

= Astrophysics Data System =

The Astrophysics Data System (ADS), developed by the National Aeronautics and Space Administration (NASA), is an online database of over eight million astronomy and physics papers from both peer reviewed and non @-@ peer reviewed sources . Abstracts are available free online for almost all articles , and full scanned articles are available in Graphics Interchange Format (GIF) and Portable Document Format (PDF) for older articles . New articles have links to electronic versions hosted at the journal 's webpage , and these are typically available only by subscription (which most astronomy research facilities have) . It is managed by the Harvard ? Smithsonian Center for Astrophysics .

ADS is a powerful research tool and has had a significant impact on the efficiency of astronomical research since it was launched in 1992 . Literature searches that previously would have taken days or weeks can now be carried out in seconds via the ADS search engine , custom @-@ built for astronomical needs . Studies have found that the benefit to astronomy of the ADS is equivalent to several hundred million US dollars annually , and the system is estimated to have tripled the readership of astronomical journals .

Use of ADS is almost universal among astronomers worldwide , and therefore ADS usage statistics can be used to analyze global trends in astronomical research . These studies have revealed that the amount of research an astronomer carries out is related to the per capita gross domestic product (GDP) of the country in which he / she is based , and that the number of astronomers in a country is proportional to the GDP of that country , so the total amount of research done in a country is proportional to the square of its GDP divided by its population .

= = History = =

For many years , a growing problem in astronomical research (as in other academic disciplines) was that the number of papers published in the major astronomical journals was increasing steadily , meaning astronomers were able to read less and less of the latest research findings . During the 1980s , astronomers saw that the nascent technologies which formed the basis of the Internet could eventually be used to build an electronic indexing system of astronomical research papers which would allow astronomers to keep abreast of a much greater range of research .

The first suggestion of a database of journal paper abstracts was made at a conference on Astronomy from Large Data @-@ bases held in Garching bei München in 1987 . Initial development of an electronic system for accessing astrophysical abstracts took place during the following two years ; in 1991 discussions took place on how to integrate ADS with the SIMBAD database , containing all available catalog designations for objects outside the solar system , to create a system where astronomers could search for all the papers written about a given object .

An initial version of ADS , with a database consisting of 40 papers , was created as a proof of concept in 1988 , and the ADS database was successfully connected with the SIMBAD database in the summer of 1993 . The creators believed this was the first use of the Internet to allow simultaneous querying of transatlantic scientific databases . Until 1994 , the service was available via proprietary network software , but it was transferred to the nascent World Wide Web early that year . The number of users of the service quadrupled in the five weeks following the introduction of the ADS web @-@ based service .

At first , the journal articles available via ADS were scanned bitmaps created from the paper journals , but from 1995 onwards , the Astrophysical Journal began to publish an on @-@ line edition , soon followed by the other main journals such as Astronomy and Astrophysics and the Monthly Notices of the Royal Astronomical Society . ADS provided links to these electronic editions from their first appearance . Since about 1995 , the number of ADS users has doubled roughly every two years . ADS now has agreements with almost all astronomical journals , who supply abstracts . Scanned articles from as far back as the early 19th century are available via the service , which now contains over eight million documents . The service is distributed worldwide , with twelve mirror sites in twelve countries on five continents , with the database synchronized by means of weekly updates

using rsync , a mirroring utility which allows updates to only the portions of the database which have changed . All updates are triggered centrally , but they initiate scripts at the mirror sites which " pull " updated data from the main ADS servers .

= = Data in the system = =

Papers are indexed within the database by their bibliographic record , containing the details of the journal they were published in and various associated metadata , such as author lists , references and citations . Originally this data was stored in ASCII format , but eventually the limitations of this encouraged the database maintainers to migrate all records to an XML (Extensible Markup Language) format in 2000 . Bibliographic records are now stored as an XML element , with sub @-@ elements for the various metadata .

Since the advent of online editions of journals , abstracts are loaded into the ADS on or before the publication date of articles , with the full journal text available to subscribers . Older articles have been scanned , and an abstract is created using optical character recognition software . Scanned articles from before about 1995 are usually available free , by agreement with the journal publishers .

Scanned articles are stored in TIFF format , at both medium and high resolution . The TIFF files are converted on demand into GIF files for on @-@ screen viewing , and PDF or PostScript files for printing . The generated files are then cached to eliminate needlessly frequent regenerations for popular articles . As of 2000 , ADS contained 250 GB of scans , which consisted of 1 @,@ 128 @,@ 955 article pages comprising 138 @,@ 789 articles . By 2005 this had grown to 650 GB , and is expected to grow further , to about 900 GB by 2007 . No further information has been published .

The database initially contained only astronomical references , but has now grown to incorporate three databases , covering astronomy (including planetary sciences and solar physics) references , physics (including instrumentation and geosciences) references , as well as preprints of scientific papers from arXiv . The astronomy database is by far the most advanced and its use accounts for about 85 % of the total ADS usage . Articles are assigned to the different databases according to the subject rather than the journal they are published in , so that articles from any one journal might appear in all three subject databases . The separation of the databases allows searching in each discipline to be tailored , so that words can automatically be given different weight functions in different database searches , depending on how common they are in the relevant field .

Data in the preprint archive is updated daily from the arXiv , the main repository of physics and astronomy preprints . The advent of preprint servers has , like ADS , had a significant impact on the rate of astronomical research , as papers are often made available from preprint servers weeks or months before they are published in the journals . The incorporation of preprints from the arXiv into ADS means that the search engine can return the most current research available , with the caveat that preprints may not have been peer reviewed or proofread to the required standard for publication in the main journals . ADS 's database links preprints with subsequently published articles wherever possible , so that citation and reference searches will return links to the journal article where the preprint was cited .

= = Software and hardware = =

The software runs on a system that was written specifically for it , allowing for extensive customization for astronomical needs that would not have been possible with general purpose database software . The scripts are designed to be as platform independent as possible , given the need to facilitate mirroring on different systems around the world , although the growing use of Linux as the operating system of choice within astronomy has led to increasing optimization of the scripts for installation on that platform .

The main ADS server is located at the Harvard @-@ Smithsonian Center for Astrophysics in Cambridge , Massachusetts , and is a dual 64 @-@ bit X86 Intel server with two quad @-@ core 3 @.@ 0 GHz CPUs and 32 GB of RAM , running the CentOS 5 @.@ 4 Linux distribution . Mirrors are

located in Brazil , China , Chile , France , Germany , India , Indonesia , Japan , Russia , South Korea , United Kingdom , and the Ukraine .

= = Indexing = =

ADS currently receives abstracts or tables of contents from almost two hundred journal sources . The service may receive data referring to the same article from multiple sources , and creates one bibliographic reference based on the most accurate data from each source . The common use of TeX and LaTeX by almost all scientific journals greatly facilitates the incorporation of bibliographic data into the system in a standardized format , and importing HTML @-@ coded web @-@ based articles is also simple . ADS utilizes Perl scripts for importing , processing and standardizing bibliographic data .

The apparently mundane task of converting author names into a standard Surname , Initial format is actually one of the more difficult to automate , due to the wide variety of naming conventions around the world and the possibility that a given name such as Davis could be a first name , middle name or surname . The accurate conversion of names requires a detailed knowledge of the names of authors active in astronomy , and ADS maintains an extensive database of author names , which is also used in searching the database (see below) .

For electronic articles , a list of the references given at the end of the article is easily extracted . For scanned articles , reference extraction relies on OCR . The reference database can then be " inverted " to list the citations for each paper in the database . Citation lists have been used in the past to identify popular articles missing from the database ; mostly these were from before 1975 and have now been added to the system .

= = Coverage = =

The database now contains over eight million articles . In the cases of the major journals of astronomy (Astrophysical Journal , Astronomical Journal , Astronomy and Astrophysics , Publications of the Astronomical Society of the Pacific and the Monthly Notices of the Royal Astronomical Society) , coverage is complete , with all issues indexed from number 1 to the present . These journals account for about two @-@ thirds of the papers in the database , with the rest consisting of papers published in over 100 other journals from around the world , as well as in conference proceedings .

While the database contains the complete contents of all the major journals and many minor ones as well , its coverage of references and citations is much less complete . References in and citations of articles in the major journals are fairly complete , but references such as " private communication " , " in press " or " in preparation " cannot be matched , and author errors in reference listings also introduce potential errors . Astronomical papers may cite and be cited by articles in journals which fall outside the scope of ADS , such as chemistry , mathematics or biology journals .

= = Search engine = =

Since its inception , the ADS has developed a highly complex search engine to query the abstract and object databases . The search engine is tailor @-@ made for searching astronomical abstracts , and the engine and its user interface assume that the user is well @-@ versed in astronomy and able to interpret search results which are designed to return more than just the most relevant papers . The database can be queried for author names , astronomical object names , title words , and words in the abstract text , and results can be filtered according to a number of criteria . It works by first gathering synonyms and simplifying search terms as described above , and then generating an " inverted file " , which is a list of all the documents matching each search term . The user @-@ selected logic and filters are then applied to this inverted list to generate the final search results .

= = Author name queries = =

The system indexes author names by surname and initials , and accounts for the possible variations in spelling of names using a list of variations . This is common in the case of names including accents such as umlauts and transliterations from Arabic or Cyrillic script . An example of an entry in the author synonym list is :

AFANASJEV , V
AFANAS ? EV , V
AFANAS ? IEV , V
AFANASEV , V
AFANASYEV , V
AFANS ? IEV , V
AFANSEV , V

== Object name searches ==

The capability to search for papers on specific astronomical objects is one of ADS 's most powerful tools . The system uses data from the SIMBAD , the NASA / IPAC Extragalactic Database , the International Astronomical Union Circulars and the Lunar and Planetary Institute to identify papers referring to a given object , and can also search by object position , listing papers which concern objects within a 10 arcminute radius of a given Right Ascension and Declination . These databases combine the many catalogue designations an object might have , so that a search for the Pleiades will also find papers which list the famous open cluster in Taurus under any of its other catalog designations or popular names , such as M45 , the Seven Sisters or Melotte 22 .

== Title and abstract searches ==

The search engine first filters search terms in several ways . An M followed by a space or hyphen has the space or hyphen removed , so that searching for Messier catalogue objects is simplified and a user input of M45 , M 45 or M @-@ 45 all result in the same query being executed ; similarly , NGC designations and common search terms such as Shoemaker Levy and T Tauri are stripped of spaces . Unimportant words such as AT , OR and TO are stripped out , although in some cases case sensitivity is maintained , so that while and is ignored , And is converted to " Andromedae " , and Her is converted to " Hercules " , but her is ignored .

== Synonym replacement ==

Once search terms have been pre @-@ processed , the database is queried with the revised search term , as well as synonyms for it . As well as simple synonym replacement such as searching for both plural and singular forms , ADS also searches for a large number of specifically astronomical synonyms . For example , spectrograph and spectroscope have basically the same meaning , and in an astronomical context metallicity and abundance are also synonymous . ADS 's synonym list was created manually , by grouping the list of words in the database according to similar meanings .

As well as English language synonyms , ADS also searches for English translations of foreign search terms and vice versa , so that a search for the French word soleil retrieves references to Sun , and papers in languages other than English can be returned by English search terms .

Synonym replacement can be disabled if required , so that a rare term which is a synonym of a much more common term (such as ' dateline ' rather than ' date ') can be searched for specifically .

== Selection logic ==

The search engine allows selection logic both within fields and between fields . Search terms in each field can be combined with OR , AND , simple logic or Boolean logic , and the user can specify

which fields must be matched in the search results . This allows complex searches to be built ; for example , the user could search for papers concerning NGC 6543 OR NGC 7009 , with the paper titles containing (radius OR velocity) AND NOT (abundance OR temperature) .

= = = Result filtering = = =

Search results can be filtered according to a number of criteria , including specifying a range of years such as ' 1945 to 1975 ' , ' 2000 to the present day ' or ' before 1900 ' , and what type of journal the article appears in ? non @-@ peer reviewed articles such as conference proceedings can be excluded or specifically searched for , or specific journals can be included in or excluded from the search .

= = Search results = =

Although it was conceived as a means of accessing abstracts and papers , ADS provides a substantial amount of ancillary information along with search results . For each abstract returned , links are provided to other papers in the database which are referenced , and which cite the paper , and a link is provided to a preprint , where one exists . The system also generates a link to ' also @-@ read ' articles ? that is , those which have been most commonly accessed by those reading the article . In this way , an ADS user can determine which papers are of most interest to astronomers who are interested in the subject of a given paper .

Also returned are links to the SIMBAD and / or NASA Extragalactic Database object name databases , via which a user can quickly find out basic observational data about the objects analyzed in a paper , and find further papers on those objects .

= = Impact on astronomy = =

ADS is almost universally used as a research tool among astronomers , and there are several studies that have estimated quantitatively how much more efficient ADS has made astronomy ; one estimated that ADS increased the efficiency of astronomical research by 333 full @-@ time equivalent research years per year , and another found that in 2002 its effect was equivalent to 736 full @-@ time researchers , or all the astronomical research done in France . ADS has allowed literature searches that would previously have taken days or weeks to carry out to be completed in seconds , and it is estimated that ADS has increased the readership and use of the astronomical literature by a factor of about three since its inception .

In monetary terms , this increase in efficiency represents a considerable amount . There are about 12 @,@ 000 active astronomical researchers worldwide , so ADS is the equivalent of about 5 % of the working population of astronomers . The global astronomical research budget is estimated at between 4 @,@ 000 and 5 @,@ 000 million USD , so the value of ADS to astronomy would be about 200 ? 250 million USD annually . Its operating budget is a small fraction of this amount .

The great importance of ADS to astronomers has been recognized by the United Nations , the General Assembly of which has commended ADS on its work and success , particularly noting its importance to astronomers in the developing world , in reports of the United Nations Committee on the Peaceful Uses of Outer Space . A 2002 report by a visiting committee to the Center for Astrophysics , meanwhile , said that the service had " revolutionized the use of the astronomical literature " , and was " probably the most valuable single contribution to astronomy research that the CfA has made in its lifetime " .

= = Sociological studies using ADS = =

Because it is used almost universally by astronomers , ADS can reveal much about how astronomical research is distributed around the world . Most users access the system from institutes of higher education , whose IP address can easily be used to determine the user 's geographical

location . Studies reveal that the highest per @-@ capita users of ADS are France and Netherlands @-@ based astronomers , and while more developed countries (measured by GDP per capita) use the system more than less developed countries ; the relationship between GDP per capita and ADS use is not linear . The range of ADS usage per capita far exceeds the range of GDPs per capita , and basic research carried out in a country , as measured by ADS usage , has been found to be proportional to the square of the country 's GDP divided by its population .

ADS usage statistics also suggest that astronomers in more developed countries tend to be more productive than those in less developed countries . The amount of basic research carried out is proportional to the number of astronomers in a country multiplied by the GDP per capita . Statistics also imply that astronomers in European cultures carry out about three times as much research as those in Asian cultures , perhaps suggesting cultural differences in the importance attached to astronomical research .

ADS has also been used to show that the fraction of single @-@ author astronomy papers has decreased substantially since 1975 and that astronomical papers with more than 50 authors have become more common since 1990 .