# THE PREPARATION AND DISTRIBUTION OF MACHINE-READABLE ASTROMETRIC DATA

Wayne H. Warren Jr.
Astronomical Data Center
National Space Science Data Center
NASA/Goddard Space Flight Center

## ABSTRACT

The acquisition, preparation and distribution of machine-readable astrometric data by the Astronomical Data Center (ADC) are described. Examples of certain general changes in data structure and format to improve compatability with other computers and software processing systems and to increase storage efficiency are discussed, as are the present data archive and request history of the ADC. Current development work in the areas of astrometric and positional catalogs is described.

## INTRODUCTION

The transition toward reduction of astrometric data almost exclusively by machine and the preparation and publication of resulting catalogs directly from computer files have resulted in the availability of an ever increasing number of machine-readable positional catalogs over the last several years. Astrometric or positional-type catalogs currently comprise about 16 percent of the total ADC archive, while their storage is approximately 30 percent of the archive, and this is without the many fundamental astrometric and meridian catalogs held by the Astronomisches-Rechen-Institut in Heidelberg and the U. S. Naval Observatory (see Jaschek 1977). It is clearly very important for new catalogs to be deposited with data centers so that future workers will have access to this valuable and often irreplaceable material. It is also important for deposited catalogs to be properly prepared and documented for distribution immediately, since original versions of catalogs are frequently not easily processable at other installations and their sources not generally known to other workers, at least after several years have passed.

New machine-readable catalogs are solicited by the ADC as they become known through publication or personal contact. They are prepared for archiving and documented as soon as possible while contact with the original author(s) or compiler(s) can be made whenever questions or



*H. K. Eichhorn and R. J. Leacock (eds.), Astrometric Techniques, 739–748.*

ambiguities arise. Generally, completed documents for new and recently acquired catalogs are sent to authors for comments and concurrence before they are distributed with catalog tapes. Procedures used to acquire catalogs and prepare them for archival are described in the following section, and examples of a few of the modifications made for homogeneity, minimization of storage and processing efficiency are given. The distribution of catalogs and the present archive are discussed in the next section, as are request statistics and the preparation of microform versions for distribution to the community. A few of the major developments concerning positional catalogs and possibilities for on-line retrieval of astrometric data in the near future are outlined in the final section.

## ACQUISITION AND PREPARATION OF CATALOGS

### 1. Acquisition and Receipt of Catalogs

Someone must have the responsibility for acquiring and archiving machine-readable astronomical data and preserving them for future generations. Logically, this responsibility should fall to the data centers because they are the only central repositories known (hopefully) to the entire community. Since many or most workers will not voluntarily take the time to prepare tapes of their catalogs and deposit them with a data center, data center personnel must remain continually aware of the literature and maintain contact with the community, and must solicit catalogs from authors and compilers as the data sets are produced. Catalogs considered to be of sufficient interest for archiving and subsequent possible distribution are solicited by the ADC. If an author agrees to prepare a tape and deposit it, ADC tapes are always supplied for the purpose, or authors' tapes are always returned after copying. The receipt of all materials by the data center is acknowledged and close contact is usually kept with an author until a catalog is archived and documented. Catalogs are identified very carefully, with specific references to the literature, to assure that all authors and compilers are given complete credit for their contributions. Individual and special treatment of each contributor and catalog help assure that future work will be deposited with the center voluntarily. Suggestions by and close cooperation with authors and compilers, who almost always have more expertise than data center personnel with the specific data being archived, help to guarantee that the data will be treated and documented in the best way possible.

### 2. Preparation of Catalogs for Archiving and Distribution

All catalogs received by the ADC are examined in detail before being archived and documented. This examination generally consists of checking the data records against a published version, if available, to insure that all data records are present and to ascertain if any data have been omitted from the machine version received. Certain data fields may be modified for increased processing efficiency, or even to

merely bring the data structure into conformity with similar data in other catalogs. If a catalog is received as a print tape, i.e. formatted with row spacing and column headings for offset printing, the data are reformatted to homogeneous records with all blank and text records removed. Similarly, if notes are included in the data file, they are detached and placed into a separate file so that the data file is homogeneous and easily sortable by machine. As examples of minor modifications made to astrometric catalogs, I describe and show changes made to two catalogs received recently.

The *Perth 75* (Nikoloff and Høg 1982) was received on floppy disks from Dr. Nikoloff. The floppy disks were read on a minicomputer and transferred to magnetic tape, on which the data were transferred to mainframe disk storage. The catalog turned out to be in excellent condition and only a few changes were made to effect uniformity with other catalogs: (1) minus signs in the coded spectral types were not always in the same column because the numbers had been written as integers; (2) the right ascension fields for 0 hours were all blank; zeros were added here and preceding zeros were inserted in hours and minutes fields where appropriate (single-digit values); (3) declination fields were non-uniform, e.g. minus signs occurred in different columns and were even in arcminutes fields for zero declination zones; these were homogenized as in (2); uniform numbers were constructed for the DA (= $\Delta\alpha \cos \delta$) and DD (= $\Delta\delta$) fields. Samples of parts of the original and modified records are given below (original first):

```
2142   5.53   5    2 00 37.633       -6 42.36   +0.00487  +0.0227   +.1
3  -.17 ....

2142   5.53   5   02 00 37.633   -00 06 42.36   +0.00487  +0.0227  +0.1
3 -0.17 ....
```

One data record (star 2764 at $9^h32^m$) was found with shifted (non-aligned) DA, DD and Epoch values during the modification process -- this was corrected.

The *Sydney Photographic Catalogue, Zone -48° to -54° (1950)* (Eichhorn *et al.* 1983) was received on magnetic tape from H. Eichhorn in 1982. Again, the data records were uniform and well formatted, but the tape had been prepared in a format suitable for printing and microfiche production, i.e. the data fields were widely spaced and there were 25 blank columns at the beginning of each record. Following consultation with the first author, who concurred with data record compression, the records were modified to remove all but one blank column between each data field; preceding zeros were added to all single-digit right ascension and declination fields. The changes decreased the logical record length per star from 104 bytes to 53 bytes, thus decreasing the storage required by almost a factor of two. The original record in the following example extends to a second line:

https://doi.org/10.1017/S0074180900077275 Published online by Cambridge University Press

```
                              -52 12251        B     16*        0   2 35.55
 -52 25 44.1       7.9      2     2

-52 12251    B     16* 00 02 35.55 -52 25 44.1   7.9 2 2
```

Following normal procedure, a descriptive document was prepared for the latter catalog (the former catalog document is currently in preparation). Each document conforms to a standard schema designed to present a brief description of the catalog, the record format, indigenous tape characteristics, remarks and modifications, and a sample listing of data records from the beginning and end of each tape file exactly as they are recorded on the tape. All changes made to a catalog, such as those illustrated above, are explained in the penultimate section so that if a user finds discrepancies between the published and machine versions of a catalog, an explanation is available. The document mentioned above was sent to Dr. Eichhorn in draft form for review, suggestions and/or comments, and concurrence; we intend to discuss the document at this meeting.

## DISTRIBUTION OF ASTROMETRIC CATALOGS

The large existing collection of machine-readable astronomical catalogs is currently divided among seven principal categories (see published lists of catalogs available in the *Bulletin d'Information du Centre de Données Stellaires*, Strasbourg and the *Astronomical Data Center Bulletin*, NASA Goddard Space Flight Center). Statistics for the archive of catalogs at the ADC are given in Table 1, where the size for each category is the total number of megabytes of data for all of the catalogs in the category. The ADC *Status Report on Machine-Readable Catalogs* actually contains 433 catalogs, but only those on master tapes or reasonably ready to distribute are included in the table.

Table 1. Statistics for the ADC Catalog Archive

| Category | Type | Number of Catalogs | Size (MB) |
|---|---|---|---|
| I | Astrometric | 59 | 1073 |
| II | Photometric | 87 | 214 |
| III | Spectroscopic | 73 | 126 |
| IV | Cross Identifications | 31 | 560 |
| V | Combined Data | 33 | 907 |
| VI | Miscellaneous | 13 | 320 |
| VII | Non-Stellar Objects | 58 | 317 |
| Total | | 358 | 3517 |

Astrometric and positional-type catalogs are certainly a major part of the archive, and well they should be, since computer processing plays at least as or more important a role in astrometry as in any other

https://doi.org/10.1017/S0074180900077275 Published online by Cambridge University Press

astronomical discipline and because the large sizes of positional catalogs make computer processing a necessity in modern data reduction and search techniques. With the exception of certain cross-identification catalogs, such as the Strasbourg *Catalog of Stellar Identifications* (Ochsenbein *et al.* 1981; Ochsenbein and Bischoff 1982) and data compilations, positional catalogs average the largest size, with the Jet Propulsion Laboratory Long Ephemeris (Newhall 1976) (included here with astrometric catalogs at 164 MB), the Astrographic Catalogs (149 MB), SAO (39 MB), Yale Zone (20 MB) and AGK3 (15 MB) catalogs heading the list.

Astronomical Data Center request activity has increased dramatically over the last six years, with positional catalogs being among the most in demand by the community. While much of their popularity can be attributed to a need for positional data in connection with space science and engineering activities, such as target acquisition and satellite tracking, many positional catalogs have been requested for purposes of automated chart preparation, object identification and conversion of catalog positions to alternate epochs and equinoxes. A summary of ADC growth in terms of the archive and requests completed is given in Table 2. Although most requests are for machine-readable data, many other products (documents, Status Reports, information) and services (data searches, bibliographical references, miscellaneous calculations, etc.) account for the large number of requests relative to the numbers of classical data center products listed in the table.

Table 2. ADC Request and Dissemination Statistics

| Year | Catalogs in Archive | Number of Requests | Catalogs Distributed Tape | Microform |
|---|---|---|---|---|
| 1978 | 75 | 40 | 43 | 2 |
| 1979 | 131 | 113 | 81 | 47 |
| 1980 | 254 | 232 | 163 | 20 |
| 1981 | 273 | 387 | 121 | 94 |
| 1982 | 360 | 548 | 257 | 76 |
| 1983 | 433 | 480 | 358 | 38 |

The reason for the low numbers of microform catalogs (mostly microfiche) distributed, aside from the fact that astronomers remain reluctant to use microform unless a catalog is unavailable in any other medium, is that most catalogs have still not been prepared in microfiche form. Presently, of the catalogs listed in the Status Report, only 46 are available in microform versions. This is due entirely to the fact that the preparation of a microform version is a difficult and time consuming task if it is to be self-contained, i.e. each microform catalog must include enough information to be conveniently used without reference to other materials or documentation. Since each catalog is unique, specific computer programs with extensive formatting must be

https://doi.org/10.1017/S0074180900077275 Published online by Cambridge University Press

written to provide a brief description of the data, column headings, page numbers and neatly arranged output. It is desirable for readability that the microform listing appear as in a printed catalog rather than the usual compressed form of magnetic tape; thus, each data record must generally be read and reformatted with decimal points, signs, and spacing before being processed to microform. Some thought has been given to the construction of a generalized software routine to enable the processing of any machine-readable catalog to microform, but this problem is extremely complex and not much progress has been made so far.

## DEVELOPMENT WORK

Most current development work involves the preparation of machine-readable versions of classical catalogs, the improvement and preparation of major catalogs already in machine-readable form, and the implementation of on-line retrieval systems to be accessible on a dial-up basis.

The large project to computerize the entire Durchmusterungen (DM) has progressed very well over the last year. This project is an international collaboration involving the Centre de Données Stellaires, the Observatoire de Nice and the Astronomical Data Center. The primary reason for producing machine-readable DM catalogs is to enable all DM stars to be added to the *Catalog of Stellar Identifications*(CSI), which is the kernal catalog for future automated retrieval of astronomical and astrophysical data. It is also envisaged that new computer generated DM charts can be produced as soon as the data can be machine processed. This would be especially useful for the southern DM catalogs, for which charts do not or only partially exist (CPD) or are not complete to the magnitude limit of the catalog (CD). The three northern zones of the *Bonner Durchmusterung* (BD) have been computerized by our French colleagues mentioned above, while the -1° to +19° zone and the *Cordoba DM* (CD) were punched at the National Space Science Data Center (NSSDC) and are currently undergoing proofreading and correcting with the help of many colleagues around the world who have proofread individual zones. Volume 1 of the CD (approximately 30 percent) has been completed and sent to the CDS, Strasbourg. The *Cape Photographic DM* (CPD) is now being punched and verified with funds provided by NSSDC and with the help of B. Rappaport, who had already prepared the first eleven most northerly zones as part of a computer graphics project. We expect to have the complete CPD by this summer. The southern BD (-2° to -23°) zone is currently being computerized at the CDS.

A new version of the *Smithsonian Astrophysical Observatory Star Catalog* (SAO) has recently been produced at the ADC and is nearly ready for distribution. The magnetic tape version contains cross identifications to the Boss *General Catalogue* and the *Henry Draper Catalogue* in addition to the DM identifications provided in the published edition. Identifications for Supplemental BD stars (lower case letters) have

been added and approximately 17000 data changes to 12000 records have been made. We hope to add positions and proper motions on the J2000 system as soon as certain problems with their computation have been solved.

We are currently preparing to load the CSI, version 1979 and sixteen associated source data files into a mass storage system on the IBM 3081 computer of the Science and Applications Computing Center at GSFC. Current ADC software for searching the CSI will then be extended to enable retrieval of specified source data from individual catalogs connected to the CSI. These source catalogs presently include the N30, Yale Trigonometric Parallax, and Nearby Stars (Gliese 1969 version), plus AGK3 and GC proper motions, in addition to catalogs in other disciplines. The source data retrieval pilot project is being undertaken partly in an attempt to make some data available to the community on a dial-up basis now, but also to give us some experience with such a system in preparation for a probable collaboraton with the CDS to transfer the new CSI and source database to GSFC for access by American astronomers. The current Strasbourg database contains about 200 connected catalogs (Egret 1983) and is now being extended to galaxies (Dubois *et al.* 1983). The *Bibliographical Star Index* (see Ochsenbein 1982) is now accessible through the database also, so that object bibliographies can be retrieved simultaneously with cross identifications and source data. These developments will hopefully provide immediate access to virtually all existing data and references and will save countless hours of library and literature searches now being performed routinely for most research projects.

## CONCLUSION

Astrometric and positional catalogs comprise a major portion of machine-readable data archives and can be an extremely valuable asset to astronomical research. Astronomical data centers have been initiated to provide data and services to the scientific community and we hope that our colleagues will take increasing advantage of these facilities. We especially hope that the authors and compilers of new machine-readable catalogs will deposit their data with any of the world network of astronomical data centers so that their co-workers can share in the fruits of their labors.

REFERENCES

Dubois, P., Ochsenbein, F. and Paturel, G. 1983, *Bull. Inform. CDS* 24, pp. 125-126.

Egret, D. 1983, *Bull. Inform. CDS* 24, pp. 109-123.

Eichhorn, H., Googe, W. D., Lukac, C. F. and Murphy, J. K. 1983, *Catalogue of 20457 Star Positions Obtained by Photography in the Declination Zone -48° to -54° (1950)*, *Astron. J.* 88, pp. 546-558.

Jaschek, C. 1977, in *IAU Colloquium No. 35, Compilation, Critical Evaluation and Distribution of Stellar Data*, ed. C. Jaschek and G. A. Wilkins (Dordrecht: Reidel), pp. 205-279.

Newhall, X. X. 1976, *Jet Propulsion Laboratory Long Ephemeris Tape of Lunar and Planetary Ephemerides for the Years 1410 BC to 3003 AD* (DE102, Jet Propulsion Laboratory).

Nikoloff, I. and Høg, E. (in collaboration with C. Ayers, Mrs. I. Nikoloff, Mrs. C. Spencer, Mrs. M. Sultana, A. Verveer and P. J. Høyer) 1982, *Perth 75, A Catalogue of Positions of 2589 FK4 and FK4S Stars* (Bickley, W. Australia: Perth Observatory).

Ochsenbein, F. 1982, in *IAU Colloquium No. 64, Automated Data Retrieval in Astronomy*, ed. C. Jaschek and W. Heintz (Dordrecht: Reidel), pp. 171-173.

Ochsenbein, F. and Bischoff, M. 1982, in *IAU Colloquium No. 64, Automated Data Retrieval in Astronomy*, ed. C. Jaschek and W. Heintz (Dordrecht: Reidel), pp. 211-213.

Ochsenbein, F., Bischoff, M. and Egret, D. 1981, *Astron. Astrophys. Supp.* 43, p. 259.

Discussion:

**GLIESE:** You mentioned that you had several Astrographic Catalogue zones on tape. What data are given in those files, spherical or rectangular coordinates?

**WARREN:** The zones of the Astrographic Catalogues that we have on tape now are those transcribed in Strasbourg by Lacroute and Valbousquet, so they contain equatorial (but not X,Y) coordinates, as well as magnitudes.

**CORBIN:** Have you been able to straighten out the southern part of the Algiers zone?

**WARREN:** No. There are problems with the positions in the southern hemisphere portion of the Algiers zone, apparently due to a bug in the reduction program used by Lacroute and Valbousquet. Tom Corbin is leading an effort at the USNO to rereduce the astrographic zones from X and Y coordinates, and he may wish to comment on this.

**EICHHORN:** The Cordoba Durchmusterung (CD) has the reputation of being not very reliable. Have you noticed that in your work with the CD or can you not comment at this point?

**WARREN:** I can comment on the fact that I have received some rather large error lists from people who have helped in the proofreading of individual Cordoba zones. I cannot comment further because I have not compared the machine version with any other catalogues (it is not finished yet). We hope to do some cross identification work once the CD and Cape Photographic Durchmusterung are completed. We can then compare them against each other to search for errors and inconsistencies. What we are doing now with the CD zones is that lists are being made of all errors detected during proofreading and individual zones are being corrected manually at a computer terminal. All published errata and corrigenda are then inserted into each completed zone, including supplemental stars (added with lower case letters appended as in the Bonner Durchmusterung). Each completed zone is then written to magnetic tape. All completed zones will be merged into a single tape file arranged north to south to prepare the final machine version.

**STRAND:** I have recently become involved in writing an article on astronomical catalogues and we have a problem here. We have finished position catalogues like the FK4, FK5 and so on; but we also have the compilation catalogue where you have compiled from many different sources, e.g., radial velocities, photometric data, and so on. Do you also call these astronomical catalogues?

**WARREN:** We do.

**STRAND:** Well, there is a significant difference between the two because one is a catalogue and the other an original catalogue which contains other data too.

**WESTERHOUT:** Such catalogues are called astronomical catalogues. You would not deny that The Catlogue of Bright Stars is an astronomical catalogue, or that Gliese's Catalogue of Nearby Stars is one?

**STRAND:** But when you have a catalogue like The Catalogue of Bright Stars, this is a finished catalogue and will not be added to, at least not for a certain number of years, unless you suddenly find some stars that were not included. When I talk about a compilation catalogue, I mean, for example, that next year somebody makes some 5000 radial-velocity measurements and they go into the catalogue of radial velocities exclusively. What do you call that?

**HEINTZ:** It has not been, obviously, the task of any data center, to modify existing catalogues apart from precorrecting errors, format work, and other technical details. These data remain the primary data catalogue until someone later compiles another section of the catalogue. But until then, they remain two different catalogues and the newly published radial velocities do not enter the previously existing catalogue.

**WESTERHOUT:** May I just make one comment for the record? As many people who work with catalogues already know, the Naval Observatory has undertaken to make sure that in another 5-7 years the entire Astrographic Catalogue X and Y coordinates will be available on magnetic tape. It depends on funding and personnel and so on when that will be done. A number of zones were done in other places; Strasbourg, for example, has done a large amount of work which was then used to make the position catalogue of Lacroute and Valbousquet. We now have almost all of the original Strasbourg X, Y catalogues and are in the process of making sure that the entire AC will be available in machine-readable form.

https://doi.org/10.1017/S0074180900077275 Published online by Cambridge University Press