

Searching for the Data

F.D. Gault, Durham

(Particle Data Group, The University of Durham)

In May of last year, *Europhysics News* reviewed the state of bibliographic data bases and their availability through networks like Euronet Diane inaugurated in February, 1980. In this article we consider some numeric data bases in research physics and the same questions of availability, cost and future impact. By a numeric data base is meant a collection of data, available in machine readable form, but not necessarily available through an information retrieval system. We also consider the differences between bibliographic and numeric data bases and the attitudes of the people who use them.

The principal difference between bibliographic and numeric data compilation in the pure research environment is economic. Books and journals are commercial products and it is a simple step from the abstract journal to the bibliographic data base searchable by computer and available, at a price, on a commercial network. A person in a company or laboratory with limited library facilities can recognise the benefits of such a service (see also N.B. Butterwick, *EN*, **11** (1980) 9, p. 8) and is willing to pay to use it. While the research physicist also recognises the value of numeric data compilations his attitude towards them is not the same.

Measurements in atomic and molecular, nuclear and particle physics are not made as a commercial enterprise and the cost is born by the public. The new data, after statistical and systematic error analysis, are preprinted and later published in a learned journal. The paper might then be listed in an abstract journal or entered into a bibliographic data base but our concern here, is what happens to the actual data. There are various possibilities which are discussed below.

Journals encourage the publication of large data sets in graphical form and it is not unknown in particle physics for the actual numbers to be lost after the disbandment of the experimental group which produced them. Some journals offer to store and make available copies of large data sets to prevent this, but they cannot provide a comprehensive data archive, as not all data are published in the same journal. In fact not all data are published, some appear in theses and some remain for years as a preprint or a preliminary report given at a conference. The purpose of the physics data base is to archive these data, as soon as they are released by the experimentalists, to preserve them for future generations of scholars.

Archiving research data is not done by commercial organisations because of the high cost and the limited market among physicists who, anyway, tend to feel that data should cost them nothing. It has been done for some years non-commercially in the areas of atomic and molecular data in support of fusion research, and in nuclear physics in support of reactor design or research into nuclear structure. Crystallographic data have also long been compiled, evaluated and, in some cases, commercially marketed. However, in each of the examples cited, there is a common element of support for other research which requires evaluated data as a starting point. The critical evaluation which produces a best value and the related uncertainty is the product which interests the user; the archiving role of the compilation work is seen to be secondary. These data are available in Europe but not, in all cases, through computer searchable data bases, accessible through networks like the bibliographic data bases mentioned earlier.

Atomic, Molecular and Nuclear Data

The IAEA in Vienna, through its Nuclear Data Section, maintains the Atomic and Molecular Data Unit which coordinates for the fusion research community, the compilation of computerized, evaluated, atomic and molecular data of adequate and uniform quality, and acceptable accuracy. A bibliography is also maintained in machine readable form but the intention is to publish it from time to time rather than to make it available as a computer searchable data base. The evaluated data are to be offered to the users on magnetic tape.

Magnetic tape is also the main method of disseminating the Evaluated Nuclear Structure and Decay File (ENSDF), an evaluated data compilation coordinated by the Nuclear Data Project at Oak Ridge National Laboratory in the USA, and in the process of being transferred to the National Nuclear Data Center (NNDC) at Brookhaven National Laboratory. The tape, which is the end product of an evaluation network, is issued twice yearly by the NNDC and is available in Europe along with its associated bibliography, the Nuclear Structure Data References, at the Daresbury Laboratory in the UK. In addition to this, the OECD Nuclear Energy Agency in Paris makes the ENSDF information available on the French CISI network, while the Fachinformationszentrum, Karlsruhe is in the process of making it available directly and through Euronet Diane.

The question of how best to access data is resolved differently in different groups and sometimes differently within one group. The Nuclear Energy Agency Data Base, for example, offers the ENSDF on a network, but its neutron data, which include the Evaluated Neutron Data File (ENDF/B) co-ordinated by the NNDC, are

PARTICIPANTS IN THE ISOLDE COLLABORATION

University of Aarhus, Institute of Physics, Aarhus
Hahn Meitner Institute, Bereich Physik, Berlin
Niels Bohr Institute, Copenhagen
Gesellschaft für Schwerionenforschung (GSI),
Darmstadt
Universität München, Sektion Physik, Garching
Chalmers University of Technology, Department of
Physics, Göteborg
Kernforschungsanlage Jülich, IKP II, Jülich
Johannes Gutenberg Universität, Institut für Kern-
chemie, Mainz
Johannes Gutenberg Universität, Institut für Physik,
Mainz
Universität Marburg, Institut für Physik, Marburg
Swedish Research Council's Laboratory, Nyköping
Institut de Physique Nucléaire, Univ. de Paris XI, Orsay

Laboratoire René Bernas, CSNSM, Orsay
Laboratoire Aimé Cotton, CNRS 2, Orsay
Institut des Sciences Nucléaires, Univ. de Grenoble,
Grenoble
Institut de Physique Nucléaire, Villeurbanne
Centre de Recherches Nucléaires, Strasbourg
Institut de Physique Nucléaire et de Physique des Parti-
cules (IN2P3), Paris
University of Oslo, Department of Chemistry, Oslo
Niels Bohr Institute, Tandem Accelerator, Risø,
Roskilde
Research Institute for Physics, Stockholm
University of Stockholm, Institute of Physics,
Stockholm
University of Uppsala, Department of Physics, Uppsala
CERN, Geneva

REFERENCES

1. Hansen, P.G. *Annual Review of Nuclear and Particle Science* **29** (1979) 69-119.
2. Jacquinot P. and Klapisch R., *Rep. Prog. Phys.* **42** (1979) 773-832.
3. Ravn H.L., *Physics Rep.* **54** (1979) 201-259.
4. Jonson B., also Otten E.W. in *Proceedings Int. Conf. on Nuclear Physics, Berkeley 1980* (North Holland, 1981) Vol. II.
5. *Proceedings of the 3rd Int. Conf. on Nuclei Far From Stability (Cargèse, Corsica (France) May 1976)* (CERN 76-13, 1976) pp. 1-608.
6. "Future Directions in Studies of Nuclei Far From Stability", *Proc. Int. Symp. Nashville, U.S.A., Sept. 1979* (eds. J.H. Hamilton, E.H. Spejewski, C.R. Bingham and E.F. Zganjar) (North Holland, Amsterdam, 1980) 1-424.

accessed through an expert intermediary. One physicist is employed full time to process requests for data — a sensible way of proceeding when the number, volume and complexity of the available data files are considered.

Whether or not an expert intermediary is used, the method of retrieval and presentation of the retrieved data can have an important impact on the use of the data. This is illustrated by the Cambridge Crystallographic Data Files (CCDF). The data can be retrieved in batch mode, or interactively, and displayed graphically. Because of the complexity of the information, graphical display and manipulation facilities are essential to the use of the CCDF in the study of chemical structure. It is a good example for a data base where the software is at least as important as the original data and evaluation, if the information compiled is to be used easily from a terminal. The CCDF originate from the Cambridge Crystallographic Data Centre and are made available on the UK Science Research Council network by the Daresbury Laboratory. The Fachinformationszentrum, Karlsruhe is also planning to offer the CCDF.

Data on Particles

A contrasting approach to that of the compilation and evaluation groups, is taken by the Particle Data Group in the UK. The PDG(UK) is committed to compiling all elementary particle scattering data and associated bibliographic information, and to making the resulting compilations easily available through computer searchable data bases. The work is complementary to that of the PDG at Lawrence Berkeley Laboratory which produces, every two years, the Review of Particle Properties.

The data bases are held on the Rutherford and Appleton Laboratory computers, and are available throughout the UK and at CERN and DESY through the SRC network. The aim of these data bases is not to present an evaluation, but to make available the data in machine readable form soon after publication (or preprinting) and to support them with graphics and manipulation facilities, to enable data to be plotted or compare in sets by overlaying them on a single plot.

The data are used in particle physics research rather than in support of a related subject, and for this reason, the user is assumed to be sufficiently expert in his understanding of the data, to address sensible questions to the retrieval package.

No expert intermediary is necessary, although expert advice and help is available from the compilers for anyone not wishing to deal with a computer terminal directly, or for people who want data not in the data bases.

The object of the PDG(UK) in the long term is to make the data bases appear as an integral part of the computing support of-

Computer Programs in Physics and Chemistry

Over 500 refereed programs in the fields of

astrophysics, atomic physics, chemistry, crystallography, electrostatics, fluid and gas dynamics, geophysics, laser physics, molecular physics, nuclear physics, optics, plasma physics, quantum chemistry, radiation physics, radiative transfer, solid state physics.

are held in the CPC Program Library at Queen's University, Belfast, with published writeups and tests in the international journal Computer Physics Communications.

Either individual programs or the complete program library may be obtained on request at charges which cover handling costs: currently £15 plus £7 per 1000 card images for an individual program or £150 per volume (approximately 46000 card images) plus the costs of tape and postage for a library subscription. Full details and costs are published in the journal or available from:

Miss C. Jackson, Department of Applied Mathematics
The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland

ferred by the UK SRC to particle physicists. In principle, the data bases should be seen by the user as a system utility like a compiler or a sort program, and by the experimentalist who supplies the data, as a marketing device. This is possible only because the SRC has taken the decision to support compilation in order to encourage the use of the data. The cost, when considered as a fraction of the budget for experiment, is quite low, but the service would not be commercially viable if the salaries of the compilers had to be recovered from a user community that is measured in the hundreds even if extended to all of Europe.

Overall Prospects

The FRG also encourages data compilation through the Fachinformationszentrum, Energie, Physik, Mathematik at Karlsruhe which, to complement its work in published data compilations and bibliographic data bases, is moving towards computer searchable numeric data bases that are publicly available. A broad spectrum of information is stored and it will be interesting to see to what extent the community of research physicists makes use of these facilities, in view of the connexion and computing charges. The charges do not reflect the real cost of compiling the information, so there is still a considerable government subsidy involved.

Physics data bases, from the examples considered, would seem to be becoming established as research tools. There is far more support for compilation, evaluation and dissemination of computer readable information than there was five years ago, and that support is growing. To some extent, this reflects the trend in computing towards data base technology and networks, but there are two other important motivations evident: to support particular projects like fusion research, and to make better or more use of the original data, as particle physics. There should be a third compelling motivation, which is the con-

servation of the original data, but one could argue that that is embedded in the other two. If physics data bases are now part of the scene one can ask what influence they are likely to have outside their role as information sources.

Data bases because they are supported by expert physicist compilers, provide a channel of dialogue between the user and the supplier of experimental information. The professional compiler is better equipped to find data which may not be published, to comment on methods of data analysis, or to indicate which of many authors of a particular paper actually did the analysis so that the user can contact him directly. The interaction between user and supplier and the comparison facilities provided by a data base, are bound to lead in time to more uniform standards of data reporting.

Database Guides

Sir,

Further to my letter in *Europhysics News* (Sept. 1980), I have received details of updates of the database guides that are mentioned there.

The details of Tomberg's guide are as follows:

Tomberg, Alex (ed.)
EUSIDIC Database Guide 1980
(Learned Information, Oxford:
Dec. 1979) ISBN 0 904933 20 2
Price: £16, paper cover
and of Hall's:
Hall, James L. and Brown, Marjorie J.
Online Bibliographic Databases:
an International Directory, 2nd. ed.
(Aslib, 1981) ISBN 0 85142 138 5
Price: £20; Aslib mbrs. £17.

N.B. Butterwick
Physics and Mathematical Sciences Librarian
Queen's University, Belfast