The A&A Tables and Abstracts: an Example of Collaboration between Data Centres and Editors

François Ochsenbein

CDS, Observatoire de Strasbourg, Strasbourg, France

James Lequeux

Observatoire de Paris-Meudon, Paris, France

Abstract

As the result of an agreement between the Centre de Données de Strasbourg (CDS) and A&A, the Abstracts of all papers published by A&A (Main Journal and Supplement) and some large tabular data published in A&A are distributed in electronic form. This service is presented and its impact on the astronomical community is discussed, based on a two-year experience for the tabular data and one year for the Abstract service.

1 The Background

The CDS (Centre de Données astronomiques de Strasbourg) has now a 25-year experience in the preparation, management, distribution and scientific usage of electronic data. If this 25-year period represents less than a human generation, it does represent many computer generations...

Since its beginning, in 1972, CDS kept, in collaboration with other institutes, electronic data initially on punched cards and magnetic tapes: in this epoch, the data were patiently "keypunched" from printed volumes to generate "machine-readable" catalogues. The usage of the computer in astronomy is effectively an old story – the first electronic version of an astronomical catalogue was created by Eckert in 1937 [?], almost 60 years ago...

The computer technology evolved a lot during this period: for instance, the compactness of the data media moved from $\sim$ 20 bytes/cm^3 for the punched cards and 50Kbytes/cm^3 for the magnetic tapes to 50 Mbytes/cm^3 for optical disks – roughly speaking, the compactness of the data media doubles each year.

Year	Catalogues
1975	116
1981	280
1983	333
1991	596
May 1993	641
June 1994	923
May 1995	1150

Table 1
Evolution of the number of catalogues saved at CDS.

Year	A&A		A&AS	
	Pages	Papers	Pages	Papers
1975	3800	540	1600	80
1985	5400	770	2500	120
1990	8000	990	3500	130
1993	10600	1100	3900	200
1994	12600	1300	3600	240

Table 2
Increase in publication.

The other major component of the evolution of the computing environment is the development of *networks*: the networks began to link the astronomical institutes in the mid-80s, and Internet came up around 1990. Now several 10^7 computers are linked to Internet, and it becomes faster to retrieve a piece of information from the other side of the planet via navigating over the network rather than moving to one's home institute library.

For an historical perspective, Table 1 summarizes the number of catalogues saved at CDS in electronic form: even though a *catalogue* may represent data amounts from a few Kbytes to hundreds of Mbytes, it is representative of the effort in collecting the astronomical data.

On their side, the editors of the astronomical publications are faced with an ever increasing volume of the published material; this evolution is illustrated for A&A with a few figures over the last 20 years in Table 2.

Year	A&A		A&AS	
	Articles	Size (Kb)	Articles	Size (Kb)
1993-1	3	69	12	1687
1993-2	22	372	41	9880
1994-1	28	1128	47	30168
1994-2	17	1388	51	12572

Table 3
Electronic tables from A&A.

2 A&A Electronic Tables

From Table 2 one can notice that the average size of the papers published in the Supplements of A&A is about twice the average size of a publication in the Main Journal, simply because the Supplements are reserved for the publication of lengthy observational results.

In fact, large tables published in A&AS were for a rather long time the final output of a computer processing, and it also happened a long time ago that the publication just consisted of an "announcement" of data in electronic form available at CDS – for instance in 1976, A&AS 26, 419.

These two conditions – usage of the computer for data processing, and an increase of the volume and of the cost of publications – naturally lead to a publication of the large data sets in electronic form.

One "ingredient" was however missing until $\sim$ 1990: the access to the electronic material was slow – about one week was required to mail a magnetic tape – and required heavy and expensive material such as tape units. The networking – should we prefer a term like *world-wide-webbing*? – completely changed this aspect.

The first announcement concerning the electronic distribution of data published in A&A was made in 1992 (A&A 266, page E1). It specifies that the material available in electronic form can be tables totally or partially printed, or material just used for constructing figures. The choice of the material to be distributed in electronic form is either proposed by the authors or sollicited by the editors.

This agreement concerns tables of papers *accepted* for publication in A&A since January 1993: this means practically that tables published after September 1993 in A&A have a good chance to exist in electronic form – as can be inferred from Table 3.

The amount of material available at CDS as a result of this agreement is summarized in Table 3; as foreseen from the purposes of the Main Journal and the Supplements, the frequency of the papers having electronic data is very high for the Supplements: over 40% of the papers published in 1994 in A&AS have associated electronic data – to be compared to the 3% for the papers of the Main Journal. The average size of the electronic data associated with a publication is also quite different: 300KB for the Supplements compared to 40KB for the Main Journal. It can also be noticed that the total size of the tabular data is currently about 50MBytes – not a large amount in today's standards.

3 *A&A* Electronic Abstracts

The publication of the Abstracts in electronic form is not motivated by the increase of the volume of publication, but was installed as a facility to advertise and locate the papers.

Following the proposal of the A&A editor, abstracts of the papers published in A&A are collected at CDS since April 1994 for the Main Journal, and January 1994 for the Supplements. The work for the generation of the electronic Abstracts is carefully scheduled and starts about six weeks before the release date of the publication, when the printers mail the existing abstracts in TeX or LaTeX format. These are interpreted at CDS, and the inevitable problems with undefined macros or complicated mathematical formulae are then approximately solved. This first run does not cover *all* abstracts, since some papers – typically the Letters in the Main Journal which are often prepared camera-ready – are not ready in electronic form at this period.

About 3 weeks before the official distribution date, a copy of the printed abstracts is received at CDS; at this stage, the Abstracts received earlier are checked – late modifications occur from time to time – and the Abstracts or the papers not previously received are entered. A final check is performed when the volumes are effectively received at the CDS library.

The possible late modifications of the Abstracts mean that the astronomers should *not* rely on the Electronic Abstracts – for instance to generate a bibliography – before the actual publication. The same remark obviously applies to usage of A&A electronic abstracts in Abstract Services; in this latter case, an agreement with the A&A Editors is obviously required.

The Electronic Abstracts are saved as plain ascii files, and do not require a lot of disk space: 1.3Mbytes/year for the Main Journal, and 0.25Mbytes/year for the Supplements. This presentation as plain ascii files, if it allows users having the minimal equipment (an ascii terminal) to have a look at the contents and facilitates the abstract indexing, implies severe limitations on the set of symbols available: for instance, we adopted the same conventions as *NED* for mathematical symbols, which can lead to hardly understandable formulae... This limitation also concerns accentuation – especially important in the authors' names.

4 Documentation of the Tables

The presentation of the tabular data for the usage by a computer differs significantly from its presentation in a paper: while the organisation in columns, and the meaning of the various column headings are more or less obvious for the reader, a computer requires complete definitions of the table structure and of how the set of bytes has to be interpreted.

The documentation of the electronic tables is therefore a critical point. In the pioneer years, it was assumed that the potential user would find the time to compare the electronic version to the printed one, and to write his own Fortran program to make use of the electronic data: the documentation just mentioned the basic information about the authors and reference of the printed version, and in a wide variety of presentations how the columns were organized.

Early progress in the documentation of electronic astronomical catalogues occured with documentation made at ADC (Astronomical Data Center at NSSDC) as booklets of *Documentation for Machine-Readable Version* of catalogues.

A further step towards more systematic and uniform documentation of catalogues occured with the publication and distribution of the *Selected Astronomical Catalogues* CD-ROM by the ADC in 1992, where the documentation has been put in the form of FITS headers (Harten et al., 1988, [?]) to allow usage by data processing systems.

With A&A Tables, the input flow of about 15 new catalogues per month implies a systematic documentation. It was decided to put all the documentation concerning a publication in a dedicated `ReadMe` file. This document file is comparable in its possible usage to the FITS headers of the ADC CD-ROM – and in fact the data tables can be copied in FITS format, as explained in the next section. The main differences between the CDS and the ADC documentations are found in the presentation – the `ReadMe` file can easily be read and understood by a human eye; and accurate definitions of table layouts in terms of column locations, labels, units, parameter limits and possible unspecified (or *NULL*) values are systematically included in the `ReadMe` file. The complete description of this documentation standard may be found in [?]; this document is also available on the Web.

The same standard has also been recently adopted for the tables published on CD-ROM by the AAS (ApJ, AJ, PASP).

5 Access to the Tables

For a long time, the distribution of electronic data from CDS relied on the post-office when tapes (or later diskettes) were mailed to the interested astronomers.

The first free access on the network – so-called *anonymous ftp* access – was installed at CDS in March 1992. This way of getting electronic data became immediately the most used one. Currently (May 1995), the monthly traffic on the FTP node *cdsarc.u-strasbg.fr* includes 10,000 files copied in 4,000 sessions by 750 different nodes; the average FTP session lasts 7 minutes, copies 120KBytes with a speed of 1.9KB/s; and there are 1.6 users connected simultaneously. These average values cover of course very different copies ranging from the copy of just a `ReadMe` file for a Web user – a session of less than one second – to the copy of *e.g.* the *Tycho Input Catalogue* – a catalogue of 262MBytes.

Using the tabular data in the popular data processing systems used in astronomy is simpler when the catalogues are presented in FITS format, as explained above; this is why CDS recently improved an enhanced FTP server which can deliver files directly in the FITS ascii table data format. The CDS FTP server also allows the compression or decompression of the data files during the transfer; for more details see Ochsenbein [?].

Another measure which comes out of the usage of data copied from CDS is the popularity of electronic tables: the most frequently copied catalogues from the CDS archives by external users during the last year are listed in Table 4: one can notice the popularity of "classics" like the Bright Star or the Nearby Star catalogues, or of the NGC catalogue. The figures concerning only the tables resulting from the agreement between CDS and A&A are given in Table 5, where the popularity of the stellar models can be noticed.

A new facility to retrieve excepts of tables from celestial coordinates addressed to a large set of tables and catalogues stored at CDS is being developed with the so-called Catalogue Server; it will be announced in the near future on the CDS WWW page.

Copies	Catalogue designation and short title	
103	V/50	Bright Star Catalogue, 5th Ed.
102	V/70A	Nearby Stars, Preliminary 3rd Version
99	I/131	SAO Star Catalogue (Edition ADC 1989)
94	VII/118	NGC 2000.0 (Sinnott)
74	VII/1A	Revised NGC (Sulentic 1974)
74	I/196	Hipparcos Input Catalogue, Version 2
71	V/32A	Woolley's Catalogue of stars within 25pc (1970)
71	I/197A	Tycho Input Catalogue (Revised Version)
66	I/193	PPM South Star Catalogue (Bastian+ 1993)
65	VII/166	QSO and AGN (Véron+ 1993)
64	I/107A	Washington Double Star Catalog (Worley 1984)
61	VII/92A	Open Cluster Data 5th Ed. (Lyngå 1987)
60	I/208	The 90000 stars Supplement to PPM (Röser+ 1994)
58	II/139	GCVS (Variable Stars), 4th Ed.
55	V/84	Strasbourg-ESO Catalogue of Galactic Planetary Nebulae (Acker+ 1992)
52	VI/66	Precession Formulae and Mean Elements for Moon & Planets (Simon+ 1994)
52	J/A+AS/100/647	Evolutionary Sequences II. (Bressan+ 1993)
51	VII/110A	Abell's Catalogue of Rich Clusters of Galaxies
50	I/204	Orbital Elements of Minor Planets (St Petersburg 1993)

Table 4
Most frequently copied data at CDS for the period May1994/May1995.

6 Interconnectivity with Other Services

A WWW access to the A&A Tables and Abstracts has been installed since November 1993 from the URL http://cdsweb.u-strasbg.fr/CDS.html. It allows a direct connection between the Tables and Abstracts: for instance one or several abstracts can be retrieved from a set of words – authors' names or words existing in the abstract – and a hyperlink provides an immediate access to the corresponding electronic tables when such tables exist. The WWW allows many more connections to Databases, Abstract Services, observation archives, etc. See e.g. Egret [?] in this conference. The present usage (May 1995) is 1500 full abstracts retrieved monthly in 1000 connections.

Copies	Reference	Title
74	J/A+AS/100/647	Evolutionary Sequences II. (Bressan+ 1993)
52	J/A+AS/98/523	Grids of stellar models. II. (Schaerer+ 1993)
40	J/A+AS/101/415	Grids of stellar models III. (Charbonnel+ 1993)
39	J/A+AS/106/275	Theoretical isochrones (Bertelli+ 1994)
39	J/A+AS/104/179	Surface photometry in Shapley-Ames galaxies (Goudfrooij+ 1994)
38	J/A+A/275/101	Chemical evolution of the galactic disk I. (Edvardsson+ 1993)
37	J/A+AS/105/311	M giants spectra and photometry (Fluks+, 1994)
36	J/A+AS/102/339	Grids of stellar models IV (Schaerer+, 1993)
35	J/A+AS/104/365	Evolutionary sequences. III. (Fagotto+ 1994)
33	J/A+AS/105/29	Evolutionary sequences IV. (Fagotto+ 1994)
30	J/A+AS/102/397	Colour magnitude diagram of M13 (Guarnieri+ 1993)
29	J/A+A/281/817	Model atmospheres for Vega (Castelli+ 1994)
27	J/A+AS/103/97	Grids of stellar models V. (Meynet+ 1994)
25	J/A+AS/105/39	Evolutionary sequences with very high metal. (Fagotto+ 1994)
25	J/A+A/272/235	Star formation in L1251 (Prusti+ 1993)
24	J/A+A/274/105	UV to X bump of Seyfert 1 AGNs (Walter+ 1993)
23	J/A+AS/104/509	Stark broadening of Lyman and Balmer lines (Stehle 1994)
23	J/A+AS/104/271	CCD survey of galaxies II. (Gavazzi+ 1994)
23	J/A+AS/102/89	uvby-beta Photometry of G5 stars (Olsen 1993)
23	J/A+A/271/515	Atmospheric parameters of A and F stars I. (Smalley+ 1993)
22	J/A+AS/106/21	Delta Sct stars: a new revised list (Rodriguez+ 1994)
22	J/A+AS/100/583	IR and Optical Photometry of galaxies (Boisson+ 1993)
22	J/A+A/265/32	Radial structure of the galactic disc (Robin+ 1992)

Table 5
Most frequently copied A&A Tables.

Thanks to the installation of Tables and Abstracts, CDS has improved its service in collecting astronomical data, but also in the management of the SIMBAD database: the inclusion of the A&A bibliography is simplified (the titles of the papers are directly included in SIMBAD from the Abstracts) and the objects quoted in the electronic tables can be more easily incorporated than in the past.

7 Long-Term Storage

The electronic version of tabular data is therefore useful in many aspects: they are now easy to get, easy to process, to combine, with characteristic times which are orders of magnitude faster than in the past – the speed of light compared to the speed of terrestrial transportation. But the question of the lifetime of electronic publication – compared to the lifetime of printed papers – must inevitably be addressed: printed papers had a lifetime of several centuries, and we only know that no electronic data storage medium lives more than $\sim$ 10 years. And even if the lifetime of optical disks is claimed to be several decades, we're almost sure that new media will come up in the next decade, and the result will likely be that no device will be available any more to read the present optical disks...

The electronic media can however easily be saved by a copy to a new medium, which leads to a new concept of a *dynamic* lifetime. As an example, several original catalogues were initially stored on punched cards – the *Astrographic Catalogue* was for instance once a huge store of 2 million punched cards at Strasbourg Observatory – and were successively copied to new and more compact media. The multiple copies existing now all around the world also provide some guarantee of perennity.

8 Conclusions

The electronic tables and Abstracts from A&A are, to our mind, a good usage of the specificities of the electronic media to facilitate access to, and advertisement of, astronomical papers. Whether fully electronic publications are desirable, and should replace completely the printed edition, is a different story...

References

[1] Eckert, W. J., 1937, Publ. Astron. Soc. Pac. **49**, 249

[2] Egret, D., 1995, "Integrating astronomical data and information services at the CDS", this conference

[3] Harten, R. H., Grosbøl, P., Greisen, E. W., and Wells, D. C., 1988, "The FITS Tables Extension," *Astron. Astrophys. Suppl.* **73**, 365-372.

[4] Ochsenbein, F., 1994, *Bull. Inf. CDS **44**, 19*

[5] Ochsenbein, F., 1995, *Bull. Inf. CDS **46**, 29*