

PART IV

CATALOGUES AND DOCUMENTATION

# INFORMATION AND CATALOGUES

C. JASCHEK

*Observatorio Astronómico, La Plata, Argentina*

The information used by scientists belongs to three basic types: (1) data, (2) procedures (techniques or description of instruments) and (3) theories and ideas. In what follows we shall consider only the first type of information, namely data. We will call 'data' all information related to or derived from observable quantities, like radial velocities, spectral types, color indices, etc.

Everyone knows by experience that the information flow has been increasing steadily over the last years and we all hope that this trend keeps on. But we know also by experience that it becomes more and more difficult to keep up with the information inflow. This poses the obvious question of which is the best way to keep astronomers informed and this is precisely the central topic of the lecture.

Probably the best way to start is to furnish some figures on the information already available and the information inflow, for the kinds of data related to the field of our Commission, i.e. spectral types, magnitudes and colors.

Let us start with spectral types. It is well known that the largest number of spectral classifications was carried out in one-parameter systems, i.e. in temperature schemes, like the ones of the Harvard Observatory. Only years later, two-parameter classifications were introduced. In what follows we will call pure temperature classifications (like the HD) systems of type I; two-parameter classifications (like the MK), systems of type III and classifications in which occasionally luminosity designations are used (in the form of prefixes *d* or *g*), systems of type II. Table I provides a gross survey of the major projects in spectral classification. Projects with less than two thousand stars classified were not listed. From the data assembled it can be seen that at least  $8 \times 10^5$  stars do have at the present some type of spectral classification. If one makes allowance for the projects not included in the table, the total would probably become  $9 \times 10^5$ . This can be compared to the total numbers of stars up to photographic magnitudes 11<sup>m</sup> and 12<sup>m</sup>, which are  $7 \times 10^5$  and  $18 \times 10^5$  respectively. This implies then that on the average the number of stars having some spectral classification is comparable to the number of stars of eleventh magnitude. This is of course an artificial comparison, since only the HD is statistically complete down to about ninth magnitude, whereas fainter stars were never systematically surveyed.

In contrast with this, we have only about  $25 \times 10^3$  stars classified in the MK system. We all hope that thanks to the efforts of our Michigan colleagues this situation will change drastically in the next few years with the reclassification of all stars contained in the HD, using the MK classification system.

One might wish to raise incidentally the question of how far one can expect to go with the present technology. Haffner (1960) has examined this question very thoroughly and has shown that if one wishes to derive MK types (dispersion around  $100 \text{ \AA mm}^{-1}$ )

TABLE I  
Large projects of spectral classification

Name and author	S	N(10 <sup>3</sup> )
Henry Draper Catalogue – Cannon and Pickering	I	225
Henry Draper Extension – Cannon	I	47
Annie Cannon Memorial Volume	I	87
AG stars in Yale Transactions	I	17
Cape Catalogue of Faint stars – Cannon	I	12
Bergedorfer Spektral-Durchmusterung – Schwassmann and Wachmann	II	173
Potsdamer Spektral-Durchmusterung – Becker and Brück	II	67
Selected Areas – Humason	II	4
McCormick –Vyssotsky, Balz <i>et al.</i>	II	60
Mt. Wilson – Adams <i>et al.</i>	II	4
Abastumani – Bartaia, Kalandadze <i>et al.</i>	I	40
Crimea – Kopylov, Brodskaia <i>et al.</i>	I	22
Kiev – Gordeladse <i>et al.</i>	I	3
Luminous stars Northern Milky Way – Stock, Nassau <i>et al.</i>	II	10
Stockholm – Elvius, Loden <i>et al.</i>	II	24

Notes: S = spectral classification scheme,  
N(10<sup>3</sup>) = number of stars, in thousands.

by using objective prisms mounted on Schmidt type cameras, it is possible to reach 10<sup>m</sup>.5 with 80-cm cameras and the existing emulsions. This limiting magnitude implies  $5 \times 10^5$  stars, a number which is just on the edge of what can be done by non-automated classification.

Let us next consider the important question of the information growth. Leaving aside the big Michigan project, one can try to answer the question separately for type I+II classifications and for type III classifications.

An order of magnitude estimate can be made by counting the stars contained in the lists published in the last years, for type I and II classification schemes. These data are given in Table II and were obtained from the IAU *Reports on Astronomy*. The annual rate is about  $9 \times 10^3$ , but by including smaller lists (excluded from Table II)

TABLE II  
Stars classified annually in  
type I or II schemes

Years	N(10 <sup>3</sup> )
1958–61	28
1961–64	10
1964–67	15
1967–70	48
Total	101
Annual rate	9

one is likely to arrive at a figure of about  $10\text{--}11 \times 10^3$  stars classified per year.

For MK classifications the figures are somewhat more difficult to obtain, because lists of stars published in this system contain very often only small numbers of objects. An earlier estimate (Jaschek, 1968) was that information doubles every 8 yr, which implies an annual increase of about 9%. Since the publication of the only existing catalogue of MK spectral types, which contained  $15 \times 10^3$  stars, an increase of about 7000 stars is therefore expected. This makes it urgent to publish a second edition of the MK type catalogue, a task which we expect to start next year at La Plata. The importance of collecting the MK type information is underlined by the fact that this information refers mainly to bright stars, which are fundamental for photometry.

Let us turn next to color indices. It should be mentioned right at the start that the information in this field is rather poorly organized, despite the fact that color indices are being observed since many years ago.

The color indices used today are either photographic or photoelectric. Since in the past photographic color indices had a much lower accuracy than photoelectric ones, the modern tendency has been to disregard them completely, an opinion which I feel not qualified to discuss. In Table III are listed the most important lists of photographic color indices, omitting all lists containing less than thousand stars. Perhaps the most impressive fact is that except in the south (Jackson and Stoy, 1954–1962) where we have practically complete data between  $-30^\circ$  and  $-80^\circ$ , no comprehensive schemes equivalent to the HD in spectral classification were attempted. This is rather surprising, since photoelectric photometry cannot be expected to rival with photo-

TABLE III

Large projects of photographic color indices

Author	$N(10^3)$
Malmquist (1927)	2
Malmquist (1936)	4
Bertaud (1939)	3
Seares, <i>et al.</i> (1941)	2
McCuskey and Seyfert (1950)	2
Kharadse (1952)	14
Jackson and Stoy (1954–1962)	68
Brodskaja (1955)	6
McCuskey (1955)	2
Eklöf (1958)	2
Brodskaja and Shajn (1958)	3
Numerova (1958)	5
Pronik (1958)	4
Kotshlashvili (1958)	2
McCuskey (1959)	3
Brodskaja (1960)	3
Metik (1960)	3
Bartkus (1964)	2
Masnauskas (1964)	4
Total	136

graphic photometry in either speed or number of stars. This can be seen very clearly through the following figures. The catalogue of Blanco *et al.* (1968) contains  $20 \times 10^3$  stars measured in the UBV system, a system which came into use in 1953 and which Blanco follows up to 1967. Our own La Plata reference catalogue which lists *all* photoelectric measurements in any color system and to which I shall refer later, includes some  $25 \times 10^3$  stars, observed between 1913 and 1968. This implies that the duplication has been very large, which is, at least partially, due to the lack of a catalogue telling what was already observed. It should however be pointed out that neither one of the catalogues includes stars in clusters, in external galaxies or stars for which no positions were published. An estimate of this number gives about  $10 \times 10^3$  additional stars.

We can therefore summarize the situation in the following way:

- (a) no effort has been made to derive photographic color indices in a systematic sky survey;
- (b) about  $2 \times 10^5$  photographic color indices exist;
- (c) photoelectric photometry exists for about  $35 \times 10^3$  of the brighter stars, many of them having been measured several times in different color systems.

The next question concerns the growth of the information. For photographic colors we have estimated the growth according to the same lines as for spectral types, and the result comes close to about  $10^4$  stars per year. This comes mainly from the fact that our Soviet Union colleagues determine usually both the spectral type and a photographic color index.

For photoelectric colors we will use an earlier estimate, which gave a figure of about 12% annual, doubling thus the information every six years.

The results for both spectral types and color indices are summarized in Table IV. If one focuses on photoelectric colors and MK types, it seems that as an order of magnitude the information increases by about 10% annually, implying that information duplicates in about seven years. This of course is very satisfactory, provided that one can use the new information-but this is, alas, not so.

A simple perusal of the literature shows that most papers, either of spectral classification or of photoelectric photometry, contain data relating to something between

TABLE IV  
Information growth

	$N(1970)$	a.g.	g
Spectral classifications, type I and II	$7 \times 10^5$	$1.1 \times 10^4$	2
type III	$3.5 \times 10^4$	$3 \times 10^3$	9
Photographic photometry	$2 \times 10^5$	$1 \times 10^4$	5
Photoelectric photometry	$3.5 \times 10^4$	$4 \times 10^3$	12

Notes:  $N(1970)$  = number of stars observed up to 1970,  
a.g. = annual growth,  
g = annual growth in percentage.

ten and hundred objects. Let us assume at the best that the average number of stars per paper is hundred. Then for increases of about  $4 \times 10^3$  stars per year, we have a minimum of about 40 new papers. The real figures are different, because already in 1960 there were for MK types 55 papers, and for photoelectric photometry 127 papers, 54 of them giving data for more than 10 stars. This means that in order to locate the parameter of a given object one has to go through at least 40 papers per year .... I am quite certain that most of us will decide either to observe the star anew or to forget about it, before going to the library.

The inescapable conclusion is that updating catalogues have to be published rather rapidly.

If one accepts that a new catalogue is in time when more than two hundred papers were published on the subject, one arrives at a minimum time interval of five years. From the abovementioned quotation of a duplication every seven years it is evident on the other hand that this time interval should be nearer to three years than to five.

Speaking mainly from the experience of the three catalogues undertaken at La Plata, it seems that this is so short an interval that one has to proceed on a continuous basis, because usually catalogues take much longer than this to assemble. The MK catalogue took four years, the Be star catalogue four years and the photometric catalogue is taking six years. Similarly the photometric catalogue of Blanco *et al.* took six years. I think therefore that my first point is proved, namely that in order to cope efficiently with the information inflow one must envisage a continuous publication system.

This really implies the creation of a Data Center, because the traditional way of an astronomer near retirement publishing a catalogue and not worrying about its continuation, is clearly not in step with the times.

It is however important to stress that the idea of data centers is really nothing new. The Index of Numerical Data Projects (1969) shows that in 1968 there existed 157 data centers in the world compiling numerical data of different kinds in sciences and technology. Even in astronomy a preliminary report of IAU Commission 1 on *Data Centers Reported* showed the existence of eighteen institutions doing data center work, on the following subjects:

Atomic Energy Levels

Cross Sections for Collisions of Electrons and Photons with Atoms, Ions and Small Molecules

Eclipsing Binaries

Ephemerides, Planetary Data, Star Catalogues

Extra-galactic Objects

Globular Clusters

Observatories, Instruments

Planetary Research Center

Star Catalogues

Radio Sources

Spectroscopic Binaries

Transition Probabilities

to which at least two very important groups should be added: the Variable Star Catalogue group at Moscow and the Czechoslovakian group (Alter, Ruprecht and Vanyseka) working on clusters and associations.

Let us now explore in more detail what the most pressing needs are in each field and what can be done. We will start again with the spectra and discuss successively literature completeness, selectivity and values to be published.

Completeness is an ideal which can never be reached, except asymptotically and at a very high cost. Usually it suffices to obtain a 'reasonable' completeness. Is it 'reasonable' to try to assemble complete spectral information of all stars? Most of the astronomical community will probably expect something of this kind from a data center, but a little reflection shows that this is rather difficult. Although HD and the 'Luminous stars' are rather easy to deal with, because they both provide positions, most of the HDE is only on charts and so are practically all classifications in special regions, like for instance those of McCuskey and Brodskaja. After thinking it over some time I feel that the best policy is to assemble only the data for stars having a published position. For work until the tenth magnitude, the position can be rounded off to  $0^m1$  in  $\alpha$  and  $1'$  in  $\delta$ . If fainter stars are to be included,  $0^s1$  and  $1''$  must be kept; a system which implies among other things an improvement of the positions of all HD stars. The application of this criterium means however that a large section of the spectral classification literature will be left out and eventually be forgotten. So my second proposal is to establish a masterlist of spectral classification programs describing:

- (a) the limits of the region studied (for some standard equinox);
- (b) the magnitude limit;
- (c) the classification system used;
- (d) the purpose of the study (search for emission line stars, galactic structure, etc.).

This list should be supplemented by charts, so that a simple glance at it can tell if there is some special field in the region one is particularly interested in.

The next question concerns selectivity. What is one to do if a star has been classified in the HD, in the Bergedorfer Spektral-Durchmusterung and in the MK system? This problem is really part of a larger one, namely to decide between critical and bibliographic compilations. In order to answer this question I have made an inquiry among my colleagues and the general opinion is that it is preferable to set up a selection scheme, which starts at the top with MK, and passes next to Mt. Wilson, HD and others. If several classifications exist for a given object, only the topmost classification is given, whereas the others are kept in files for reference.

A complication arises when there are systematic differences between the series from which the particular classifications come. On one side it is better to have a single system, while on the other hand original values are always more valuable than reduced values. Up to now it is not entirely clear (at least to me) what is the best procedure.

The third question is related to the preceding ones, namely what to publish. What does one do for instance with MK types if several differing ones are available? In the La Plata catalogue we listed everything, feeling that this is the only fair procedure

to all colleagues. In the introduction to the catalogue some general rules were laid down, of what to do when several discrepant classifications were available. Because of this 'bibliographic criterium' we have been under criticism from several sides, the main objection being that only a single classification – 'the best one' – should be provided. In order to see what could be done, I consulted again several colleagues and the opinion was that although difficult, a weighting could possibly be performed, in the sense that everything should be listed, but some type be underlined, stressing that *in the opinion of the compiler* this is the best type. This combines probably the best of two conflicting tendencies (and leaves one with the smallest number of enemies). As a practical consequence we expect to start in 1972 the continuation of our MK type catalogue from 1962 on, along these lines.

As for the other things mentioned, we will also attempt to publish soon the masterlist of regions where spectral classifications were done, but we will leave for the very last the attack upon the 'complete' bibliography.

Assuming that all this will be done in a reasonable time, one could think perhaps that the information problem is solved. But there is still one important aspect, namely the one of the non-normal spectra. In this case, bibliographic completeness should be attempted, because here even small bits of information can be important. A list of the partial catalogues made in this field is given in Table V.

TABLE V  
Bibliographic catalogues of special groups of stars.

Be stars	Merrill and Burwell (1933, 1943, 1949, 1950) Jaschek <i>et al.</i> (1971)
Emission line stars	Wackerling (1970) Bertiau and McCarthy (1970)
Late type emission line stars	Bidelman (1954)
Combination spectra	Hynek (1938)
Ap and Am stars	Bertaud (1959, 1960, 1965)

A glance at this table shows that several important groups were never covered like WR and Of stars, F peculiar star, late type giants, etc. This is an important omission and eventually it should be closed.

We will consider next color indices. Although I am not a photometrist, my justification for discussing this problem is that we are in the process of publishing the La Plata photometric catalogue, a project which has put me in touch with some of the problems one is likely to find in this area.

Let us start with the completeness problem. It has been doubted by many, if it is really a good idea to collect all the photographic color indices, mainly because of the large errors they have. On the other hand it would be only fair to say that the publication of a masterlist of all the regions in which photography photometry was once done, together with an assessment of the errors, would be a very valuable under-



taking. To say the least, they could constitute a convenient starting point for photoelectric photometry. Since the effort to put up such a list is several orders of magnitude less than collecting all measurements, it is probably a step in the right direction. Reduction to a common system, like for instance reduction to something like the  $B-V$  index, would enhance certainly the value of published photographic color lists, but again here it would be necessary to examine if the effort is worthwhile, and this problem I leave to my photometric colleagues.

If one turns now to photoelectric photometry, the situation is slightly more encouraging. Of the 1900 papers listed up to 1968 in our bibliography, 870 deal with  $UBV$  photometry, 72 with multicolor photometries, 69 infrared measurements, 760 with unspecified systems and 84 with narrow band photometry. The majority of the stars which have been measured photoelectrically do have  $UBV$ , this being shown by Blanco's catalogue in which he lists 20000 stars measured in the  $UBV$  (and  $U_cBV$ ) while our catalogue listing everything at a somewhat later date, contains only about 25000 stars. For some time in the future,  $UBV$  will be thus the most important photoelectric system; Strömngren's narrow band photometry following in second place and Golay's system third, sharing its place almost with Johnson's  $UBVRI\dots$  and Stebbins, Whitford and Kron's six color photometry. I have quoted these details because of the need to underline that there are few photoelectric systems containing more than say three thousand measured stars.

The implications for data centers are obvious, namely that one can concentrate on about half a dozen photometric systems and probably disregard the rest, except for very special purposes. This is essentially the idea of Dr B. Hauck and his group at Geneva. Since Dr Hauck will give later a report of his work, I will not go into more details.

Until now we have not discussed the more technical aspects, namely the way the data are stored and the way they are distributed. We will now mention them briefly.

With regard to data storage it seems that most people are inclined to punch cards, because this way the information can be subsequently handled easily, and even published without further complication. At La Plata we have always worked with handwritten cards, because it seemed simpler and because we had no punching unit the time we started. Since I have no experience with handling punched cards, I can only make the comment that whatever one does, cataloguing involves a great deal of cross checking, correcting (or trying to correct) misprints and misidentification, and it seems to me that this can be done with ease on handwritten cards. Of course the matter is quite different when it comes to printing, because then one has essentially to duplicate the work already done when writing the card. At this stage at least the punched card file is definitely superior.

With regard to publication policy, the inquiry I mentioned before has not revealed a definite tendency among prospective users. In principle one can envisage a passive data center or an active center. The passive data center stores information and releases it only on request. This system is good only if the information is stored on punched card or tape. It is a cheap system, in the sense that the customer pays what he buys,

but on the other hand the usefulness of the center is restricted to those who ask and can pay for the questions. For instance for countries with exchange restrictions this difficulty could spoil the usefulness of the center. Also it would make it more difficult for non-specialists to obtain the occasional data they need, because by experience one knows that if one has to write a letter for each object which comes up, generally one prefers to drop the subject. In active centers instead all information is published at intervals and distributed to all colleagues. This policy seems to be more in accord with the general psychology. Within this general policy, one can envisage a regular catalogue every X years and supplements at shorter intervals. The financing could perhaps be covered by the sale of the catalogues.

Let me now come to the conclusion of my paper and state briefly what is being done in the field of our Commission on the data center problem.

European astronomers have decided to create, a few years ago, the 'Centre de Données Stellaires' (Stellar Data Center) at Strasbourg, through the collaboration of the Observatories of Paris, Heidelberg, Marseille, Lausanne, Strasbourg and La Plata. Each of these observatories will take one chapter of stellar data and act as a data center for this field. B. Hauck, Lausanne, will be in charge of the photometry and we, at La Plata, will do the same for spectral classification. F. Spite, Paris, will be in charge of the 'bibliographic file' which consists of an inventory of all bibliographic mentions of a given star in the literature. This very ambitious project will start with the publication of the mentions contained in a dozen leading astronomical journals in the last decade; later on it will be extended to more journals and back in time. T. Lederle, Heidelberg, will be in charge of the positions and proper motions and Mme. M. Barbier, Marseille, of the radial velocities. J. Jung, Strasbourg, finally will be in charge of the Center itself and of the combined files of all sub-centers.

Attention is called to the Information Bulletin of the Stellar Data Center, which provides the latest news about the project.

We all hope that this data center constitutes a first step toward our goal, which is to make available all stellar data to all colleagues.

### References

- Bartkus, R.: 1964, *Bull. Vilnius* **13**.  
 Bertaud, Ch.: 1939, *Ann. Obs. Paris, Sect. Meudon* **8**, f.3.  
 Bertaud, C.: 1959, *J. Obs.* **42**, 45.  
 Bertaud, C.: 1960, *J. Obs.* **43**, 129.  
 Bertaud, C.: 1965, *J. Obs.* **48**, 211.  
 Bertiau, F. C. and McCarthy, M. F.: 1969, *Specola Vatic. Ric Astron.* **7**, 525.  
 Bidelman, W. P.: 1954, *Astrophys. J. Suppl.* **1**, 175.  
 Blanco, V. M., Demers, S., Douglass, G. G., and Fitzgerald, M. P.: 1968, *Publ. U.S. Naval Obs., 2nd Ser.* **21**, 3.  
 Brodskaja, E. S.: 1955, *Comm. Crimea* **14**, 139.  
 Brodskaja, E. S. and Shajn, P. F.: 1958, *Comm. Crimea* **20**, 299.  
 Brodskaja, E. S.: 1960, *Comm. Crimea* **24**, 160.  
 Eklöf, O.: 1958, *Medd. Uppsala Astron. Obs.*, No. 120.

- Haffner, H.: 1960, *Astron. Nachr.* **286**, 28.
- Hynek, J. A.: 1938, *Contr. Perkins Obs.* **1**, No. 10, 185.
- International Compendium of Numerical Data Projects*: 1969, ed. Comm. data for Sci. and Tech., Springer Verlag.
- Jackson, J. and Stoy, R. H.: 1954–62, *Ann. Cape Obs.* **17**.
- Jaschek, C.: 1968, *Publ. Astron. Soc. Pacific* **80**, 654.
- Jaschek, C., Ferrer, L., and Jaschek, M.: 1971, *La Plata Publ. Ser. Astron.* **37**.
- Kharadse, E. K.: 1952, *Bull. Abastumani* **12**.
- Kotshlashvili, T. A.: 1958, *Bull. Abastumani* **22**, 67.
- Malmquist, K. G.: 1927, *Lund. Medd, Ser. II* **37**.
- Malmquist, K. G.: 1936, *Stockholm Obs. Ann.* **12**, No. 7.
- Masnauskas, J.: 1964, *Bull. Vilnius* **12**.
- McCuskey, S. W. and Seyfert, C. K.: 1950, *Astrophys. J.* **112**, 90.
- McCuskey, S. W.: 1955, *Astrophys. J.* **122**, 359.
- McCuskey, S. W.: 1959, *Astrophys. J. Suppl.* **4**, 1.
- Merrill, P. W. and Burwell, G. G.: 1933, *Astrophys. J.* **78**, 87.
- Merrill, P. W. and Burwell, G. G.: 1943, *Astrophys. J.* **98**, 153.
- Merrill, P. W. and Burwell, G. G.: 1949, *Astrophys. J.* **110**, 387.
- Merrill, P. W. and Burwell, G. G.: 1950, *Astrophys. J.* **112**, 72.
- Metik, L. P.: 1960, *Comm. Crimea*, **23**, 60.
- Numerova, A. B.: 1958, *Comm. Crimea* **19**, 230.
- Pronik, I. I.: 1958, *Comm. Crimea* **20**, 208.
- Seares, F. H., Ross, F. E., and Joyner, M. C.: 1941, Carnegie Inst. Washington, Publ. 532.
- Wackerling, L. R.: 1970, *Mem. Roy. Astron. Soc.* **73**, 153.