

GLOBAL NETWORK ACCESS AND PUBLICATION OF SURVEY DATA

N.E. WHITE

*High Energy Astrophysics Science Archive Research Center,
Laboratory for High Energy Astrophysics, GSFC*

1. Introduction

The results of astronomical surveys include 1) catalogs containing anywhere from a few to many million objects, 2) data products used to generate the catalogs (*e.g.*, images or spectra), 3) publications and 4) object based compilations of information from many sources. The ubiquitous growth in the Internet and the dramatic reduction in the cost of mass storage systems now allows instant global access to this information. Astrophysics on-line services have grown up with the Internet, and represent an invaluable resource whose access is a routine part of any research project. Unfortunately users are also faced with searching and accessing multiple sites with different content, access and response methods. There can also be inconsistencies amongst the various systems, which can involve extra effort to resolve. A network-integrated astrophysics system has long been sought because it would remove multiple user interfaces and allow much simpler integration of services. In an era of shrinking budgets, the issue is how to achieve this in a cost effective manner. This review gives an overview of the current situation and discusses the likely evolution towards a network-integrated astrophysics system.

2. Current Capabilities

There are currently three levels of access to the results of scientific investigations 1) the literature search, 2) on-line catalog and data services, and 3) value added services. The literature search provides the top level published knowledge base, which typically denotes the start of any research project. The catalog and data services provide access to the raw catalogs

and data products that are the result of an investigation. A value-added service takes elements of the first two and combines them to produce a new product, *e.g.*, combining catalogs into a single object based compilation. In this section the current on-line astronomical services are discussed in terms of these classifications. This is not a comprehensive list of services available, but rather examples used to illustrate the three levels of service currently provided.

2.1. THE LITERATURE SEARCH

2.1.1. *Electronic Journals*

Over the past year the publishers of astronomical journals have started experimenting with providing WWW access to their publications. The *Astrophysical Journal* at <http://www.journals.uchicago.edu/ApJ/> provides one of the first examples of direct online access to a journal. This has proved a very effective and popular method for rapid dissemination of papers. We expect that all the major journals will follow this lead. One major advantage of these electronic publications is that they can include the original tables from the papers in a machine readable form, so they can be more readily integrated into databases. Ultimately, one might expect that links to the original data may also be included.

The Los Alamos National Laboratory (LANL) and other servers distribute preprints ahead of publication in the journals. They too have become very popular. However, in the near future one might hope that the turn around-time between acceptance in a journal and electronic publication will be reduced to a few days. This would make the preprint a relic of the past.

2.1.2. *The Astrophysics Data System Abstract Service* (http://adsabs.harvard.edu/ads_abstracts.html)

The NASA Astrophysics Data System (ADS) abstract service based at the Smithsonian Astrophysical Observatory (SAO) provides WWW access to astronomical journal abstracts going back to 1975. This system is becoming the glue that ties together the on-line journals. It is also taking the lead in scanning back issues of journals, so they are network accessible. The ADS currently includes 240,000 astronomical articles and 410,000 space instrumentation and engineering articles. A search engine allows a search by author, object name, words in the title, and/or abstract text. The results are ranked by how closely the paper matches the query and in many cases contain links to the scanned images of the original journal articles, SIMBAD object pages, and even in some cases to the original data in the archive. The ADS has given a major boost to the productivity of

astronomers. It is rapidly making obsolete the need to visit the library to make a literature search.

2.2. CATALOGUE AND DATA SERVICES

2.2.1. *CDS*

(<http://cdsweb.u-strasbg.fr/CDS.html>)

The Centre de Données astronomiques de Strasbourg (CDS) collects and distributes astronomical catalogs related to observations of stars and galaxies, as well as other galactic and extragalactic objects. At the time of this writing it includes 1622 catalogs of which 1182 are available via ftp in ASCII or FITS format. Very recently a new catalog browser called VizieR has made it possible to search 995 catalogs via the a WWW interface. The search options include by coordinates, object name, mission name, wavelength and other catalog parameters.

2.2.2. *Astronomical Data Center (ADC)*

(<http://adc.gsfc.nasa.gov>)

The Astronomical Data Center (ADC) at the NASA GSFC acquires, verifies, formats, documents, and distributes files containing astronomical data in computer readable format. The current holding is 800 astronomical catalogs plus 750 data tables from various astronomical journals. It is possible to make a search via a Web interface for catalogs and then download them directly from the ftp site. The ADC has produced two CD ROMs with the most often requested catalogs. These CD ROMS have proved very popular.

2.2.3. *HEASARC*

(<http://heasarc.gsfc.nasa.gov>)

Access to data and catalogs from 22 different X-ray and Gamma-ray astrophysics satellite observatories is provided by the High Energy Astrophysics Science Archive Research Center (HEASARC) at NASAs GSFC. There is ftp access to over 500 Gb of data in FITS format from both active missions (ASCA, CGRO, ROSAT, RXTE) and newly restored data (*e.g.*, HEAO-1, Einstein, EXOSAT, Ariel V, Vela 5B, Cos B and SAS-2). The archive can be searched via a user interface called Browse (command line) and W3Browse (WWW interface). The HEASARC database also contains many of the most popular astronomical catalogs. The command line and WWW versions both allow a remote user to search by coordinates, name and all database parameters. A key feature of the HEASARC browse system is to link the catalogs to the data. This allows the user to view data, make quicklook analysis and retrieve selected items that were used to generate the catalog. Also possible in the command line version are

cross-correlations between catalogs, plots and filtering of parameters (these will appear in the Web version in early 1997).

2.2.4. *Astrophysics Multi-spectral Archive Search Engine (AMASE)* (<http://amase.gsfc.nasa.gov>)

The Astrophysics Multi-spectral Archive Search Engine (AMASE) is a WWW accessible multi-mission and multi-spectral catalog being jointly developed by the GSFC Astrophysics Data Facility and the University of Maryland. It allows the search of space mission data in the NASA public archives. It is built using an object database methodology. This is currently a prototype system and so only has IRAS, ROSAT and a few other catalogs. It is a hybrid system where some value is added by trying to use an object orientated system to link catalogs from different wavelength regimes to the data archives. This system can be searched using a WWW interface by coordinates and object name.

2.3. VALUE ADDED SERVICES

2.3.1. *The SIMBAD* (<http://cdsweb.u-strasbg.fr/SIMBAD.html>) and *NED* (<http://www.ipac.nasa.gov/ned/ned.html>) *Object Databases*

SIMBAD provided by CDS was the first and is the most comprehensive of the online astronomical services. It is an object based compilation of basic data, cross-identifications, observational measurements and bibliographic information. It contains all non-solar system objects (stars, galaxies and nonstellar objects). SIMBAD provides information on an object by object basis, and in doing so has integrated the information in lower level catalogs together with an ongoing object based literature search. SIMBAD can be queried via telnet, an X-windows client-server interface or by e-mail. Very recently a Web interface was also made available. Using these interfaces it is possible to search by object name or position. The results contain detailed information about the object, as well as a bibliographic listing.

The NASA Extragalactic Database (NED) contains an object based compilation of extragalactic objects based on combining catalogs and published results in the literature. Also available are the notes from catalogs and abstracts from 1988 onwards. The NED database has rapidly become an essential tool for extragalactic astronomers. It can be searched by name, near name, position, literature reference and various object parameters. Recently a skyplot and spectral energy distribution plot functions were added. NED is accessible via telnet, X-window and e-mail. A WWW interface is planned for the near future.

While NED and SIMBAD are invaluable tools, they are not well optimized for use with large surveys, *e.g.*, the ROSAT X-ray sky survey where

the identification of a large number of sources is required. A typical starting point is to cross-correlate a new source list against SIMBAD and NED. But the output from such a search presents a number of problems, that can be very time consuming to resolve.

- NED and SIMBAD are dynamic systems, that are constantly being updated and corrected. This makes it difficult to obtain a reproducible result at different times. A static catalog available via ftp that represents a snapshot of their current object catalogs at a particular time would make it much simpler to track objects that have been deleted, changed or added.
- The two databases can give inconsistent results for the same object. It is essential that the two teams establish a procedure to make regular consistency checks and correct errors.
- The output from the two systems are poorly structured. A standardized tabular form output would make it much easier to utilize these results within a large survey.

2.3.2. *Skyview*

(<http://skyview.gsfc.nasa.gov/skyview.html>.)

Skyview is a virtual telescope that can provide, on demand, images from any portion of the sky at wavelength regimes from radio to gamma-ray. It is the first example of an astronomical service that dynamically generates data on the fly in the form that is required by the astronomer. It makes trivial the once complex operation of obtaining views and overlays of the sky in different wavebands. Skyview currently contains 21 surveys accessible. The user selects which part of the the sky, the scale, the coordinate system, map projections, and equinoxes. In a more advanced interface they can also make contour, image and catalog overlays. The catalogs are provided via a client-server interface to the HEASARC database. The latest release allows image manipulation using a Java based interface.

3. Source Classifications

A typical goal of a survey is object classification. This can be done in a number of ways including 1) directly determining the class via source properties, *e.g.*, taking a spectrum, 2) making a cross-correlation with existing catalogs, and 3) by requesting telescope time to optically identify a source detected in another band. The major problem with source classification is that there is no standard classification scheme. Also most existing schemes are character based, which is not efficient for a database search engine. This makes class based searches of databases very difficult and gives inhomogeneous results. To circumvent these problems different numerical schemes have been

introduced by both the IUE and the EXOSAT projects. The EXOSAT class scheme (originally developed by P. Giommi) is currently in use at the HEASARC. It assigns each catalog entry a four digit numeric code to represent its classification. The first digit describes the global classification (*e.g.*, AGN or star). The following digits assign further classifications or properties such as spectral type, or AGN sub-class. Each sub-class is chosen to contain a unique set of properties. This class system, which is more fully described at <http://heasarc.gsfc.nasa.gov/docs/xray/class.html> provides the ability to both search using numerical methods and to allow combinations of properties. It is however limited to a few sub-classes and does not allow complex class combinations. To make further progress in this important area requires an international agreement on a numerical based classification system. This system could then be adopted by the IAU in a similar fashion to the standards adopted for source naming conventions, for example.

4. A Network Integrated Astronomy System

It has been long recognized that having a single access point to astronomical catalogs and data potentially could increase the productivity of astronomers. In the late 1980s NASA created the Astrophysics Data System (ADS) and ESA the European Space Information System (ESIS) with the objective to create a layer that connected together the diverse astronomy services available on the network at that time. But both projects failed to meet their objectives and were not considered a success. In the case of ESIS the project was discontinued. The ADS program was descoped, leaving only the abstract service, the only feature of the system that became widely used. A shortcoming of these global services was the fact that they attempted to compete against the existing, and well established, user interfaces. There was insufficient incentive to use them over the original more familiar interface. Another issue was the lack of standards to connect to data files with the database tables that are used to access them, which made it difficult to access data through any generic interface. ADS was eventually overtaken by the WWW, which effectively made redundant the machine dependent and vendor specific user interface that had been developed. After spending substantial sums of money, both NASA and ESA retreated from trying to provide a unified service as an expensive moving target that was not widely supported by the community they were designed to serve.

Despite the failure to integrate the existing astrophysics services together by NASA and ESA we (and the funding agencies) should not be discouraged. The Web browser addresses the problem that ADS and ESIS

were unable to solve, *i.e.*, a simple multi-platform interface that the astrophysics service providers are willing to adopt. The current WWW services are not yet as capable as the originals, because of the stateless nature of the WWW client-server interaction. But this is being solved with the development of Java and plug-ins, which will allow fully interactive interfaces. The development of the URL, which underpins the success of the WWW, provides the mechanism required to achieve a network integrated astrophysics system. The URL is a networked extension of the filename concept. The HEASARC has already assigned each dataset in its archive a unique URL, which specifies the path to the data which can be directly accessed by any user interface. In addition to this, the URL can be used to access programs and scripts over the network giving transparent access to specialized services. By making available URLs that access given services, both data centers *and* users can develop user interfaces that are best suited to their applications. The most simple example is the name-coordinate resolver services already provided by NED and SIMBAD. This is a simple network service that given an object name returns the coordinates, and is commonly found integrated within other on-line services. In the future one can expect an increasing number of services that are cross linked to provide access to a resource on the network, *e.g.*, calling up images from skyview, within NED. This approach will naturally lead to a user-driven network-integrated astrophysics system that will both increase data access and boost astronomer productivity.

Acknowledgements

I thank Tom McGlynn, Sherri Calvo, and Gunther Hasinger for useful comments and discussions.