AIDS TO THE RETRIEVAL AND EVALUATION OF ASTRONOMICAL DATA

G. A. Wilkins
Royal Greenwich Observatory
Herstmonceux Castle, Sussex, England

ABSTRACT

Authors, editors and data centres should ensure that data on astronomical objects and systems can be readily located and that their accuracy can be properly judged by the user. In particular, authors and editors should endeavour to ensure that the title, abstract and other retrieval aids are accurate and informative and that the procedures used are described in sufficient detail to allow a proper evaluation of the results. Data centres should, where appropriate, provide additional aids to allow the users to find data efficiently and should, in all cases, ensure that each datafile contains adequate information about the sources of the data and about the significance of the fields in the records of the file.

INTRODUCTION

The main purpose of this paper is to draw attention to some of the ways by which authors, editors and data centres may significantly increase the probability that published data will be found and used by those who require such data. The paper also contains some information that may be helpful to those who wish to retrieve astronomical data. The adoption of the practices recommended here would lead to reductions in the amount of unnecessary observing and computing that now takes place because the astronomer has either failed to find already published data or has not been able to judge whether the data are of adequate quality for his purposes. Their adoption would lead to an increase in the amount of effective use of the data and a decrease in the amount of unnecessary publication and storage. More importantly, their adoption could lead to more rapid advances in our knowledge and understanding of the universe.

For simplicity of presentation we suppose that an astronomer wishes to find data about stars of a particular type, but similar considerations apply to searches for information about other types of astronomical object or system or technique. We are here concerned only with data

C. Jaschek and W. Heintz (eds.), Automated Data Retrieval in Astronomy, 193–198.
Copyright © 1982 by D. Reidel Publishing Company.

https://doi.org/10.1017/S0252921100082920 Published online by Cambridge University Press

that are published in printed form and/or on magnetic tape. We recognise, however, that the published information may only refer to a source of further data about the stars, and that these data may be in non-numerical form, eg a photographic plate or as traces from a chart-recorder. The same principles apply: the information given should be sufficient to allow the astronomer to obtain access to the data and to judge whether they will be suitable and of sufficiently high quality to justify the effort involved in obtaining them.

AIDS TO RETRIEVAL

There are two phases in the retrieval of data, namely "search" and "fetch": firstly, it is necessary to establish where the data may be found and, secondly, it is necessary to obtain the data from that place, which may be in a journal, or on a magnetic tape, or in the store of a computer. In a fully automated system the fetch-phase will be a trivial operation once the search-phase has been successfully completed, but this is not yet the usual situation.

The most commonly used search technique is probably that in which the astronomer starts from his own knowledge of relevant material and then expands his search by looking up appropriate references given by each paper. This technique can involve much wasted time and expense unless the following conditions are satisfied: (a) each reference should include the title of the paper, a complete and unambiguous statement of the title of the journal, and the limiting page numbers, not just the number of the first page; and (b) the title of the paper should give a clear indication of its contents. The astronomer making the search can use clues given by the original author, but these may not be sufficient to avoid a wasted fetch when the reference is incomplete or the title is misleading. Review papers and the triennial reports on the progress of astronomy published in the Transactions of the IAU are very useful starting-points for such searches.

The second commonly-used search technique involves the use of abstracting journals and other published indexes, such as the Science Citation Index. Such searches may be carried out manually or by computer techniques since the contents of many abstracting journals are stored in computer databases. Astronomy and Astrophysics Abstracts provides the most useful single source of information about the current literature of astronomy, but it is not yet available in a database. Other English-language series containing abstracts of astronomical interest include Physics Abstracts and Scientific and Technical Aerospace Reports (STAR); these are accessible by computer techniques. Astronomical Abstracts are published in Russian in Referativnyi Zhurnal, part 51, and in French in Bulletin Signaletique, part 120.

The effectiveness of this technique depends on the quality of the information given in the entries as well as on the facilities provided

https://doi.org/10.1017/S0252921100082920 Published online by Cambridge University Press

in the abstracting journal or by the computer program. In particular the author should provide information in the abstract that will show the nature of the contents of the paper more clearly than the title is able to do; in particular he should indicate whether or not the paper contains observational data and results. In some disciplines, but not in astronomy, there are formal procedures for the "flagging" and "tagging" of data to indicate their presence and facilitate their retrieval (CODATA, 1976). The principal results or conclusions should be given in the abstract wherever this is practicable.

The arrangement of the abstracts in the journal and the supplementary indexing facilities that are provided also affect the effectiveness of searches. Unfortunately there are several different classification systems in use in astronomy. Since astronomers need to use information from other disciplines it appears to be extremely desirable that the Universal Decimal Classification should be adopted as the standard classification scheme for astronomy. Revised schedules for astronomy in class UDC 52 were prepared by an IAU/FID Joint Working Group quite recently (BSI, 1977). The UDC system usually allows the contents and format of a paper to be specified clearly and compactly and is quite suitable for use in information-retrieval systems. A guide to use of UDC in astronomy is in preparation at the Royal Greenwich Observatory.

At present many information-retrieval systems make use of keywords to describe the contents of a paper. A list of astronomical terms that are useful for this purpose has been prepared by the editors of Astronomy and Astrophysics Abstracts (Schmadel, 1979), and an extended list is being prepared by a Working Group of IAU Commission 5 (Lantos, 1981); this list should be consulted by editors and others responsible for the assignment of keywords. A multi-lingual version of this list would be useful. Some systems do not use assigned classification numbers or keywords but rely on scanning the title, and possibly the abstract, for words chosen by the person making the search. This should be borne in mind by authors when choosing titles for their papers.

Neither of the main techniques so far described is suited to the common astronomical requirement of finding data about, say, a set of stars satisfying specific selection criteria since, for example, all the relevant information cannot be given in the reference or abstracting journals. Instead, it is necessary to search the numerical datafiles themselves, and not merely bibliographic datafiles. Such searches are now usually carried out by computer techniques since many printed astronomical catalogues have been transcribed to cards or magnetic tape and most new catalogues are prepared by using computers. Datafiles should be made available in both printed and computer-readable form; extensive datafiles may be issued conveniently and economically on microfiche; magnetic tapes should be made available in standard formats.

https://doi.org/10.1017/S0252921100082920 Published online by Cambridge University Press

The Stellar Data Centre at Strasbourg and its associated institutions have built up an extensive collection of astronomical datafiles and provide associated services. Information about these datafiles and services is distributed regularly in the Bulletin d'Information du Centre de Donees Stellaires (Observatoire de Strasbourg). Two of the datafiles deserve special mention because of their great value in searches for data: they are the Catalogue of Stellar Identifications, which provides cross-references between the names used for individual stars in different catalogues, and the Bibliographical Star Index, which aims to provide bibliographic references to papers giving information about individual stars. Even so the variety of astronomical objects and the bewildering number of different ways in which they are identified and by which their properties are described means that astronomers will continue to need patience, skill and a good knowledge of the literature of astronomy if their searches for published data are to be effective and efficient. Authors should follow the recommendations of the International Astronomical Union (1980) and ensure that the astronomical objects to which their data refer are unambiguously identified; they should seek the advice of the Union before introducing new styles of designation for new or unusual types of object.

A directory of sources of astronomical data has been compiled by Jaschek (1977) and republished by CODATA (1980) with amendments and additions.

Much useful astronomical data is not published in journals but in special series of reports issued by the observatories themselves. It is important that such reports be sent to the Astronomisches Rechen-Institut in Heidelberg so that they are included in Astronomy and Astrophysics Abstracts. A union list of the holdings of the principal astronomical libraries in the United Kingdom is in the early stages of preparation so as to facilitate the fetch phase of the retrieval of older material.

The original plates or other records invariably contain more information than the published reports of the observations. The retrieval of valuable information from such sources will be greatly facilitated by the compilation by IAU Commission 5 (Hauck, 1980) of a list of the plates, etc, that observatories hold and are prepared to make available for re-examination and re-measurement.

AIDS TO EVALUATION

The adoption of the following principles would make possible the proper evaluation of published data obtained from astronomical observations.

1. The paper must contain an adequate description of the observational procedures used to obtain the numerical data.

https://doi.org/10.1017/S0252921100082920 Published online by Cambridge University Press

2. The paper must contain an adequate description of the procedure used to derive the reported results from the actual measurements.

3. The paper must contain full numerical results in a form that is as free from interpretation as possible and in such a manner that (a) the uncertainties of the data can be independently assessed and (b) the observational data can be re-analysed in terms of a hypothesis that is different from that considered by the author.

4. The data must be presented in such a way that the objects, system or phenomena observed and the quantities tabulated can be unambiguously identified and so that the results can be readily related to other data for the same or similar systems.

These principles have been developed from the guides prepared by CODATA (1973 & 1979) on the presentation of data derived from experiments and from observations in the geosciences. A full explanation of these principles, with specific examples of their application in astronomy, has been prepared (Wilkins, 1981) for submission to the IAU for approval prior to its publication by CODATA.

Similar principles apply to the publication of data derived by combining the results of different series of observations. The value of such data is very much diminished if the author does not provide adequate information about the sources of the original data and about the methods used to combine the data. Results based on computations for theoretical models must also be accompanied by sufficient information about the formulation of the model and the numerical values of the parameters to allow another person to reconstruct the computation and verify the published data.

Data centres should also ensure that the sources of the data that they store and distribute are specified with the data, preferably in a header block with information about the significance of the fields in the records of the file.

CONCLUSIONS

The advice given in this paper is not original but it is hoped that its presentation in this form will be found to be helpful to authors, editors, and the users of the astronomical literature. The costs of obtaining and analysing astronomical data are often very great and so it is clearly desirable that the published data should be presented in such a way that they can be readily found and fully utilised by those who are interested in them. The sourcebook prepared by Rossmassler & Watson (1980) contains several articles that are relevant to the retrieval and evaluation of data in the wider context of science and technology.

https://doi.org/10.1017/S0252921100082920 Published online by Cambridge University Press

REFERENCES

BSI, 1977. Universal Decimal Classification: UDC 52 Astronomy Astrophysics Space Research Geodesy. ii + 43pp. BS 1000 [52]: 1977. British Standards Institution, London W1A 2BS.

CODATA, 1973. Guide for the presentation in the primary literature of numerical data derived from experiments. CODATA Bulletin No. 9, 6pp.

CODATA, 1976. Flagging and tagging data to indicate its presence and facilitate its retrieval. CODATA Bulletin No. 19, ii + 22pp.

CODATA, 1979. Guide for the presentation in the primary literature of numerical data derived from observations in the geosciences. CODATA Bulletin No. 32, i + 6pp.

CODATA, 1980. CODATA Directory of data sources for science and technology. Chapter 3. Astronomy. CODATA Bulletin No. 36, v + 56pp.

CODATA publications are available from CODATA Secretariat, 51 Boulevard de Montmorency, 75016 Paris, France.

Hauck, B., 1981. Preliminary results of an enquiry for better access to plate vaults. Bull. Inform. Centre Donnees Stellaires No. 20, 68-71.

IAU, 1980. Resolutions concerning the identification of astronomical objects in publications and the abbreviations of catalogues are given in Trans. IAU 17B, 42-43.

Jaschek, C., 1977. Survey of existing facilities. See Jaschek & Wilkins, 1977, 205-279.

Jaschek, C., and G. A. Wilkins (eds), 1977. Compilation, Critical Evaluation and Distribution of Stellar Data. D. Reidel Publishing Company, Dordrecht-Holland.

Lantos, P., 1981. Draft IAU vocabulary. 18pp. Observatoire de Paris, 92190 Meudon, France.

Rossmassler, S. A., and D. G. Watson (eds), 1980. Data handling for science and technology: an overview and sourcebook. xvi + 184pp. ISBN 0-444-86012-6. North-Holland Publishing Company, Amsterdam.

Schmadel, L. D., I. Heinrich and G. Zech, 1979. Astronomy and Astrophysics Abstracts: Vocabulary 06.79, vii + 152pp. Astronomisches Rechen Institut, Heidelberg.

Wilkins, G. A. 1981. Guide to the presentation of astronomical data. Preprint, 9pp, available from Royal Greenwich Observatory.

https://doi.org/10.1017/S0252921100082920 Published online by Cambridge University Press