Baltic Astronomy, vol. 6, 176–182, 1997.

THE EVOLUTION OF DISSEMINATION OF ASTRONOMICAL DATA

Françoise Genova

CDS, URA CNRS 1280, Observatoire de Strasbourg, 11 rue de l'Université, 67000 Strasbourg, France

Received July 23, 1996.

Abstract. The ways and means for the dissemination of astronomical data have changed rapidly in recent years, owing in particular to accelerating technical evolution. The role of the "general" and the "specialized" data centers, the impact of the WWW development and the consequences of electronic publishing of astronomical journals on traditional activities of data centers are discussed.

Key words: databases – electronic publishing

Astronomy has a very long tradition of data conservation and dissemination: Ptolemy's Almagest, published in the 2nd century, contains a star catalog. Chronicles from ancient China, Japan and other countries are still in use to track back observations of transient astronomical phenomena. Astronomers have a long-term habit of keeping at least part of their data in the form of catalogs and using data and results produced by their colleagues. Data collections are in particular used to build up statistics on many objects, to identify families of objects and to study their properties. It is also used to study variability and to compile the "best value" from measurements of the same parameter taken from different sources. It is used to go back to observations to test new theories.

The importance of collecting and disseminating astronomical data is growing every day, due in particular to the general trend towards multiwavelength astronomy studies. The understanding of the physical processes at work in astronomical objects involves a comparison of more and more data acquired at different wavelengths.

No one can be a specialist of many different techniques used, and an important fraction of the data needed often comes from earlier observations by other astronomers. Thus, it is necessary to allow non-specialist astronomers to re-use data obtained by others.

Apart from these scientific reasons, the trend towards large ground-based projects, such as the VLT, Keck, Subaru, Gemini, etc., and large space observatories of different space agencies makes it more and more necessary to optimize the scientific return of these costly projects. This was realized many years ago by NASA, which founded the National Space Science Data Center (NSSDC) in 1966, to provide access to data from the NASA space flight missions (Roman 1997). The growing awareness of the funding agencies can be also illustrated by the fact that "maximal exploitation of data" was decided to be one of the priorities of French astronomy in a recent national prospective report.

Several kinds of specialists intervene in the collection, preservation and diffusion of data:

- *the data producing teams*, which know their instruments and methods, have to preserve at least part of their data in a form usable by other astronomers. The situation is improving in this domain, since systems for data production, diffusion and archiving become more and more available at large observatories (see e.g. the HST on-line services at STScI, ESO-ST-ECF and the CADC).
- in some cases, the *specialized centers* build access to and information about different subfields of astronomy. NASA, in particular, has developed several specialized centers, e.g. HEASARC (High-Energy Astrophysics Science Archive Research Center) or IPAC (Infrared Processing and Analysis Center). There are also specialized databases, such as NED (NASA Extragalactic Database) at IPAC or LEDA (Lyon-Meudon Extragalactic Database) at the Lyon Observatory in the extragalactic domain, or the GCPD (General Catalogue of Photometric Data) built at the Lausanne Observatory by J.-C. Mermilliod and his colleagues. As it is demonstrated by these examples, one deals now with information rather than with data. This information includes data, results, documentation and the connections between them.
- the role of the *data centers* is to bridge the gap between the specialized approach of the scientific teams and the general approach of the astronomical community. In practical terms, this

means the definition, development and distribution of tools helping astronomers to retrieve the information they need from the growing number of possible sources.

Tools permitting the retrieval of astronomical information have, of course, been developed in the past: the compiled catalogs containing observations, measurements or bibliography, the systematic collection of abstracts published since 1899 by the Astronomisches Rechen-Institut in Heidelberg, the collection of catalogs prepared by astronomical data centers in the last 25 years, the "Yellow Page services", such as the one produced by Stroobant et al. (1907) at the Observatoire Royal de Belgique (Heck 1991) or the Star*s Family of Astronomy and Related Resources developed by A. Heck, a growing collection of directories, dictionaries, databases and related products on astronomy, space sciences as well as on related fields, organizations and people (Heck 1995).

The question of information retrieval in astronomy is very closely linked to the techniques used for preserving and disseminating information and to the type of information to be retrieved. In recent years, the problem has evolved towards increasing complexity. The volume and diversity of information increases very rapidly, and this will continue in the coming years. The large ground-based and space-born observatories, the large surveys such as DENIS, 2MASS, SLOAN or the large Schmidt telescope catalogs (PPM, GSC I and II, APM, APS, etc.), produce larger and larger amounts of data. In parallel, the volume of results published in journals is also increasing rapidly. The evolution of the computer storage capacity, the development of networks and webs, lead to many "individual" sources of information (e.g. personal Web pages). Important "organized" sources will be the data archives from the large space and ground-based observatories, as well as the information gathered by specialized centers, data centers, journals and books.

A very important task is to render the scattered information retrievable. For this, WWW navigation is an important tool. In contrast to Net Browsers, information for specialized scientific usage is better retrieved through specialized tools, developed with an "expert eye" allowing an adequate selection of information in the particular field. This expertise on astronomical data will be important at all steps: the expertize of the data producing teams on their own data and techniques, the expertize of the specialized centers and specialized database providers in their field and general expertize of data centers on astronomical data, results and information retrieval methods. The data centers in particular develop "metadatabases" (e.g. SIMBAD). Information retrieval tools rely mostly on selected, organized, homogenized data and results (not on raw data): lists of observations (observatory "logs"), catalogs of observations or results, published papers, reference images, etc. Another very important point is the need for standards and tools, as will be shown in the following (see also Ochsenbein 1997; Schmitz et al. 1995).

Technical evolution indeed plays a very important role in the evolution of astronomical information services, allowing significant improvements of the service to the users. This is illustrated by the following examples which were taken from activities of the CDS. It is clear that all the data centers and data providers share this kind of experience. Many other examples can be found in the papers from these Proceedings.

- An international network of data centers (which are described in other papers presented at this Colloquium) takes charge of the collection, preservation and distribution of astronomical catalogs. This activity has evolved very significantly in the last few years, owing to the development of networks, the Internet and the WWW, and owing to the availability of powerful and affordable commercial Data Base Management Systems. A few years ago, the catalogs were distributed (e.g. on magnetic tapes or microfiches) by regular mail. Since March 1992, anonymous FTP has been used, with a WWW access at the end of 1993. This led to an explosion of the diffusion of catalogs. At CDS it has increased from about 100 orders per year by surface mail to more than 10 000 files per month (mid-1996), from 4000 different nodes. The access tools have also evolved, from the query by keywords or browsing of the catalog list, allowing one to retrieve entire catalogs, to a DBMS system which permits the search of any catalog or table by any of its fields, the CDS/ESRIN VizieR Catalogue Browser. The key for cooperation between the data centers and with the journal editors is the standard description of tables, first defined by CDS and now shared by all partners.
- The evolution permitted by the development of new tools to build graphics user interfaces is illustrated by different versions of the SIMBAD user interface. The "classical" command line interface requires the user to know the specific command line query language and it allows one to perform simple and complex queries. XSimbad is a menu-driven, XWindow interface which was distributed by the end of 1993. It keeps most of the

functions of the command line interface. A WWW interface, WSimbad, will be open within a few months with the most frequently used functions as a first step. This new development, and the definition of a general CDS data exchange model which is under way, will offer to the user the possibility of navigating between the CDS services. In parallel, a set of client/server routines, which gives access to the fundamental information from SIMBAD (object names, position, bibliography), has been installed in several services (ADS, HEASARC, StarView, STARCAT, ISSA/PS, IRSKY, WISARD, AMASE, SAX, etc.). This access mode to SIMBAD, in which the information from SIMBAD is used for the purposes of other services, is used most frequently.

The rapid development of hypertext links between distributed information, and the client/server access from one service to another, demonstrates the growing interconnection between astronomy services. This allows the user to navigate between different types of information, e.g. from bibliography to objects (e.g. ADS to CDS) and reciprocally (CDS to ADS). WWW access to services, data and publications develops very rapidly. This opens new possibilities: the access tools to observatory archives such as VizieR (see Payne 1997 for description of access tools to archives, software and preprints), as well as the links between the electronic journals and databases. The increased partnership between data centers and editors of journals leads to important consequences for the journal editing process (with new "automated" consistency checks for the contents), for the data center procedures and contents and the emergence of ADS as a reference bibliography data center for astronomy.

This evolution is certainly a challenge for astronomical data centers. They have to adapt their activities in a very rapidly changing world and keep up with the evolution of astronomy itself. This requires them: (1) to maintain their general expertise and to take into account the new subjects which arise, (2) keep up with the evolution of the "political" context, which means to be ready for the large projects and to adapt their cooperations to the policy of the funding agencies, (3) keep up with the increase in the volume and complexity of the available astronomical information, which means adapting their procedures (in particular increasing the automation, but preserving the quality) to develop new tools and new standards, (4) keep up with the very fast and significant evolution of techniques and methods, which requires a sustained activity of technological and methodological watch. This means managing two different types of activities, with very different time scales and constraints: (1) the continuous enrichment of the database contents which is an everyday, time consuming activity over many years, essential for the quality of the services although not spectacular and (2) the development of new services and new interfaces in phase with the technical evolution and the users' expectations with time scales of a few months.

In this rapidly evolving context, the data centers will continue to ensure their "mission". They will develop general tools for information retrieval and for "metadatabases", as well as their general expertize in astronomical information and the use of this expertize in the services, participation and initiative in the definition of standards, partnership with the journal editors, the data providers and the specialized centers. The present situation offers to data centers very interesting opportunities for giving access to new types of information and the capacity of building up innovative services, in particular, by building links between electronic journals and databases and links between distributed services and archives. Clearly, the increase in interconnection made possible by the technical evolution will result in the increase in world-wide cooperation in the domain of data dissemination.

Table 1. A few WWW URLs to begin with.

ADS	http://adswww.harvard.edu/abs_doc/abstract_service.html
ADC	http://adc.gsfc.nasa.gov/
ADAC/NAOJ	http://adac.mtk.nao.ac.jp/index.html
ApJ	http://www.aas.org/ApJ/
CADC	http://cadcwww.dao.nrc.ca/
CAD/INASAN	http://www.inasan.rssi.ru/CAD.html
CDS	http://cdsweb.u-strasbg.fr/CDS.html
GCPD	http://obswww.unige.ch/gcpd/gcpd.html
HEASARC	http://heasarc.gsfc.nasa.gov/
IPAC	http://www.ipac.caltech.edu/
LEDA	http://www-obs.univ-lyon1.fr/base/
NSSDC	http://nssdc.gsfc.nasa.gov/
STScI	http://www.stsci.edu/top.html
ST-ECF	http://arch-http.hq.eso.org/ESO-ECF-Archive.html

REFERENCES

Heck A. 1991, in Information and On-line Data in Astronomy, eds. M. A. Albrecht & D. Egret, Kluwer Academic Publishers, Dordrecht, p. 211

Heck A. 1995, in Information and On-line Data in Astronomy, eds. D. Egret & M. A. Albrecht, Kluwer Academic Publishers, Dordrecht, p. 195

Ochsenbein F. 1997, Baltic Astronomy, 6, 221

Payne H.E. 1997, Baltic Astronomy, 6, 236

Roman N.G. 1997, Baltic Astronomy, 6, 283

Schmitz M., Helou G., Dubois P., LaGue C., Madore B., Corwin Jr. H. G., Lesteven S. 1995, in Information and On-line Data in Astronomy, eds. D. Egret & M. A. Albrecht, Kluwer Academic Publishers, Dordrecht, p. 259

Stroobant P. et al. 1907, Les observatoires astronomiques et les astronomes, Hayez, Bruxelles