LIBRARY COLLECTIONS OF MACHINE-READABLE ASTRONOMICAL CATALOGS

Wayne H. Warren Jr.
Astronomical Data Center (ADC)
National Space Science Data Center (NSSDC) /
World Data Center A for Rockets and Satellites
NASA Goddard Space Flight Center

ABSTRACT. Major astronomical facilities generally use a large number of machine-readable catalogs and data sets. These are requested over a long period of time, usually by individual staff members who seldom request data through their library services. An important drawback of this procedure is that the same catalogs are often requested by different persons at the same institute, thus requiring a duplicate effort on the part of the service organization (usually the data centers) and longer than necessary delays for the persons requesting data. It is suggested that libraries at major facilities build collections of important data sets and maintain library catalogs of these collections for use by staff astronomers. This paper discusses simplified and efficient methods of building such collections and details possible ways of storing and retrieving particular catalogs easily when they are required by staff members.

1. INTRODUCTION

The use of machine-readable astronomical data in basic research, telescope and satellite pointing and guidance, automated data collection and reduction, and educational applications has grown dramatically over the last decade. The wide availability of data through the international network of astronomical data centers and the dedication of data center personnel to the tasks of acquiring and disseminating high quality and well-documented data have contributed significantly to this increased usage of large computerized catalogs and other data to study problems that were done manually by former generations of astronomers.

The increasing use of centralized library facilities for computerized data and bibliographical searches has opened possibilities for the maintenance of library collections of machine-readable data sets that can be made available to all staff members of observatories and research institutions. (A similar project is currently underway for software documentation [Rhodes, Kurtz, and Rey-Watson 1988].) While the current trend in data acquisition is for individual astronomers to request data for their own use from central repositories, increased workloads on data centers having limited staff and facilities have had the general effect of producing longer turnaround times for the processing of requests. One solution to this problem is for observatories and institutes having major library facilities and qualified staff to interact with the data centers directly to build collections of machine-readable data sets and documentation for use by all local personnel needing these data. This would not only decrease the currently heavy burden now placed upon central data repository personnel, but would also open possibilities for the exchange of data and documentation among the astronomical institutes themselves without intervention by the data centers.

This paper outlines a method for building and maintaining a collection of machine-readable astronomical data sets at a library facility in order that they may be made available to local researchers and educators as needed.

https://doi.org/10.1017/S0252921100003158 Published online by Cambridge University Press

2. BUILDING A LOCAL ARCHIVE

A localized archive of astronomical data sets can be accumulated by requesting individual catalogs and other data from the nearest central repository (Astronomical Data Center [ADC], Centre de Données de Strasbourg [CDS], Soviet Centre for Astronomical Data [SCAD]) as they are needed. The local archive would currently consist of a set of master magnetic tapes and their backups corresponding to the categories of data maintained at the data centers and given in, for example, the ADC *Status Report on Machine-Readable Astronomical Catalogs* (SR) published regularly in the *Astronomical Data Center Bulletin* and distributed separately to requesters. Of course, very large data sets, such as spectra from the *International Ultraviolet Explorer* (*IUE*) and survey images from the *Infrared Astronomical Satellite* (*IRAS*) would require separate collections of tapes, but most of the astronomical catalogs, atlases, and collections of spectral data included in the SR can be stacked on a single set of tapes corresponding to the categories of astrometric, photometric, spectroscopic, cross identifications, combined and derived, miscellaneous, and nonstellar data used in the SR. As additional data sets are acquired from the data center, they are written to the correct tapes corresponding to their designations in the numbering system of the data centers. Thus, a set of fourteen tapes (seven "use" and seven "backups") can be used to maintain a large part of the collection of frequently used catalogs.

As individual master tapes and their corresponding backups become full, additional tapes are added for that category only. This procedure keeps the archive "under control," so to speak, and allows individual data sets to be located easily (as shown in the next section). All backup tapes should be stored in a separate location, of course, to decrease the probability of destruction of both sets of tapes simultaneously.

3. MAINTENANCE OF THE ARCHIVE

3.1 The Archive Medium

The master tape archive should, of course, be maintained on the highest density tapes available to the local institution. The quantity of data makes at least 6250 bpi almost mandatory, and higher densities, such as the 38,000 bpi cartridge-type tapes now being marketed, are a distinct advantage. This is the case not only for storage capacity, but also because of the more reliable recording techniques of the higher densities. While 6250 bpi quality tapes can be expected to last several years between failures with fairly heavy use, the new cartridge tapes are guaranteed to develop ≤ 2 I/O errors per decade. As the more advanced technologies, such as Write Many Read Many (WMRM) optical disks and higher density tape, become available, procedures for archive storage and maintenance will be simplified.

In the current world of magnetic tapes, however, one does need to be concerned with I/O errors and tape replacement. Tapes should be stored in a constant temperature and reasonably dust-free environment. When a tape goes bad, its corresponding backup should replace it as the "use" tape and its first use should be to create a replacement backup on a **new** tape. This procedure ensures that the master tapes are recycled to decrease the possibility of both primary and backup tapes developing I/O errors simultaneously.

There is also the possibility of maintaining the archive on a mass storage system with only a set of backup tapes. The current computer being used by the ADC to maintain its archive and to disseminate the data is the IBM 3081K system of Goddard's NSESCC (NASA Space and Earth Sciences Computing Center). This system has an attached IBM 3850 mass storage system with a capacity of 360 gigabytes; it currently holds the entire *IUE* archive of extracted spectra (some 61,000 spectral images). However, while magnetic tapes allow block sizes of up to 32K bytes (32,760 bytes actual maximum), the mass storage system uses IBM 3330-type format that only allows a maximum block size of 13,030 bytes. Thus, because data transfer charges are assessed per block of data, it costs less to transfer data from tape to tape than from mass storage to tape, and this is significant for the large catalogs maintained by the ADC. This constraint will probably no longer prevail for the next generation of mass storage devices, which will allow larger block sizes.

3.2 The Archive Record

To enable the retrieval of specific data sets when they are required by local users, a complete and concise record of the contents of each master tape must be kept. At the ADC, this is done by maintaining an online data set for each tape. These data sets contain information about each file of every catalog on every archive tape. The information contained in each data set includes catalog number, file number within the catalog, catalog name, file name, block size, logical record length, and number of logical records. A software program then computes additional parameters and produces a **MAGNETIC TAPE FILE RECORD** for each master tape. An example of such a record is shown in Figure 1. The program computes the length of each file and the total length of tape used, which is necessary information for transferring individual catalogs to users' tapes and for keeping a record of the amount of space available on the master. An asterisk (*) indicates that a file has been superseded by a later edition of a data set. Master tapes are periodically cleaned up and re-created, whence all flagged files are removed from the tapes and migrated to a tape of superseded catalogs. Since the tape files are not included in the file record data set, it is only necessary to delete the superseded record and rerun the program for that tape. Records are added to the master tape data set from another data set containing file records for all catalogs in the archive. The program is also used to generate file records for requesters' tapes that are generated by funneling catalogs from the master tapes to each requester's tape. Thus, users of the facility receive a complete record of their tapes that is generated automatically after the tape is created.

4. SUMMARY AND CONCLUSIONS

Major centers of astronomy having centralized library facilities are encouraged to produce and maintain local archives of machine-readable astronomical data sets by acquiring the data from central repositories and creating their own archiving and retrieval systems. Such a procedure will allow local users to obtain data that they need much faster and will decrease the burden now placed on the astronomical data centers. A regional data center, the Canadian Astronomy Data Center (CADC), is already operational at the Dominion Astrophsyical Observatory and has produced a user's manual (Justice, Durand, and Crabtree 1988); a similar data center has been proposed for Australia (Tuohy 1987). The CADC maintains not only a wide variety of data sets for distribution to Canadian astronomers, but also provides a basic collection of services as well, including access to the STARCAT (Space Telescope ARchive and CATalog) system developed at the ST European Coordinating Facility and the European Southern Observatory, the SIMBAD (Set of Identifications, Measurements and Bibliography for Astronomical Data) data bank of the CDS, and a variety of local data reduction programs.

Further information and assistance with the creation of local archives of astronomical data can be obtained by contacting the author.

REFERENCES

Justice, G., Durand, D., and Crabtree, D. 1988, *CADC Services Reference Manual* (available from the Dominion Astrophysical Observatory).

Rhodes, C., Kurtz, M., and Rey-Watson, J. 1988, in *Astronomy from Large Databases*, ed. F. Murtagh and A. Heck, *ESO Conf. and Workshop Proc.* No. 28, p. 459.

Tuohy, I. 1987, *Proc. Astron. Soc. Australia* 7, 80.

A S T R O N O M I C A L   D A T A   C E N T E R

MAGNETIC TAPE FILE RECORD

Name: Master Tape 5A1 ; Date: 24 JUL 1988; Density: 6250 bpi; Coding: EBCDIC; Tracks: 9

Catalog Number	File	Catalog Name	File Name	Tape File	Number of Blocks	BLOCKSIZE Max	BLOCKSIZE Min	Blocking Factor	RECFM	LRECL	Number of Records	Length (feet)
5032		Stars < 25 pc (Woolley et al. 1970)		1	9	32670	28890	242	FB	135	2150	4
5031		Kinematic Data O-B5 Stars(Mead,Hill 1981)		2	8	32757	28461	183	FB	179	1440	4
5801		Nearby Stars (Gliese 1969) Selected		3	5	32720	20320	409	FB	80	1890	2
5802		Yale Bright Star 3rd Ed (Hoffleit 1964)		4	57	32724	7676	162	FB	202	9110	26
5902		Yale Bright Star with GC Data (Nagy 1978)		5	93	32634	31302	98	FB	333	9110	43
5825	1	Yale Bright Star 4th Ed (Hoffleit 1982)	Data	6	60	32648	5088	154	FB	212	9110	28
5825	2	Yale Bright Star 4th Ed (Hoffleit 1982)	Remarks Intro	7	1	4480	4480	56	FB	80	56	0
5825	3	Yale Bright Star 4th Ed (Hoffleit 1982)	Remarks	8	33	32736	3036	248	FB	132	7959	15
5701		Nearby Stars (Gliese 1969)		9	8	32750	18209	250	FB	131	1889	4
5836	1	Yale Bright Star Suppl (Hoffleit 1983)	Introduction	10	2	32720	240	409	FB	80	412	1
5836	2	Yale Bright Star Suppl (Hoffleit 1983)	Data	11	17	32648	31164	154	FB	212	2611	8
5836	3	Yale Bright Star Suppl (Hoffleit 1983)	Remarks	12	8	32704	256	511	FB	64	3581	4
5045	1	Comb List Astron Sources 3.1 (1985)	Data	13	955	32702	22908	197	FB	166	188076	440
5045	2	Comb List Astron Sources 3.1 (1985)	Notes	14	1	18320	18320	229	FB	80	229	1
5045	3	Comb List Astron Sources 3.1 (1985)	Flux Notes	15	1	3920	3920	49	FB	80	49	0
5039	1	Fourth Orbits Visual Binaries (1983)	Introduction	16	1	21040	21040	263	FB	80	263	1
5039	2	Fourth Orbits Visual Binaries (1983)	Data	17	12	32736	9372	248	FB	132	2799	6
5039	3	Fourth Orbits Visual Binaries (1983)	Notes	18	2	32720	16400	409	FB	80	614	1
5038	1	Ultraviolet Star Cat UCL (Carnochan 1978)	Description	19	1	13360	13360	167	FB	80	167	1
5038	2	Ultraviolet Star Cat UCL (Carnochan 1978)	Data	20	249	32760	10920	126	FB	260	31290	115
5840	1	Spectroscopic Bin 7 (Batten et al. 1978)	Data *	21	7	32640	3672	160	FB	204	978	3
5840	2	Spectroscopic Bin 7 (Batten et al. 1978)	References	22	3	32720	12800	409	FB	80	978	1
5048	1	Field Type II Cepheids (Harris 1985)	Table I	23	1	10640	10640	152	FB	70	152	0
5048	2	Field Type II Cepheids (Harris 1985)	Table II	24	1	3920	3920	56	FB	70	56	0
5705		Colors, Lum, Mot Nearer K-M Giants (1966)		25	3	32760	7344	455	FB	72	1012	1
5051	1	Data Inv Space-Based Obs 1.1 (DISCO) 1986	Catalog	26	271	32686	3658	277	FB	118	74821	125
5051	2	Data Inv Space-Based Obs 1.1 (DISCO) 1986	References	27	1	10480	10480	131	FB	80	131	0
5840	1	Spectroscopic Bin 7 (Batten et al. 1978)	Data	28	7	32640	3672	160	FB	204	978	3
Totals: Number of Catalogs: 16				28								837

Figure 1. An example of a MAGNETIC TAPE FILE RECORD for an ADC master archive tape.

https://doi.org/10.1017/S0252921100003158 Published online by Cambridge University Press