ACCESS TO ASTRONOMY DATA AND SERVICES VIA THE ASTROPHYSICS DATA SYSTEM (ADS)

Guenther Eichhorn
Smithsonian Astrophysical Observatory
60 Garden Street
Cambridge, MA 02138

Presented at the October 1993 Meeting

Abstract

The goal of the Astrophysics Data System (ADS) is to provide the astronomy researcher with easy access to a wide variety of astronomical data and services. The objective of the ADS is to explore the unique value of each supplier of data and services and to provide access to that value by the science community at large via the Internet. The ADS is a NASA-sponsored project and is available free of charge to the astronomical community. Data currently available include over 250 catalogues, access to data archives—e.g., IPAC/JPL's IRAS all-sky survey, SAO's Einstein archive, NSSDC's NDADS archive—access to over 160,000 abstracts from astronomy and astrophysics journals, and access to the SIMBAD astronomical objects database in Strasbourg, France. Services available include a local table editor, plot and image display services, and a coordinate conversion tool.

1. Introduction

The idea of a system to provide access to provide general astronomical data was conceived in the 1980s, when concerns arose about the lack of tools for accessing the mushrooming amount of data from NASA's space missions. First discussions about the implementation of a data access system were held in 1988. After several workshops, NASA funded a project to develop such a data access system: the Astrophysics Data System (ADS).

Early on it was clear that it was not feasible to collect all the data and store them in one central location. We therefore designed a distributed data system that maintains the data at the data centers that produced them. The user has a client software package that runs on his/her workstation that can access these data holdings and transfer data to the local computer on request. This architecture allows the data centers to keep their data up-to-date without the need to redistribute the data when they have been changed (e.g., after implementing improved calibration procedures or after adding more data). The user always has the latest version of the data available.

The first version of the ADS was released in 1991, and a revision appeared in 1992. Both used a character-based user interface. The third release in November 1992 introduced a graphical interface which was much more user-friendly. The latest release, in April 1993, provided access to 200 catalogues and 150,000 abstracts from astronomy and astrophysics journals. It also included a data plotting service.

The latest release in January 1994 increased the number of catalogues to about 250 and added access to data archives at the Infrared Processing and Analysis Center (IPAC) at the Jet Propulsion Laboratory (the IRAS all-sky survey), the Smithsonian Astrophysical Observatory (the Einstein data archive), and the National Space Science Data Center (NSSDC) at the Goddard Space Flight Center (NSSDC Data Archive and Distribution Service [NDADS]).

The ADS's goal is to enable multi-spectral science by making available a wide range of data and services. However, the ADS is not a data analysis system. It provides access to the data and permits previewing data and performing simple data manipulations. For in-depth data analysis there are several specialized systems, such as IRAF, AIPS, and MIDAS. The data retrieved through ADS are available in FITS format, a standard in the astronomical community.

2. System Architecture

The ADS is a distributed system communicating across the Internet. It is based on a client/server architecture. The users have the client software on their local workstation. The servers are at the data center's computers, so the users do not have to store the complete data set at their local computer. They need to retrieve only the data that are needed.

Figure 1 shows a schematic of the system architecture. The left part of the figure is the client part, the right side shows the server side. The gray boxes show software distributed by the ADS project. The white boxes are software developed by the data centers. These programs are customized for each specific database at the data centers.

When a user logs into the ADS, his/her password is checked by the Authenticator. The password transmission is in encrypted form for security reasons. We use the KERBEROS authentication system to handle all security aspects of the ADS.

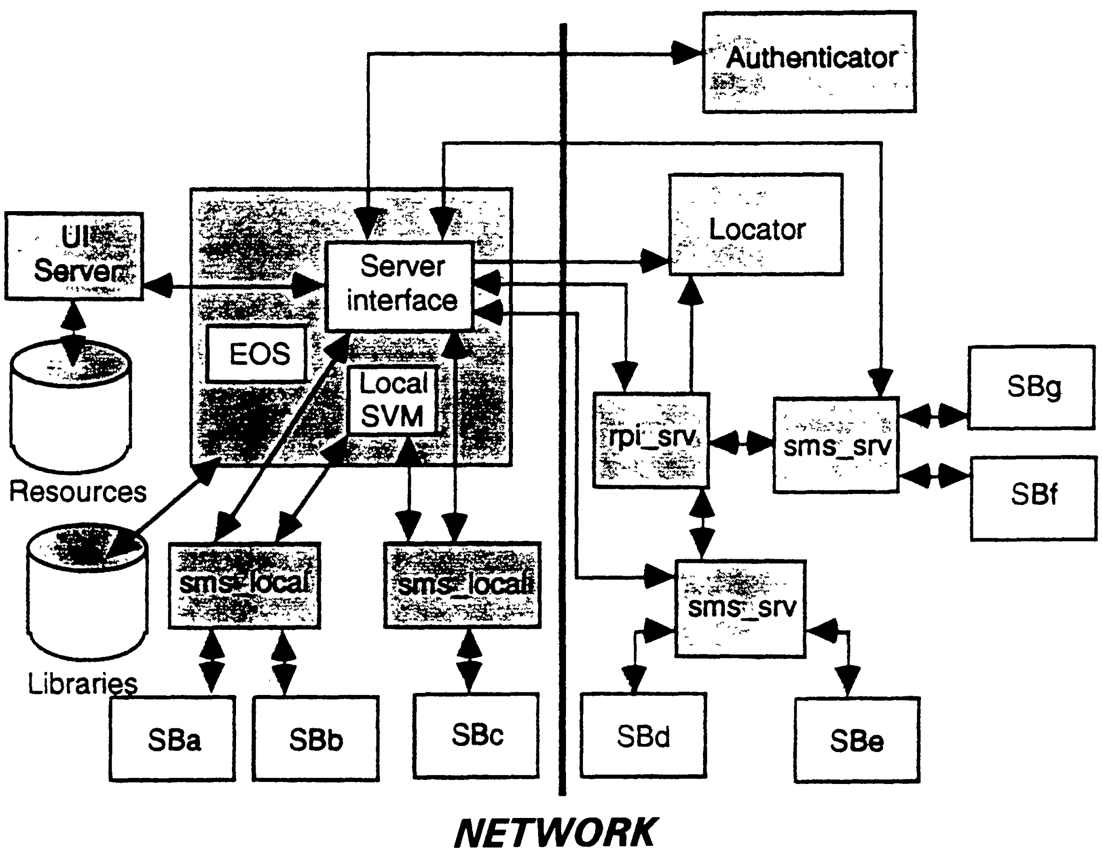

Figure 1. ADS System Architecture. The gray parts are software distributed by ADS. The white parts are software developed by the data centers. EOS (Ellery Open Systems) is the underlying operating software that handles the interaction between the user and the network. The schematic is described in detail in the text.

The Locator is a server that locates data and services. Whenever a data server or other service is started up by running the Remote Process Invocation Service (*rpi_srv*), it broadcasts its existence to the Locators (several for redundancy) along with a property list of the service. Subsequently, any client may find this service through any of the locators.

The client part of the system is started when a user requests an ADS session. The EOS (Ellery Open Systems) kernel requests a password from the user. After authentication is completed, the graphical user interface (GUI) is started and the main window is displayed. The user interacts with the system through this graphical user interface.

When the user requests a service, the system checks whether it is a local or remote service. If it is a local service, a local Session Manager (sms_local) is started. The sms_local in turn starts the service (e.g., SBa in Figure 1). The user then interacts with this service on the local machine. If the service is a remote service, the client contacts a Locator to find the location of this service. It then contacts the Remote Process Invocation Service on the remote computer. The rpi_srv starts a Session Manager Server (sms_srv) on the remote machine and forwards its address to the client. From then on the client interacts directly with that sms_srv. The sms_srv in turn starts the program that provides the service. This program is called a Server Body. This server body performs the actual service, be it data base access, file transfer, or remote calculations, to name a few.

When a session is terminated, the client sends a quit command to shut down the remote server body and sms_srv and then closes the network connection. The communication between the GUI and the server body is through ASCII commands. The server body receives commands through Standard Input and returns results through Standard Output. This makes it very easy to develop and debug a server body interactively from the terminal. It also makes it easy to incorporate an existing service into ADS since most services can already communicate through a terminal interface. The user interfaces are developed within ADS. They use X/Motif as the underlying windowing software. ADS contains an interpreter called C-Lite. This interpreted language is used to write the user interface code. It closely resembles C, making it very easy to develop and debug GUI code.

3. Examples

Following are some examples of using the ADS. The first set of screen dumps are from querying the AAVSO catalogue, which is hosted at the Smithsonian Astrophysical Observatory. This catalogue contains information about data available from AAVSO. Figure 2 shows the main ADS window (top left), the catalogue selection window with the AAVSO catalogue selected (top right) and the catalogue description window. The main ADS window includes pulldown menus to access all the capabilities as well as "Quick Buttons" to access quickly some of the frequently used functions. The catalogue descriptions are served from an ADS node. To get and display these descriptions, we implemented a Mosaic client in ADS. Using public domain software allowed us to implement the document service without having to develop our own GUI/server combination.

Figure 2 shows the index to the AAVSO database, while Figure 3 shows a look at the AAVSO data. The top right window of Figure 3 is an ADS table editor window. This utility allows the user to manipulate catalogue tables. The window in the left middle shows an example of a table manipulation. In this case all entries in the table were selected that had a period greater than 0.0. This was done to eliminate entries that have no known period in the catalogue. The lower left window is a 2D-Plot widget. It allows the user to plot any 2 columns of a table against each other. In this case the period was plotted against the galactic latitude. The plot in the lower right shows that the very long period

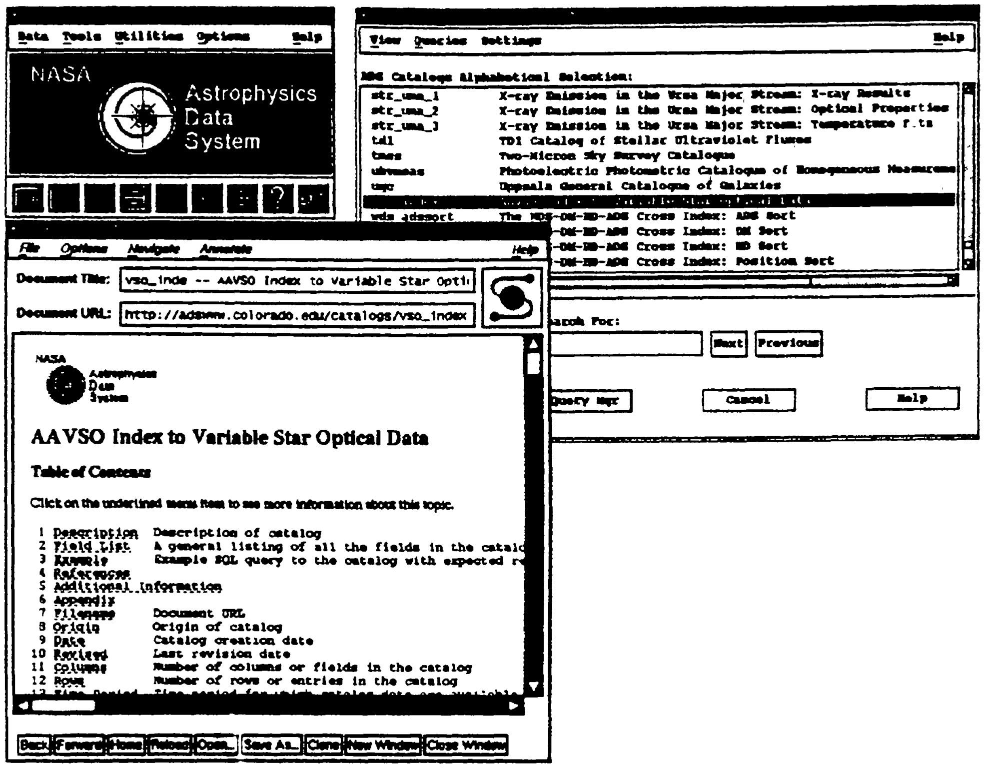

Figure 2. Screen dump of catalogue selection and description for AAVSO catalogue.

variables seem to be concentrated in the galactic plane. While producing these plots I noticed that one star, T CrB, stood out in the catalogue data. The catalogue specified its period as 29000 days. This seemed very long, so further checks were done to ascertain that this was not a bad catalogue entry.

Figure 4 shows a query to the SIMBAD database in Strasbourg, France. The top left window shows the catalogue with the entry for T CrB highlighted. The bottom left window is the query window for SIMBAD, querying by object name with T CrB specified. The top left window shows the result of this query. It has the name of the object, its class (Nova), its coordinates, magnitudes, aliases and available measurements. The measurement for a variable star (V*) is selected and requested. The result from this request is in the middle window on the right. It shows the period of the star again as 29000 days. The entry in the AAVSO catalogue is therefore consistent with other data sources. The SIMBAD service also provides literature references.

Figure 5 shows the result of a reference query. In the SIMBAD results window (lower left) the "Get Refs." button was used to query the SIMBAD server for references. The result is in the list of references in the top window. From this list, the connection between these references and the ADS abstract service was used. By highlighting a reference and requesting the full abstract, a query is sent to the abstracts service. If the abstract is available, it is retrieved and displayed (lower right window).

To complement this reference list, a direct query to the abstracts service was used (Figure 6). This was a text query for T CrB and cataclysmic variables (lower right window). The result is in the top window. The column with the score indicates how well the particular abstract matches the query. A score of 1 means that all the query elements were found in the abstract. In this case it means that the selected abstract contained the words "T," "CrB," "cataclysmic," and "variables." Scores less than 1 indicate that not all the conditions have been met and therefore are a measure of how relevant the abstract is to the query. The abstract service can in turn get references from SIMBAD by

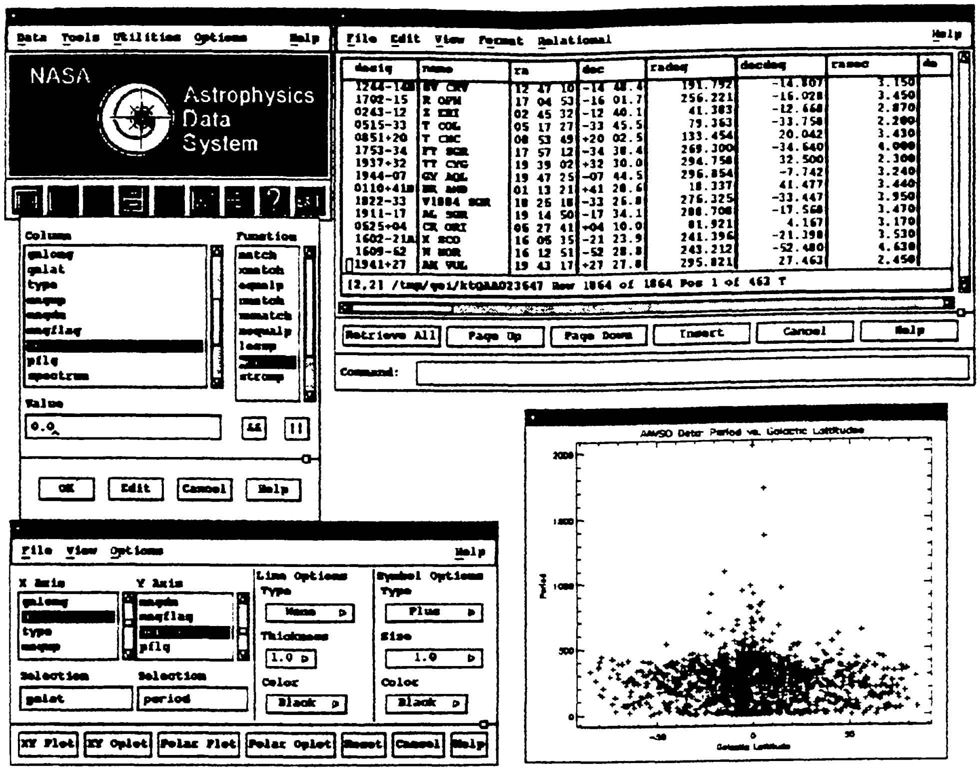

Figure 3. Screen dump of a plot from the AAVSO catalogue data.

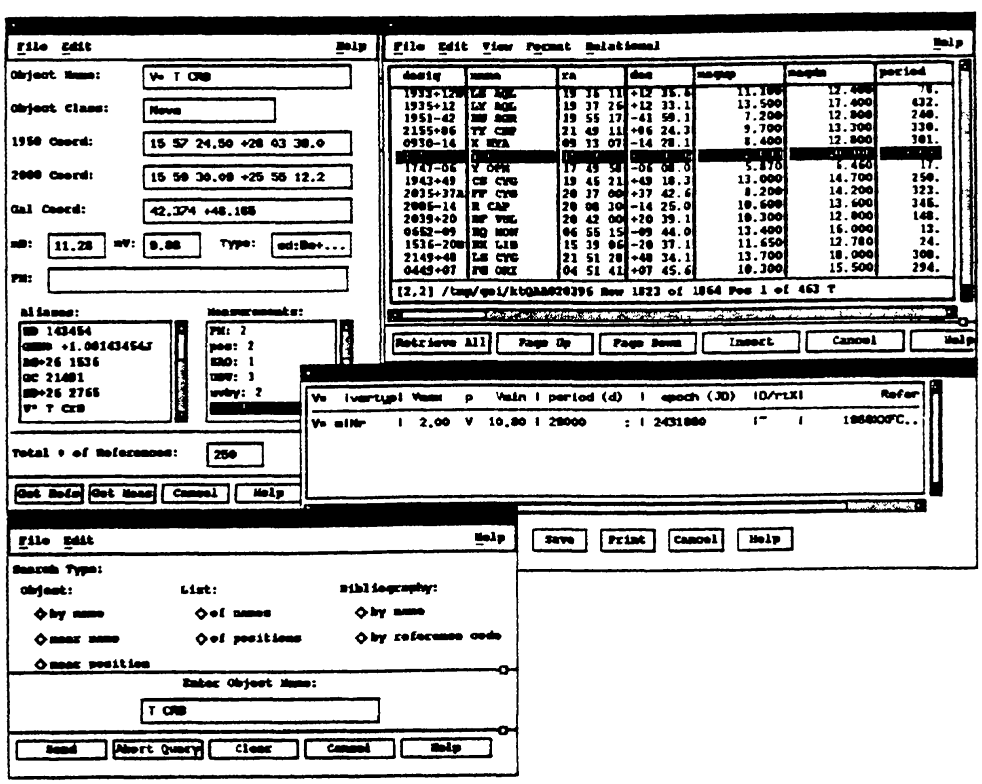

Figure 4. Screen dump of SIMBAD query.

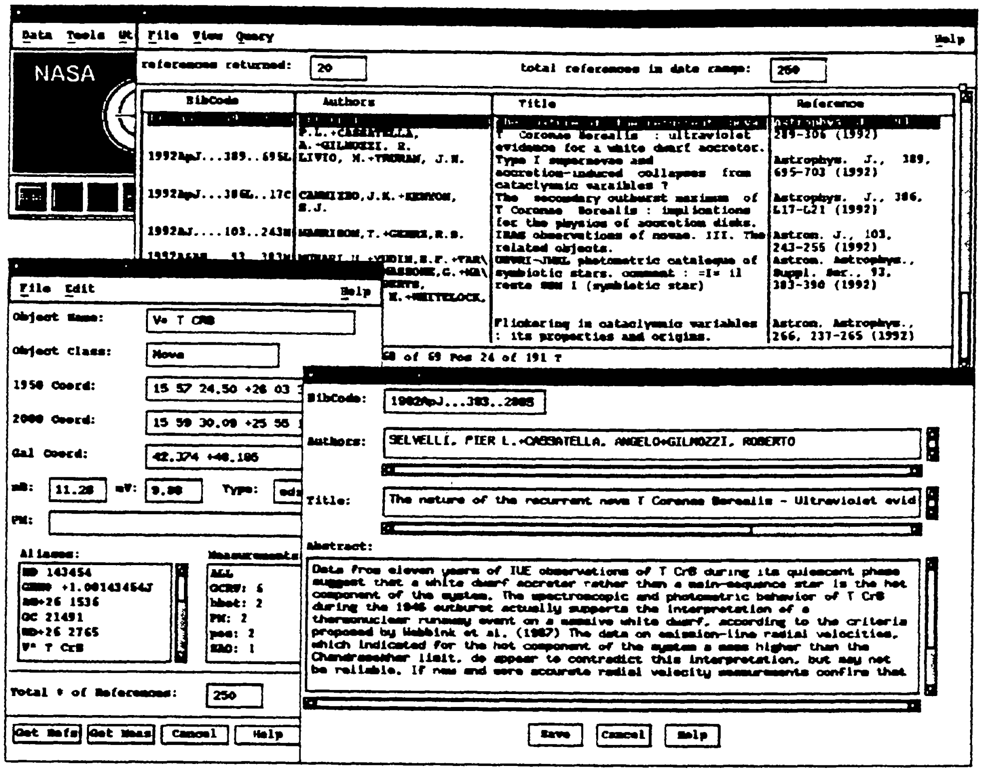

Figure 5. Screen dump of SIMBAD reference query.

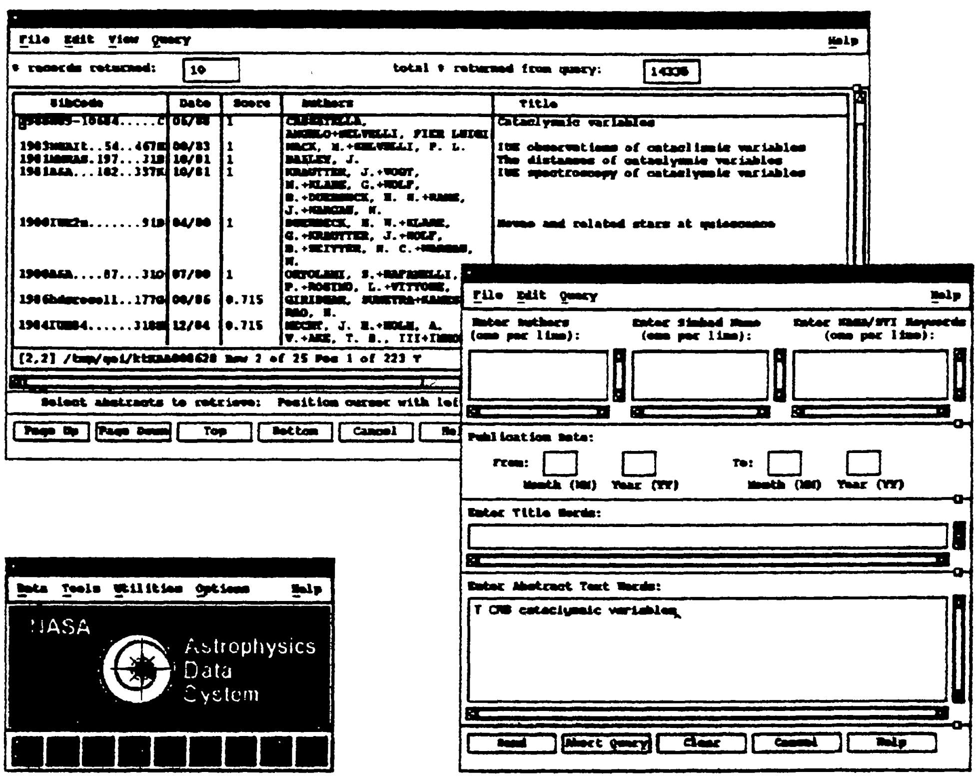

Figure 6. Screen dump of abstracts query.

specifying an object name (here T CrB).

The next set of screen dumps describes the use of an archive service and the simultaneous display of catalogue and image data. Figure 7 shows again the catalogue query window (top right) and the catalogue documentation window (left), this time for the CfA redshift catalogue. The lower right window shows a positional query. The position selected was for the Coma cluster of galaxies. This query returned about 750 objects in the area specified. This table was then separated in two tables with the table editor. One table contained the elliptical galaxies, the other the spirals. The coordinate conversion service was used to add columns with B1950 coordinates since the table itself contains only J2000 coordinates, whereas the Einstein archive uses B1950 coordinates.

Figure 8 shows an archive query to the Einstein data archive. The lower left window is the query window. It has selected IPC and HRI images. The query is by field title with "coma" as the query string. The result is in the upper left window. The query resulted in 8 files that matched the query. One of the images was selected and a transfer requested. The file transfer window is in the lower right. It shows that the transfer status is "done." The results of the last two queries are combined in Figure 9. The upper left window is the control window for one of the image display services. It displays the Einstein image of the coma cluster. The "Table Overplot" function was used to overplot the table with the elliptical galaxies (circles) and spiral galaxies (crosses). The display shows that the elliptical galaxies concentrate in the center of the cluster, whereas the spiral galaxies are distributed around the perimeter.

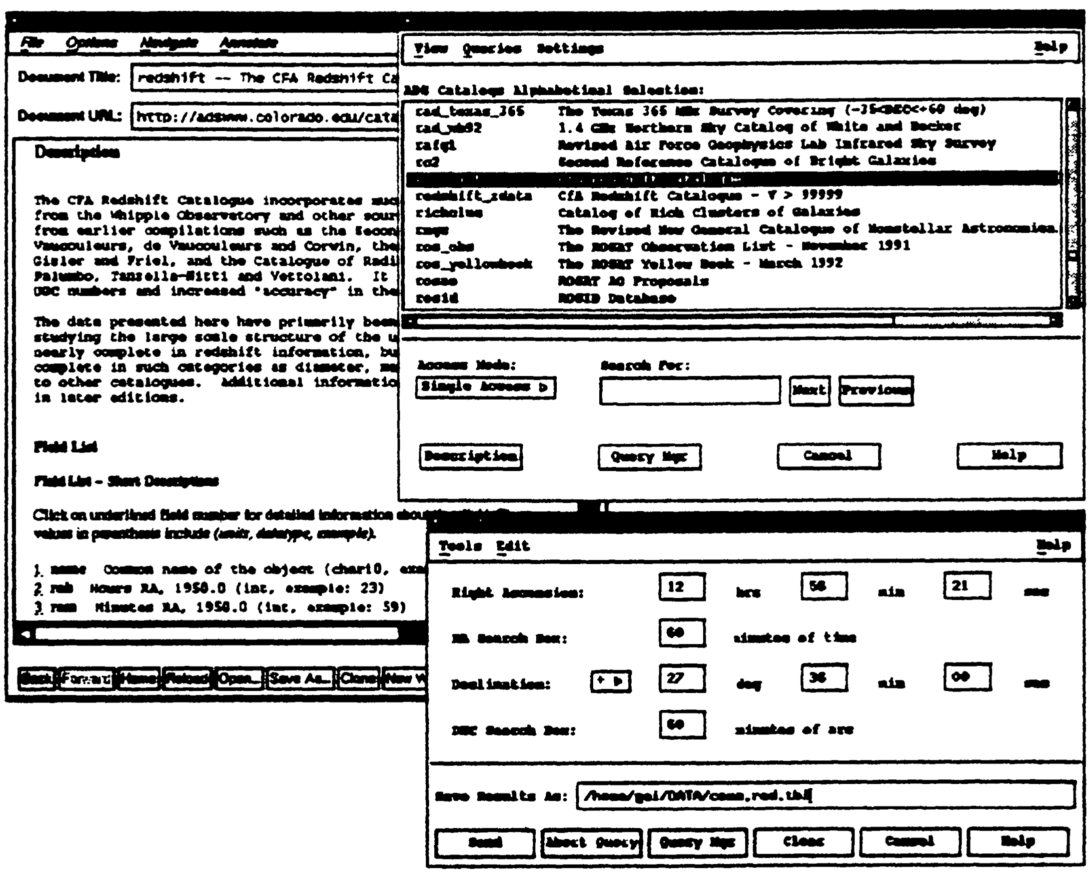

Figure 7. Screen dump of catalogue query for the CfA redshift catalogue with positional query window.

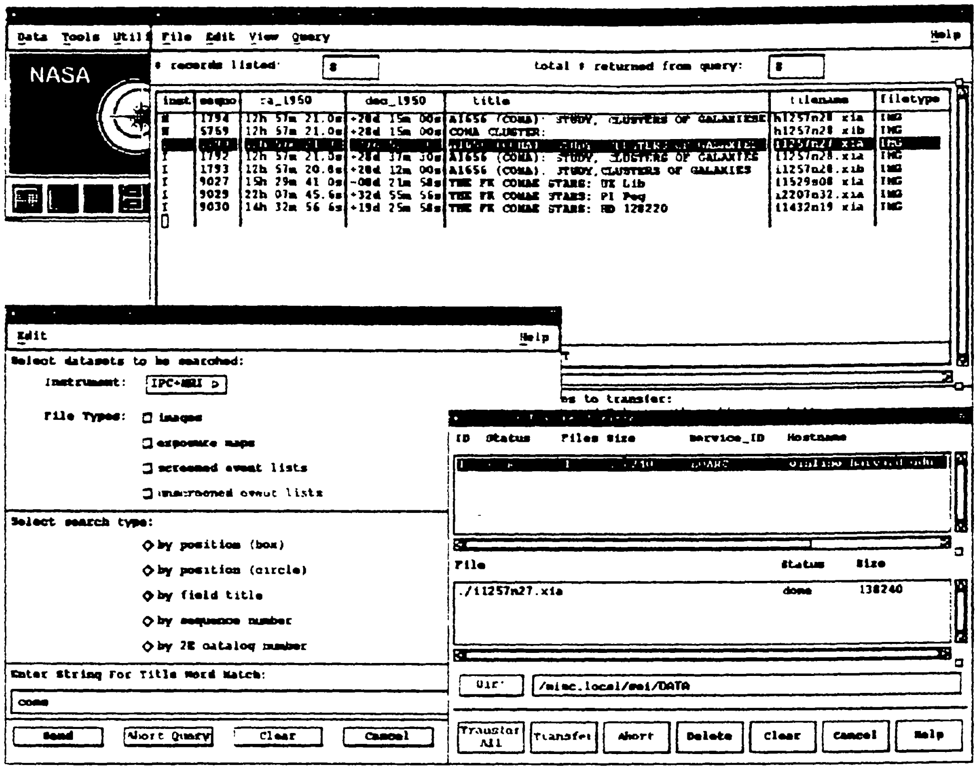

Figure 8. Screen dump of archive query for the Einstein data archive.

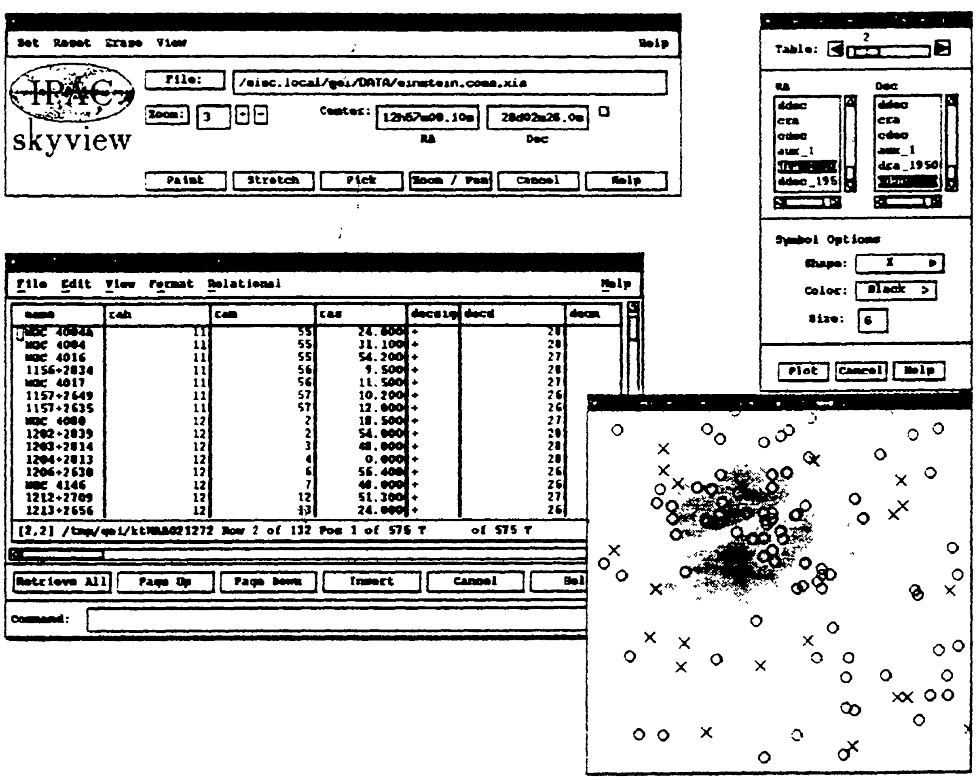

Figure 9: Screen dump of combined image and catalog display.

4. Summary

These were just two examples of what the ADS system can do. It is constantly expanding the number of data sets and services. The number of registered users is currently (December 1993) over 1,500. We have approximately 6,000 logins per month and over 10,000 different queries.

To get more information about ADS, here are some contact addresses:

For user and registration information contact:
Jacque Anderson
303-492-0466
ads@cuads.colorado.edu
CASA/University of Colorado
Campus Box 391
Boulder, CO 80309

For data supplier questions contact:
John Good
818-397-9539
jcg@ipac.caltech.edu
IPAC/Caltech
Pasadena, CA 91125

For programmatic and general questions contact:
Stephen Murray (Project Scientist)
617-495-7205
ssm@cfa.harvard.edu

or

Guenther Eichhorn (Project Manager)
617-495-7260
gei@cfa.harvard.edu
Smithsonian Astrophysical Observatory
60 Garden Street
Cambridge, MA 02138

5. Acknowledgments

This project, which is a cooperative venture of several different groups, is supported by NASA grant NCCW-0024. The major development groups are the Smithsonian Astrophysical Observatory, Cambridge, MA; the Center for Astrophysics and Space Astronomy at the University of Colorado, Boulder, CO; and at the Infrared Processing and Analysis Center at the Jet Propulsion Laboratory, Pasadena, CA. Other groups (too numerous to list here) are contributing both to the development and the data in the system. I want to express my appreciation to everybody involved with the ADS project.