

THE ARCHIVING OF SPACE ASTRONOMY DATA

Jaylee M. Mead
Space Data and Computing Division
NASA-Goddard Space Flight Center
Greenbelt, MA 20771
U.S.A.

When considering the archiving of space astronomy data, one has only to look at the current and projected astrophysics missions to be aware of the enormity of the archival requirements and the urgent need for careful planning in order to properly manage all the data which is and will be acquired. The current list of astrophysics missions for which NASA supports observing/archival phases during the next ten years numbers 23! Even when a mission's observing period was less than a year, as for the Infrared Astronomy Satellite (IRAS), or only a week, as for Astro-1, the archival phases last for many years, especially if the archives are user-friendly and easily accessible.

Long before a U.S. space mission receives a "new start" approval, data specialists from NASA's National Space Science Data Center (NSSDC) are assigned to work with the project office to help in developing the first version of a Project Data Management Plan. This plan defines the projected availability of the data and supporting information on a timely basis for use by the science community, the data rights of the scientists directly involved in the mission, and the proprietary-use period. The data plan serves as an interface document between data producers and archivers so that proper software, manpower, equipment, and other requirements can be provided; it also enables resource planners to assure adequate and timely funding for the archive and distribution facilities. The data plan continues to be developed during the study phase of the project and thereafter. Such a document is required for all NASA missions.

One of the most important sections of the data plan describes the derived data products and catalogs. The formats and standards which will be followed are specified. The archive catalogs usually include the planned and completed observations, the data logs, and a list of sources detected above a certain threshold. Typical data which might be produced include tables of observational results, images in various energy bands, exposure maps, and background maps. Where appropriate, as in the case of surveys, there should be a cross-correlation against existing catalogs, the light curve during the observation in the same energy bands as the images, and the results of a simple timing and spectral analysis.

NASA's long-term data archive and distribution facility is the NSSDC, located at the Goddard Space Flight Center. Recently the High Energy Astrophysics Science Archive Research Center has also been established at Goddard. It supports multi-mission X-ray and gamma-ray archival research. Charged with the acquisition of archival data from the

and gamma-ray archival research. Charged with the acquisition of archival data from the major recent missions, such as the Broad Band X-Ray Telescope (BBXRT), the Roentgen Satellite (ROSAT), and the Gamma Ray Observatory (GRO), it will also be responsible for data from future missions, such as Astro-D and the X-ray Timing Explorer (XTE). Products to be acquired include the raw data, the derived data products, catalogs of results, the calibrations, the analysis software and the documentation. This archive research center also provides the scientific interface needed by the community to obtain meaningful results from the data.

Before data are permitted to enter the public domain, products from most missions are stored on a dedicated device, such as an optical disk "juke box," which is accessible only to authorized project personnel. Among the responsibilities of any astronomical archive are the creation/maintenance of the astronomical standard, such as FITS, for transporting data and keyword standards compatible with other astronomy missions.

In comparing characteristics of the archival systems for several well-known astronomy missions, one finds the most variety in the data base management systems which have been adopted. Most of these differences have developed because of historical reasons in that workers in a given spectral regime were familiar with certain ways of processing and archiving data. For instance, the data system INGRES has been adopted by three of the high-energy astrophysics missions, HEAO-2, ROSAT, and GRO. The Hubble Space Telescope (HST) uses Omnibase and Sybase, whereas the International Ultraviolet Explorer (IUE) depends on a DEC-based system known as REF-11. The IRAS uses the Cyber-based IM/DM, and the Cosmic Background Explorer (COBE) uses Datatrieve, a custom front end for DEC computers. Data from most of these missions are available on tape and optical disk. User access is over the major networks with NSI-DECNET being a common link. Output products vary from magnetic tape, microfiche and hard copy to optical disks and CD-ROMs.

In an effort to improve the handling of, and access to, the rapidly increasing astrophysical databases, two systems are being developed: the European Science Information System (ESIS) by the European Space Agency, and the Astrophysics Data System (ADS) by NASA. Both systems use high-speed communication links and a common set of analysis tools to make the astrophysics data holdings more broadly and efficiently accessible. By providing a common information system infrastructure for science analysis, the duplication of effort should be reduced while increasing the scientific return. A server/client architecture allows services and data to be added or replaced easily. It is expected that this modular, layered design will accommodate both anticipated and unanticipated growth and will thus enable smooth evolution of the hardware and software without interruption of service to the user community.

The ultimate value of archived data lies in the quality of the data placed in the archive, the documentation accompanying the data, the ease of access, and the convenience of distribution. Time spent by the mission planners in designing the archives for space astronomy data will undoubtedly be rewarded by better and more extensive use of the data by a larger number of astronomers who will benefit from access to data obtained by a variety of instruments across multi-wavelength regions.