

A. EXISTING DATA CENTERS

THE STELLAR DATA CENTER

Jaschek C.
Centre de Données Stellaires - Strasbourg (France)

ABSTRACT

The paper presents an overview of the problems astronomical data centers are facing today.

When one organizes a colloquium, one can do some fairly outrageous things, like putting one's own contribution first. However as an excuse for such a procedure, I shall say that I want really not to discuss that much what the C.D.S. itself does, but what problems we have been confronted with, specially those which came up since the last Strasbourg colloquium, or which got more complicated to handle.

Until very recently, Data Centers were mostly distributors of magnetic tapes. This is a nice way of sending out information, if the user and the provider live in the same country and use compatible computers. If they live in different countries usually there exist mailing and custom problems. Tapes from Paris to Madrid have been found to have velocities of 20 km/day which is less than stage coach velocities of the XIX century. However recently direct access to data bases became feasible, in real time, like for instance through EURONET in Western Europe. One can say confidently that within a few years all major observatories shall be linked to some data center - regional or national - so that many operations which seem unrealistic today will be regarded as routine operations in a near future. For instance checking from the observers dome if the star one sees in the finder as double is really double or if one got the wrong star, or if the other star brighter than 9^m near the object one is observing is a nova or a known object. Similarly one will be able to ask from one's desk for all the known data about the star one wants to study, and for the list of all papers mentioning the object. Such a facility is already available at Strasbourg and can be seen in the adjoining room. This is the facility which was already provided to European astronomers working on I.U.E. at Madrid, and so far the reactions have been very positive - each space observer saves a lot of

time knowing which of the parameters of the object are already known from ground based astronomy. It is curious that several ground based observatories which have been offered the same facility as E.S.A., have hesitated to use it. Why ? Because they think these data should be obtained and paid by the observers before they use the telescope. Let us hope that in the next decade we can convince our colleagues of the usefulness of the facility !

It is sure that the direct interrogation of data bases in real time from any observatory constitutes an important advance in data accessibility and I am sure it will enhance the importance of the work done at the data centers. This is certainly an important point for all of us, since we know that up to now data collecting activities have not been exceedingly popular projects in many countries. There exists a widespread feeling that it is a somewhat second order activity. Although rightly no one could say that we are doing "frontier research" (as the people with the largest telescopes) it is important to remark that in natural sciences like astronomy, any single astronomer cannot afford to start from scratch - he has to rely to a very large extent upon the work of his predecessors or his colleagues. Therefore data center activities, which represent the collective memory should be regarded as being work of the same level of importance as any other scientific project.

One thing which makes data center activities less attractive to many people is the fact that it is largely anonymous. Whereas the author of a catalog is usually well remembered, people at data centers are usually oftener forgotten than remembered, and the ready accessibility to data via networks will contribute even a little bit more to this. It seems to me that one should insist explicitly on the recognition of the services provided. If it is common use to thank the directors of the observatory where one made his observations, or his computations, I do not see why one should not ask for a similar recognition if data centers or bases have been used extensively.

One can ask next as to what shall happen at the moment every one can interrogate data banks from his working place. Shall we insist on delivering only standard data of a given type, or should we condense or process the data further ? Up to now generally we have been content to store data and to disseminate them, but sooner or later the problem of condensing data will come up, specially if satellites provide us with many more data. We hope to learn more on this point on Thursday from the Director of the largest existing facility.

Another instance where data processing could be very useful is for surveys like H α or 21 cm - here the possibility to generate iso-

photos would certainly enhance the value of the data. Now this question opens others, because if we do that, we will be asked to provide superposable displays of different kinds and so on and so forth. Thus it seems logic to start discussing where one should put the limit. This is closely related to the libraries of programs which are available at some computer centers. We discussed once the idea of having a library of astronomical programs, but our scientific council pointed out that the programs have to be either very simple (for the general user) or to be rather complicated (for the specialist). Since the specialist will very probably prefer to do his own program, the complicated programs are not very useful and therefore the whole idea of the library of programs collapsed. By now it seems to me that we should not mix our data base activities with too much additional work, but I would be happy if the subject be discussed at the meeting.

Let us leave now this area and shift our attention on what we have achieved so far. From the lists published in the reports of the data centers we know that there are about three hundred catalogs available on magnetic tape. It is important to recall that this is NOT the product of one center or even one country, quite to the contrary, these catalogs come from at least a dozen countries. I think that we should be proud of this international cooperation, reached through a number of bilateral agreements, which greatly facilitate the exchange of data. At the present time there are agreements relating the data centers in U.S., U.S.S.R., Japan, G.D.R. and we plan to extend this further, to arrive at the point where all astronomers from all countries can have access to ALL existing data. This was a dream, but thanks to your cooperation it is about becoming true.

Nevertheless, we should not be too self content, because first we have not yet on tape all what we need. To quote but a few examples of what is missing : the complete Durchmusterungen, the Astrographic Catalogs Luytens various surveys. Second, because not all countries have data centers, and use very little the existing facilities and thirdly, because we have not been able so far to update our catalogs and files with the necessary speed. Let me provide some examples. I start with a case familiar to me, the "The Bibliographic Star Index" : we are running behind, since we have only completed the literature to 1979, o. Well, we hope to do better in the next future, but that is something everybody hopes to do Photometry also is a rapidly moving field, and infrared catalogues are some years behind, as also u v b y photometric catalogues. In the latter case the annual growth is of the order of 20 %, counting only the number of new stars added. Similarly in spectral classification we have not yet on tape the excellent work done by our Soviet colleagues ; also here the number of stars classified is very large. In astrometry we have still nothing better than S.A.O. for positions and proper motions over the whole sky, despite the fact that the material for such an enlarged S.A.O. exists - even

on magnetic tape. And let us not forget some astrophysical parameters like magnetic fields, whose measurements were never collected in a catalog.

I think that the data centers should play a more active role in astronomy, in the sense that they should encourage people to do work which is needed. This implies contacting the best specialists in the field and ask them to do the job, be it a catalog, a selection or a bibliography. I insist strongly, and you should excuse me, on the principle that a good catalog can only be made by a good specialist. It is quite useless (but happens unfortunately) that colleagues who never used a given technique compose a catalogue using this technique. This I feel is a bad practice, since surely he will not be aware of the many pitfalls on the road, and produce something of inferior quality. So in my idea, data centers should attempt to be ahead of developments, trying to develop new catalogs and/or bibliographies. I know that this involves a great deal of sometimes frustrating work, but I think that those colleagues who work at data centers have a much better feeling for what is missing than the average astronomer. Let me also insist that data centers should try also to act as diffusers of the specialities being developed in their country. This means simply that if in the country X there is a strong development of the speciality Y, then the local data center should feel obliged to see that the results of this development reaches the mainstream of astronomical information. Curiously this simple idea needs much implementation in many countries, specially in those in which astronomy is just beginning to be developed.

If we consider next what happens in data activities in our field, I think that we have been organizing successfully the stellar area mainly. This is certainly a reasonable approach, since some problems are much less complicated for stars than for non-stellar objects. A first obvious problem is the existence of instrumentally defined nomenclatures, for instance in the case of an HII region with dimensions defined very much by the resolution of the receiver. If one changes the receiver, the picture alters in a significant way and the problem arises if one is dealing with exactly the same object. Multiple names attached to the same entity, or to different parts of the same objects occur frequently and there is no clear cut resolution to the problem. Probably all these objects need a detailed cartography together with the nomenclature and the only way to produce this is through images on microfiches. We hope to hear more on this from those colleagues involved in non-stellar data centers. There is one important point however which I hope shall not forgotten in the discussion, namely that even more than for stars we need detailed bibliographics of each object. Whereas in stars many parameters can be catalogued because they can be summarized in a compact way, this is not so in the non-stellar area. Just as an example think of the "diameter of a star" as opposed to the "diamater of a galaxy". The only reasonable solution seems to be a complete bi-

bibliography of the galaxy one is interested in, so that one can get the papers which contain that part of the information which cannot easily be summarized.

One should also think of the creation of a data bank for the solar system. I am sure that a bibliographic file on each comet or asteroid would be greatly appreciated. But is somebody doing such a thing ?

A last point I want to consider is the question of how to publish the catalogs : i.e. to what could be called "editorial policies". It is true that we are primarily interested in computer readable data, but please do not forget that first of all we are astronomers and as such working for the future generations, which implies that we have to preserve our observations. How long can a tape be preserved, if stocked properly and re-read say, once a year ? I have found different opinions but nobody said that they can be stored without alterations more than ten years, with the present technology. Now this is nothing in face of the time over which an astronomical observation is used. Probably we shall hear more on this subject in the bibliographic section, but let me anticipate that we cannot liberate us yet from the need of a permanent support, like paper, film or plastic. You can answer that scientific magazines are just there to print catalogs. But if you consider the page charges which increase and the typography which decreases in some journals with time, you feel that catalogs cannot be any more printed the traditional way, if you want to have them updated frequently. This is a "must" in certain fields, like photometry, where one gets an increase in the number of objects by 10 - 20 % per year. So if you do not want to have a large number of data "lost" for practical purposes because they are spread out over two hundred publications, you must find a "quick" solution and that is why I believe that the microfiches are to stay with us, despite the inconvenience of the need of a microfiche reader. To underline this, just consider the case of any printed catalog. The traditional way of a type-set catalog produces an inevitable delay of two years at least after the catalog is compiled, so that by the time it is distributed, it is already outdated. We need definitely to remedy this and I hope we shall hear more on this topic in the session on "Editorial policies".

I shall stop here ; I have listed many open questions and I hope that this meeting shall help to clarify them or to produce answers.