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

WIYN Data Distribution and Archiving

Rob Seaman

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

Ted von Hippel

U. Wisconsin/WIYN

Abstract. The NOAO/IRAF Save the Bits archive has been operating for over three years at Kitt Peak National Observatory and at the National Solar Observatory's nighttime program. Since that time, the W. M. Keck Observatory and the Cerro Tololo Inter-American Observatory have also adopted the software. These first generation Save the Bits installations rely on Exabyte tapes as the archival medium, typically using pairs of drives to produce duplicate copies of the data for heightened protection against data loss.

The upgrade of Save the Bits that is currently in progress to support writable CD-R drives is discussed. In addition to another media option, this expands the role of the package to include data distribution as well as data archiving. Dual CD-R copies are produced as with tapes. One copy is retained for archival purposes, but the second copy of each nightly CD is released to the appropriate institution as the principle means of data distribution from the telescope. The four individual institutions are free to handle their copy of the data in any appropriate way, such as by mounting the disks into a jukebox as they are received. Both raw and mountain-reduced data are included in random access FITS files on the ISO 9660 CD-ROMs. Planned future improvements include support for DVD format disks.

Save the Bits is freely available to outside institutions and is straightforward to install and manage. Hardware requirements are minimal and other storage media should be straightforward to support.

1. Introduction

The WIYN Consortium consists of the University of Wisconsin, Indiana University, Yale University and NOAO—the National Optical Astronomy Observatories. The Consortium manages the 3.5 m optical alt-azimuth WIYN telescope on

[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.

Kitt Peak near Tucson, Arizona. The WIYN telescope supports two primary facility instruments, a wide field CCD imager and the Hydra multi-object spectrograph. These instruments are mounted at the two Nasmyth foci of the telescope allowing both instruments to be used throughout a given night. The primary mirror system takes advantage of modern active optics technology through 66 separate actuators that push or pull on the back face of the mirror to maintain the best possible optical figure. A thermal control system maintains the surface of the mirror within 0.2 C of the ambient air temperature, eliminating mirror seeing, which is caused by turbulence in cool air over a warmer mirror surface.

The motivations for a WIYN Archive and Data Distribution System are to enhance the total science output of the telescope over its lifetime and to ease the data distribution and handling process at all four WIYN institutions.

An archive increases the total science output of the telescope as investigators attempt new problems with old data, and as statistically large samples of different classes of objects are accumulated, often from projects initially undertaken for a wide variety of purposes. Indeed, as a large digital imagery and spectroscopy library is accumulated, many new uses will be made of it. One can expect not only statistical studies of commonly observed objects, but also test reductions and analyses by observers contemplating new projects.

Currently WIYN data are recorded by the observers on Exabyte or DAT tape using IRAF tasks. While this procedure does store data, it does not allow for easy recovery or dissemination of the data. Since tapes do not allow random access, since they may not last more than a few years even if carefully stored, and since every observer writes their data to tape differently and has different styles of logging their observations, the result is a cumbersome and heterogeneous data depository. Even the original investigator must load their data to disks, often repeatedly, and often they must spread their data across many disks in order to have access to an entire run at one time.

The system described here is based on an automatic and homogeneous data storage process using random-access media (CD-ROM). The long-term goal is a robust and easy to use WIYN data archive. The near-term goal is a robust and easy to use data distribution system. The raw data CD-ROMs are mounted in jukeboxes at the institution of data ownership and are immediately and continuously readable without transferring the data to hard disk. Additionally, CD-ROMs are stable media and the data are in a standard format.

Observers relying on this mechanism exclusively will be responsible for verifying the content of the archival media containing their data. Tools will be provided to check the image headers and display the images directly from the archival media. Additionally, NOAO backs up all data taken at WIYN and with the KPNO telescopes using *Save the Bits*. This software automatically stages CCD data taken at all Kitt Peak telescopes to a single Sun workstation and its disks, then writes these data as FITS files to Exabyte tape.

2. System Overview

The physical approach proposed is to write WIYN data to two CD-ROMs (for redundancy) simultaneously using two CD-ROM writers, mounted on a dedicated Sun workstation using a version of *Save the Bits*. A separate directory will be written for each instrument (currently Hydra and the Imager) on each CD-ROM as well as directories for mountain-reduced data separate from the raw data. The data files will be copied into these directories as individual FITS files. The data will be staged to a hard disk before writing the CD-ROMs, and the resulting CD-ROMs verified against the hard disk data set and against each other. The mountain-wide *Save the Bits* archive will also still back up WIYN data.

At some point during each day the (now archival) CD-ROMs will be ejected from the drives, such that individual CD-ROMs belong to individual nights, and thus to individual institutions. One copy of each night's data will be kept at NOAO, and the other copy returned to the institution which owns the data. For the CD-ROMs which return to the Universities, the media can be mounted in a jukebox and the observer will effectively have new disk space with their observations. For the CD-ROMs which are stored at NOAO, the media will be available as a backup and for repeated duplication and distribution to the requesting observers.

The advantages of CD-ROMs as the archive media are multiple: they are dependable, easy to store and use, they are random access media, and they are forward compatible with the next generation of DVD "CD-ROMs." Random access is important for an archive as well as for data distribution purposes, as it greatly decreases access time and allows someone to recover, copy and distribute data from widely disparate media locations. Forward compatibility also seems assured with the next generation of higher density DVDs as the principal manufacturers have agreed to this.

The major disadvantage of CD-ROM as an archival media is that a single disk only holds 650 MB. Based on current rates of data taking, this data quantity is large enough, however, such that one CD-ROM will be sufficient to hold the data from a single night on 60% of all nights, while three CD-ROMs will only be needed approximately ten nights a year.

The software to perform the archiving will be based on *Save the Bits*, which will be modified to archive only WIYN data, and to write to CD-ROM writers. *Save the Bits* will also perform the function of creating a growing index file from the headers of all archived data. This will allow for simple searching and post-processing programs to be written. *Save the Bits* is proven software which has preserved more than one terabyte of data from seven telescopes at Kitt Peak over the last two and a half years, and it is also being used at CTIO and at Keck. This software needs only minor adaptations, its use requires minimal interaction from mountain staff, and it already can handle a number of contingencies, including system crashes.

3. Daily Maintenance Chores

The archival system requires the following tasks, generally once, at some point between dawn and mid-afternoon:

1. check the monitor program, *bitmon*, to see that the archive system wrote a good pair of CD-ROMs,

2. eject, date-stamp, and log the just written CD-ROM pair, and

3. load new blank CD-ROMs into the CD-R writers.

In the event that one or both of the CD-ROM disks of the pair were bad, as indicated by the verification passes of *Save the Bits*, new blank CD-ROMs are inserted into the drives, and that dataset is re-archived from the staging disk. Occasionally two (and less often three) CD-ROM pairs per night of data will be required, and the above process will have to be repeated a second (or even a third) time during a given 24 hour period. The system can be allowed to fall behind by several CD-ROMs, if necessary, though this is not desirable.

At the end of each run, the CD-ROM disks will be transferred downtown for a final read verification using a separate CD-ROM drive. This will also provide the opportunity to read detailed observing run information from each disk which will be used to generate an informative label to be printed onto each disk using a thermal printer designed to print directly to normal CD media.

References

NOAO/IRAF *Save the Bits* archive[3]

WIYN Consortium[4]

Exabyte Corporation[5]

CD Information Center[6]

Stinson, D., Ameli, R. & Zaino, N. 1995, Lifetime of KODAK Writable CD and Photo CD Media[7]

DVD: Inside Story[8]

---

[3]http://iraf.noao.edu/projects/stb

[4]http://www.noao.edu/wiyn/wiyn.html

[5]http://www.exabyte.com

[6]http://www.cd-info.com

[7]http://www.cd-info.com/CDIC/Technology/CD-R/Media/Kodak.html

[8]http://www.sel.sony.com/SEL/consumer/dvd/index.html