Integrating Astronomical Data and Information Services at the CDS

Daniel Egret, Françoise Genova, Pascal Dubois,
André Heck, Soizick Lesteven, François Ochsenbein,
Michel Crézé, François Bonnarel,
Gérard Jasniewicz & Marc Wenger

CDS, Observatoire astronomique de Strasbourg, Strasbourg, France
question@simbad.u-strasbg.fr

Abstract

The CDS is providing several unique services to the world-wide astronomical community: the catalogue service, the SIMBAD database, the ALADIN project, bibliography and literature search, yellow pages, etc. We describe how the CDS works at providing a global perspective on astronomical data and information, with the help of recent technological developments.

1 Introduction

The objective of the Strasbourg astronomical Data Centre (CDS) is to provide on-line data and information for the world-wide astronomy community in a multi-wavelength and multi-mission approach.

In the present context, with an unprecedented accumulation of data being collected by ground- or space-based telescopes and observatories, the role of the data centres is to bridge the gap between the specialized approach of the scientific teams (where the detailed expertise about the specific data resides), and the general approach of the wider community of researchers (who need an easy access to data, calibrated as far as possible in meaningful physical units). This is more specifically critical for multi-wavelength or panchromatic astronomy (see Wells, 1992), a key to the understanding of many astrophysical processes, which implies to cross-compare data acquired for the same astronomical objects through several instruments working at different wavelengths.

The concept of data centre has to be somehow extended to the one of 'information hub' where the users will find many kinds of tools to help them retrieve the data and information they need among the many possible distributed sources.

We present here the current efforts of CDS in this direction.

2 The Astronomical Catalogues

The key rôle of CDS, as one of the first major astronomical data centres, is to collect the catalogues, control the data integrity, and act as a depository of catalogues and archives.

In the recent past, data centres could conceivably plan to store in magnetic tape collections whatever piece of information was produced electronically in astronomy. Obviously the evolution of observing techniques, including modern telescopes with their digital receivers, radio telescopes, or space missions, now imposes a different approach: both storage capabilities and expertise on data cannot be concentrated at data centres anymore.

To face these evolutions, a number of initiatives, taken by the CDS, contribute to a more efficient approach of the collection and distribution of astronomical catalogues.

2.1 Exchange Agreements

The CDS has signed international exchange agreements with NASA Astronomical Data Center (ADC), Japan's National Astronomical Observatory in Tokyo, the Russian Academy of Sciences, the SERC Starlink network in Great-Britain, China's Beijing Observatory, the University of Porto Allegre in Brazil, the University of La Plata in Argentina, and India's InterUniversity Center for Astronomy and Astrophysics.

2.2 Standardization of the Format and Documentation

A new standard for the format and documentation of ASCII tables has been proposed by Ochsenbein (1994). The corresponding catalogue descriptions are designed to be read easily by humans and by computers.

A full library of documented catalogues, available on-line for electronic distribution, is gradually being built following this format. Procedures are also provided to translate this format into FITS format, the common standard for images and binary data (Wells et al., 1981; Grosbøl, 1991).

The new standard is now also used by other data centres (NASA/ADC) and publishers (AAS CD-ROMs).

2.3 Tables from Astronomy & Astrophysics and Other Major Journals

Some of the major astronomical journals are heading towards electronic publication, at least of their data tables (see e.g., the contributions given in a specific session of this meeting).

Since January 1993, following an agreement with the Editors of *Astronomy & Astrophysics*, a number of tables from the main journal and from the Supplements Series are deposited at the CDS, and made available for electronic distribution (see Lequeux and Ochsenbein, in this meeting).

The tables published in the AAS CD-ROMs are also made available on-line at CDS, after an agreement with the Editors.

2.4 A Catalogue Alerting Service

New catalogues and tables made available on-line, and more generally catalogues for which an updated version has been produced in the current month are listed in a specific file on the CDS server. Users and data managers can subscribe to an automatic two-weekly alerting service through electronic mail.

2.5 The Catalogue Service on the World-Wide Web

The CDS started by 1991 an effort to move most frequently requested catalogues, together with their documentation, into a fully electronic archive accessible on-line. New catalogues and tables are now directly entered into this system. This effort was made possible through the collaboration between all data centres, and especially NASA/ADC which had in the meantime installed its own magnetic tape collection into a near-line system, and had produced a subset of most frequently used catalogues on CD-ROM.

The "Astronomer's Bazaar" made available at the end of 1993 (Egret and Ochsenbein, 1994) is a fully interactive on-line service, based on the World-Wide Web (WWW) allowing remote users:

– to query the list of catalogues, by keyword, or in browse mode,
– to display the corresponding documentation,

– and to retrieve the complete electronic files (possibly compressed), from the `anonymous ftp` space of the CDS server.

More than 1200 catalogues and tables, for a total of several Gigabytes of data are already available through this interface.

2.6 Dictionary of the Nomenclature of Celestial Objects

The Second Reference Dictionary of the Nomenclature of Celestial Objects has been recently published by Lortet et al. (1994). It can be queried through different keys (catalogue name, author, object type, format, etc.) on-line on the SIMBAD host, and through the World-Wide Web server. The Dictionary is also used as an auxiliary database in Simbad.

2.7 The ESIS-CDS Catalogue Browser

As a result of a collaboration with the ESA/ESIS project, ESRIN and the CDS are setting up a Catalogue Browser offering additional functionalities, namely the search by parameters through individual catalogues, or through a full set of catalogues. These functionalities will be integrated within the existing CDS services. The catalogue browser will also support links to other archives, such as the ESO-ECF and CADC archives.

The operational phase of this improved catalogue browser is planned for early 1996.

3 The SIMBAD Database of Astronomical Objects

The specificity of the SIMBAD database is to organize the information per astronomical object, thus offering a unique perspective on astronomical data. This can only be done through a careful cross-identification of catalogues, lists, and journal articles, a task which has made SIMBAD a key tool used worldwide for all kinds of astronomical studies.

The SIMBAD database has been described by Egret et al. (1991) and recent developments are regularly presented in the "Simbad News" papers in the CDS Information Bulletin.

We will just recall here, for completeness, the main features of this astronomical object-oriented database:

- a database of more than 1 million astronomical objects (stars, galaxies and all astronomical objects outside the solar system);
- a cross-index to more than 1400 astronomical catalogues and observation logs of space missions;
- observational data from some 25 different types of data catalogues and compilations;
- a bibliographic survey covering the astronomical literature since 1950 for stars, and since 1983 for extragalactic objects;
- an interactive object-oriented database system offering several user interfaces (the classic command-line interface, Xsimbad for XWindow, e-mail batch mode, etc.), all of them using a common client/server mode for remote access;
- an efficient management of the possible variations in the naming of astronomical objects with the *sesame* module and the *info* database of nomenclature of celestial objects;
- a *name resolver* integrated within other applications (such as STARCAT, STARVIEW, ISSA-PS, HEASARC, IRSKY, etc.), benefiting from the client-server approach previously mentioned.

Unlike the other CDS services, SIMBAD is a charged service. Users have to register, and receive a userid/password from the CDS staff (or from the U.S. agent for American users). The charges are covered by NASA for all U.S. users, and, starting January 1995, are covered by ESO and ESA for all European users from ESO or ESA member states.

SIMBAD is kept up-to-date on a daily basis. New data (bibliographical references, identifiers, basic data), and new acronyms are being entered day after day; this is done as a result of a continuous survey of the astronomical literature, under the responsibility of Institut d'Astrophysique de Paris with the collaboration of Paris and Bordeaux observatories (see Laloë et al., 1993, and the communication by Laloë at this meeting).

In 1994, about 500 new bibliographical references per month were added to SIMBAD, while the rate of growth of the number of objects in SIMBAD was about 2500 new objects monthly.

Large astronomical catalogues, carrying their specific identifiers and measurements, are added after a cross-matching procedure which frequently spans many months.

The data contained in SIMBAD are also permanently updated, as a result of errata, remarks from the librarians (during the scanning of the literature), quality controls, or special efforts from the CDS team to better cover some specific domains. Requests for corrections, errata, or suggestions are regularly received from SIMBAD users through a dedicated e-mail address (question@simbad.u-strasbg.fr).

4 The ALADIN Interactive Sky Atlas

ALADIN is a new project, currently under development (Paillou et al., 1994), to create an interactive atlas of the digitized sky allowing the user to visualize on his/her own workstation digitized images of any part of the sky, to superimpose entries from astronomical catalogues or user data files, and to interactively access the related data and information from the SIMBAD database for all known objects in the field.

This new tool will be particularly useful for multi-spectral approaches such as searching for counterparts of sources detected at various wavelengths, and for a number of applications related to the database quality control and the careful identification of astronomical sources.

The ALADIN project has set up collaborations with the major groups providing digitizations of sky surveys (see e.g., McGillivray, 1994).

The software architecture of ALADIN is based on the client/server philosophy, each set of stored data (astronomical catalogues, SIMBAD database, and image pixels) being accessed through a dedicated server.

Currently (May 1995), ALADIN is only available as a local prototype; it will eventually become a public interactive tool, available for all laboratories through networks. It will allow any astronomer to point a region of the sky, to display the corresponding digitized images (with accurate positions and fluxes), and to overlay catalogued data from the SIMBAD data base and CDS catalogues.

5 Bibliography and Literature Search

The bibliography of astronomical objects is one of the unique features of the SIMBAD database. This service is gradually being extended in order to provide the user with a wider perspective of the current astronomical literature.

5.1 The SIMBAD Bibliography

The SIMBAD bibliography for the astronomical objects includes references to all published papers from some 90 periodicals covering the whole astronomical literature. Articles are scanned in their entirety, and references to all objects mentioned are included in the bibliography. References, authors, and titles are stored for more than 80,000 papers since 1950 providing some 1,4 million references to astronomical objects.

5.2 Abstracts of Recent Papers

More recently, through agreements with the Editors, the abstracts of *Astronomy & Astrophysics*, main journal and Supplement Series, as well as the abstracts from the *Publications of the Astronomical Society of the Pacific*, are made available on-line a few weeks before publication.

The collection of abstracts from these two journals, starting from January 1994, is available on the World-Wide Web, and can be queried by keywords or author names. Links are made between the abstracts and the electronic tables, when available (see contribution by Ochsenbein and Lequeux at this meeting). The bibliographic 19-digit reference code (Schmitz et al, 1995) also shared with NED and ADS, is the common index key.

This bibliography service is expected to be developed in the future, with the planned inclusion of other abstract collections, and the implementation of advanced search mechanisms.

The analysis of bibliography linked to astronomical objects is an efficient tool for scientific investigations and database quality control (as shown by Lesteven, in this meeting).

6 Yellow-Page Services

A growing number of astronomical resources are made available through the Internet. At the same time, modern tools such as the World-Wide Web makes possible a real navigation between these resources, using carefully collected pointers. Combining yellow-page services and meta-databases of active pointers is a crucial solution to the data retrieval problem.

The CDS hosts the following databases:

- **The Star*s Family of Internet resources** is the generic name for a growing collection of directories, dictionaries and databases which is described in more detail by A. Heck in a contribution to this meeting (see also Heck, 1995).

- **AstroWeb** (Jackson et al. 1994) is a collection of pointers to astronomically relevant information resources available on the Internet. It is maintained by the AstroWeb consortium, a group of scientists from CDS (Strasbourg astronomical Data Center), MSSSO (Mount Stromlo and Siding Spring Observatories), NRAO (National Radio Astronomy Observatory), STScI (Space Telescope Science Institute), ESO/ST-ECF (European Southern Observatory, Space Telescope–European Coordinating Facility), VILSPA (IUE Observatory), and CERN/La Plata.

  The browse mode of AstroWeb gives a very good perspective of the efforts currently developed for making astronomically related information available on-line through the World-Wide Web.

7 Towards a Global Astronomical Data and Information Service

The various activities of the CDS, as described above, are the pieces of a puzzle, all directed towards the organization of a simple and unified view of astronomical data and information.

In order to achieve this goal, the CDS is gradually building the needed links between the different CDS services in order to provide more powerful tools to the users. These new developments imply a dedicated effort in terms of research, in order to derive the best solutions according to the existing information technology.

Examples of such links are:

- the ability to overlay SIMBAD and catalogue data on ALADIN sky fields;
- the access to abstracts and to catalogue information from the SIMBAD database;
- the access to data tables for recent papers listed in the abstract service.

A first version of the CDS information service is already available: this is the WWW server. It provides the following pages:

- General information on CDS
- Access to the Catalogue service
- SIMBAD documentation
- Abstract service for *A&A* and *PASP*
- On-line version of the CDS Information Bulletin
- Star*s Family of Internet resources and AstroWeb
- etc.

The CDS, hosted by the Strasbourg astronomical Observatory, is funded by the French 'Institut National des Sciences de l'Univers' (INSU, CNRS).

References

[1] Egret, D., Wenger, M., Dubois, P. (1991), in *Databases & On–line Data in Astronomy*, Albrecht & Egret (Eds.), Kluwer Acad. Publ., 79.

[2] Egret, D., Ochsenbein, F. (1994), CDS Inform. Bull. **44**, 57.

[3] Grosbøl, P. (1991), in *Databases and On-line Data in Astronomy*, M.A. Albrecht & D. Egret (Eds.), Kluwer Acad. Publ., 253.

[4] Heck, A. (1995), "The Star*s Family – An example of comprehensive yellow-page services", in *Information & On-line Data in Astronomy*, D. Egret and M. A. Albrecht, Eds., Kluwer Acad. Publ., in press.

[5] Jackson, R., Wells, D., Adorf, H.M., Egret, D., Heck, A., Koekemoer, A., Murtagh, F. (1994), *Astron. Astrophys. Suppl.* **108**, 235.

[6] Laloë, S., Beyneix, A., Borde, S., Chagnard-Carpuat, C., Dubois, P., Dulou, M.R., Ochsenbein, F., Ralite, N., Wagner, M.J. (1993), *CDS Inform. Bull.* **43**, 57.

[7] Lortet, M.C., Borde, S., and Ochsenbein, F. (1994), *Astron. Astrophys. Suppl.* **107**, 193.

[8] MacGillivray, H.T., Editor (1994), "Astronomy from Wide-Field Imaging", Postdam, Germany, Kluwer Acad. Publ.

[9] Ochsenbein, F. (1994), *CDS Inform. Bull.* **44**, 19.

[10] Paillou, Ph., Bonnarel, F., Ochsenbein, F., Crézé, M. (1993), "Aladin Deep Sky Mapping Facility", Project Report, CDS, Strasbourg, May 1993.

[11] Paillou, Ph., Bonnarel, F., Ochsenbein, F., Crézé, M. (1994), in *Astronomy from Wide-Field Imaging*, Postdam, Germany, H.T. MacGillivray Ed., Kluwer Acad. Publ., 347.

[12] Schmitz, M., Helou, G., Dubois, P., LaGue C., Madore B., Corwin Jr. H.G., Lesteven, S. (1995), in *Information & On-line Data in Astronomy*, D. Egret & M.A. Albrecht (Eds.), Kluwer Acad. Publ., in press.

[13] Wells, D. (1992), in *Astronomy from Large Databases, II*, ESO Proceedings **43**, 143.

[14] Wells, D. C., Greisen, E. W., Harten, R. H. (1981), *Astron. Astrophys. Suppl.* **44**, 363.