other data streams such as messaging telemetry from the telescope and instrument control systems.

These mostly non-imaging data sources could potentially be archived in their native format, but this would complicate the pipeline design. The *Save the Bits* software handles FITS objects, either primary header data units (HDUs), conforming FITS extensions, or FITS multi-extension files (MEFs). The IRAF group has implemented a new FITS extension type, *FOREIGN.* , that is described in Nelson Zarate's paper for ADASS IX. A small amount of effort was expended to fully integrate the handling of these foreign FITS objects into the STB data handling procedures.

Propagating these metadata as foreign FITS objects provides a great simplification in the handling and storage procedures. Only one archive medium is required and the complete holdings of the archive can be maintained in a single unbroken sequence of serially numbered FITS objects. Maintaining duplicate copies is straightforward since STB already supports this feature. Each foreign FITS object inherits the pre-existing FITS features such as keywords to provide time stamps, unique archive identifiers, serial numbers as well as explicitly expressing the size of each object. The FITS checksum mechanism (described in the author's "FITS Checksum Verification in the NOAO Archive" from the ADASS IV proceedings) can be used to ensure the integrity of these metadata foreign FITS extensions, as well as the archival raw and reduced data.

Finally, adding a new stream of metadata into the archive is very straightforward - requiring only that the metadata be encoded as foreign FITS objects and sent to the common STB data queue.

The Mosaic Archive Pipeline can be configured to either operate untended at the telescope, or offline by an archive operator away from the telescope. The latter mode of operation may rely on the physical transport of data from the telescope due to the large quantities of data involved (tens or hundreds of MB per exposure) and the remote location and resulting limited bandwidth available between the telescope and the central archive data center (in downtown Tucson, for instance).

Since this physical transport of data is slow, and perhaps irregularly scheduled, messaging capabilities are required for the real time updating of archive index and catalog entries between the telescope and the data center. These have been provided in a generic manner that is currently implemented as a simple unix mail queue. After each Mosaic exposure, or after a foreign FITS metadata object is ingested by the STB server on the mountain, a pair of messages are sent to a separate remote MAP server downtown to update the archive catalog and the separate archive index that cross-references the catalog to the actual data tape, tape file and FITS extension within that file.

These FITS catalog entries will be ingested on the MAP server into the database that is used to control the day-to-day MAP operations. This can be any generic database software that supports normal features such as SQL queries. We are currently building these facilities around the publicly available MySQL package.

The downtown MAP server supports its own *Save the Bits* system that will be used for queued storage of the reduced data products generated by the pipeline. The output tapes, or potentially DVD-R disks in the future, will then be shipped to a remote data center that will supply contracted services such as web-based data searches and retrievals. The familiar STScI multi-mission archive will likely be used, but no final contract has been negotiated.

The integration of STB with the new Mosaic pipeline generates benefits for both systems. The MAP inherits the STB features discussed above, but the STB data holdings will also benefit from the new infrastructure being implemented for the pipeline. In particular, the capability is being added to enter the STB index and header information into a SQL database. The normal SQL queries can then be performed to generate the wide variety of archive management reports that have previously been compiled by hand (that is - by general unix software tools).

References

NOAO/IRAF *Save the Bits* archive, *http://iraf.noao.edu/projects/stb*

Seaman, R. 1993, "Managing an Archive of Weather Satellite Pictures" in ASP Conf. Ser., Vol. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes (San Francisco: ASP), 113

Seaman, R. 1994, "NOAO/IRAF's Save the Bits, A Pragmatic Data Archive" in ASP Conf. Ser., Vol. 61, Astronomical Data Analysis Software and Systems III, ed. D. R. Crabtree, R. J. Hanisch, & J. Barnes (San Francisco: ASP), 119

Seaman, R. 1995, "FITS Checksum Verification in the NOAO Archive" in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 247

Seaman, R. 1997, "WIYN Data Distribution and Archiving" in ASP Conf. Ser., Vol. 125, Astronomical Data Analysis Software and Systems VI, ed. G. Hunt & H. E. Payne (San Francisco: ASP), 306