

The Astrophysics Data System

Guenther Eichhorn, Alberto Accomazzi, Michael J. Kurtz, Carolyn S. Grant

Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138, USA, e-mail: gei@cfa.harvard.edu

Abstract. The NASA Astrophysics Data System has been very successful in providing the researcher and librarian the capability to effectively search the astronomical and space science literature from their desktop. It currently provides access to four searchable databases of scientific bibliographic references and a large archive of full-text documents which includes all the major astronomical journals.

1. Introduction

The Astrophysics Data System is a NASA funded project tasked with providing access to astronomical information through the World Wide Web (WWW). There are two main parts of the ADS, the Abstract Service and the Article Service. This article describes these two services and gives a brief overview of their contents and user interfaces.

2. Abstract Service

The Abstract Service provides access to four databases containing more than 1.1 million abstracts: Astronomy and Astrophysics, Instrumentation, Physics and Geophysics, and the Los Alamos preprints in Astronomy. Each database is separately searchable through different interfaces, allowing users to easily identify the information in which they are interested.

2.1. Data Holdings

As of April, 1998, the ADS databases contained the following datasets: Astronomy and Astrophysics (400,347 references), Space Instrumentation and Engineering (480,188 references), Physics and Geophysics (289,752 references), and the Los Alamos Preprints in Astronomy (2,442 references). The databases contain bibliographic references and abstracts from all the major journals and several minor journals, conference proceedings, NASA reports, and many PhD theses. Coverage for most of these publications goes back to 1975, but as more data is converted in digital form from hardcopy, coverage for most astronomy journals will be extended back to volume 1.

Each entry in the databases is uniquely identified by a 19-letter label called the bibliographic code (Schmitz et al. 1995) for that entry. The adoption of

this code as an identifier for bibliographies by the majority of the astronomical institutions and data centers has allowed the integration of the ADS resources with other data providers' services (Eichhorn 1998).

This allows, for instance, linking between the references in the ADS databases and online electronic journals, data tables, and astronomical object lists.

Over the years, the ADS databases have been expanded by incorporating references from different sources, while still maintaining information about their origin and grouping. Among them is worth mentioning approximately 100,000 entries from the SIMBAD astronomical database, over 90,000 entries from the Society for Photo-Optical Engineering, 18,000 entries from the Lunar and Planetary Institute database, 10,000 entries from the European Southern Observatory Astronomy Library, and 17,000 entries from the Library of Congress.

In early 1997 the AAS purchased from the Institute for Scientific Information (ISI), producer of the Science Citation Index, a list of references and citations for the major astronomical journals published between January 1981 and June 1996 for use by the ADS. For the articles that fall in this dataset, the ADS provides links to the list of their references and citations. However, because the cross-references used to create these lists are limited to this dataset, we cannot guarantee their completeness (not even in a specific time-span). We are working to add more cross-references on a regular basis by cooperating with the publishers of electronic journals in the creation of links to and from them, and by automatically recognizing references from the scanned images of older journal articles.

2.2. Query Interfaces

Browsing and searching the ADS bibliographies is possible through two different WWW interfaces: the Abstract Query form and the Table of Contents form.

The Abstract Query form allows users to perform sophisticated fielded queries by simply typing the name of the item to be searched for in the proper input field. Users can submit queries by author, object name, words in the title, and words in the abstract text. Following an intuitive syntax in use with other search engines, users can require that a particular term be present or missing from the list of results, or can use boolean constructs to formulate each fielded query.

Upon issuing a query, the ADS search engine ranks all the publications in the database against the input query by assigning them a relevance score. This score is computed as a weighted sum of the normalized frequencies of the search terms. The search interface provides default weights but also gives users great flexibility by letting them modify the value of such weights.

A set of "filters" to be applied to the resulting list can be specified to either eliminate or select only part of the results. For instance, users can specify to return bibliographies that have appeared only in refereed publications, or only in one or more specific journals. Another useful application is the possibility to require that the results from a query be part of one of the predetermined datasets mentioned earlier. For instance, users can require that the results from a query belong to the SPIE dataset, which in essence allows them to limit the search to the SPIE publication list.

By default the list of references that are displayed by the abstract service are sorted by descending score, the entries with highest scores being listed at the top of the list. A recently added option allows users to override this ordering by selecting one of the following sorting criteria: first author name, chronological order, inverted chronological order, date of entry of references in the database.

The Table of Contents form provides an easy-to-use interface that produces a list of references published in a particular journal given an input volume number or publication date (year and month). A separate graphical user interface is available to have the system return only the references published in the most recent issue of a journal or in the last issue not read by the user, depending on the user preferences setting.

2.3. Results List

A query to any of the ADS databases generates a listing of the references that were found to be matching the search criteria. For each reference in the list the following items are displayed: bibliographic code identifier, matching score, publication date, author names, title, and a list of hyperlinks showing the items available for that reference. Each item in the list of available items is represented by a letter as a shorthand notation for the type of data linked to (Eichhorn 1998). Because of the heterogeneous nature of our databases and the resources that are linked to and from it, following some of these above links will cause the user to be sent to a different site (e.g. when retrieving the electronic version of an article or information about the astronomical objects cited in the paper).

Options available at the bottom of the results list allow users to determine which format should be used to display the full information available for each reference. The formats currently available include: HTML tables, plain text, tagged records, Bib-TeX, and custom format, which allows users to specify the format string to be used for creating the output text. Also available at the bottom of the results list is the option of returning either the entire list of references or a subset of it sorted in a different way.

2.4. Display of Abstracts

The complete information available for each bibliographic reference can be displayed by following the bibliographic code link (1989ApJ...342L..71R) of each individual reference or by selecting several entries from a list of results and requesting that all their records be returned. Fields displayed for each publication typically include title, authors, journal, volume and page, publication date, bibliographic code. Keywords and abstract are also displayed when available.

Once a publication of interest is found, the user has the option to query the database for a list of papers similar to that one. This option (named "find similar abstract") is available at the bottom of each page displaying an individual abstract, and provides a powerful way to probe the database for information relevant to a published paper. The similarity criteria can be easily defined by the user by using the menu provided, which selects the records and database to be searched for.

3. Article Service

The ADS Article Service provides access to scanned images of astronomical journal articles in a variety of formats and resolutions. The project initially focussed on scanning the last 20 years of the major astronomical journals, seeking permission from the copyright holders and publishers. Subsequently, and as a result of the project's success, both smaller journals and older volumes started being digitized and made available. We currently have plans to place all the historical astronomical literature online (Kurtz & Eichhorn 1998).

The articles available online are cross-linked with their entries in the bibliographical databases, allowing users to download the full text of a bibliographic reference returned by a database search. Alternatively, a separate form listing the journals that have been scanned can be used to access an individual paper by specifying journal, volume and page to be displayed.

Users having access to a graphical World Wide Web browser can view the contents of each article page as a sequence of grayscale images on their screen. Different options are available for printing either an entire article or a single page, depending on the type of printing capabilities and bandwidth available to the user.

4. Summary

The Astrophysics Data System as described in Eichhorn 1997 and Kurtz & Eichhorn 1998 and this article is a powerful tool for the working astronomer in need of researching the literature and accessing other on-line information sources. It enables scientists to easily search the literature and data associated with it. With the inclusion of a large part of the historical astronomical literature in the near future the ADS will extend these capabilities significantly.

The success of the project is exemplified in the deep penetration that it has had in the astronomical community, changing the way astronomers keep abreast of their field. Access statistics for the Abstract Service and the Article Service are described in Eichhorn 1998.

Acknowledgments. This work is funded by the NASA Astrophysics Program under grant NCCW-0024.

References

- Eichhorn, G. 1998, Linking in the Astrophysics Data System, these proceedings, 118
- Eichhorn, G. 1997, The Digital Library of the Astrophysics Data System, *Ap&SS*, 247, 189
- Kurtz, M. J. & Eichhorn, G. 1998, The Historical Literature of Astronomy, via ADS, these proceedings, 293
- Schmitz, M., Helou, G., Dubois, P., LaGue, C., Madore, B., Corwin Jr., H. G., & Lesteven, S. 1995, in *Information & On-line Data in Astronomy*, eds. D. Egret & M. A. Albrecht (Dordrecht: Kluwer Acad. Publ.), 271