Searching the Literature

An enquiry by Europhysics News

Basic to the method of scientific development — whether this concerns fundamental research or commercial application — is the search of the published literature to discover what is already known, or at least believed. For most physicists, this is a continuing process of scanning certain of the journals covering their own field, as they are published, studying in detail a few of the papers and references cited, and then from time to time making sorties into unfamiliar territory. For the designers of equipment, a further operation should be added, namely consulting sources relating to the specification of components that may be built into an assembly. It is in this last context that the word data seems most aptly to describe the information sought, although in the jargon, the word is now used to cover any information about a publication.

Against this background of user practice, the publishers of scientific literature have organized their production and we have seen the rather general publications of former times supplemented by more and more specialized journals, letter journals, and various forms of reference documents listing new publications, with or without abstracts. At the same time, review journals have become popular and services have been introduced to draw the attention of subscribers to new material appearing in defined subjects. Publishers of the Trade and Technical Press have reacted similarly, concentrating on «news» and the characteristics of new devices and offering a service designed to put potential customers in touch with manufacturers.

For the individual, the wealth and the cost of all these information sources are overwhelming — but then they always were. The scientist has always been dependent upon a wide spectrum of information sources, relying on his institutional library to supplement his direct contacts. Within most university physics departments, the rôle of the library has been in its essentials unchanged for a hundred years and more, while modern research laboratories have followed the tradition the Universities established. Periodicals come in and are displayed and/or circulated, books are acquired and are duly catalogued and put on shelves, references are hunted and copies of publications not in stock are obtained

from other libraries. Reference books are as up to date as the budget permits, indexes as up to date as the librarian is efficient. A familiar scene, which evokes the question: does it have to change?

The problem is one of scale. Can systems that were established years ago, cope with the exponential rise in information production for very much longer? Many of the bigger libraries have already introduced computer cataloguing with refined indexing systems that permit interrogation by relatively unskilled users working in everyday language. Many companies have their own current awareness services designed to lighten the reading load of the staff. In both cases a considerable in-house effort is involved, and the range of subjects covered must still be restricted. If the in-house activity can be narrowly defined, this may be no disadvantage provided, however, the ability to browse is not eliminated. The literature should offer inspiration every bit as much as information and for this, the user must be able to pick up and put down, leaf through a document, skim through the titles, linger over something unexpected, all of which is a very personal activity, difficult to delegate to someone else. It also demands a certain atmosphere — or is this just old-fashioned thinking?

Before waxing sentimental about the warmth and scholastic pleasures of the traditional library, thought has to be given to the basic problem of keeping the stored information relevant. Unused and unusable literature does not make for good science, and the modestly funded library required to serve a clientele of wide-ranging interests, has no possibility of supplying a comprehensive service without recourse to external help. The most economic and most convenient way of obtaining such help may well lie in establishing a connexion to one or more of the «data base hosts» that have made their appearance over the past few years. Choosing the best is not easy, but the following comments may provide some useful background.

Euronet Diane

Towards the end of 1979, a new information network, Euronet Diane, came into operation in western Europe and was formally inaugurated on 13 February of this

year. Following agreements between the Post Office authorities in the countries of the European Communities, an international data transmission network has been set up which links the nine member countries, while an extension to Switzerland should be ready by Autumn, this year. This is Euronet. Diane is an acronym for Direct Information Access Network for Europe. It is designed to make accessible online, and in an interactive mode, individual users with any of a series of «host» institutions which maintain in computerized memory banks, «data bases» consisting of files of information on published literature.

The Communities' part in all this has been to promote the idea and create the conditions for arriving at a standardized and cheap communication system. From then on, national and commercial interests take over. There is in consequence the same competition between producers of files as between publishers, the same competition between hosts as between booksellers and evidently competition between the manufacturers of equipment needed by the user. This competition should be a guarantee of quality, but it has the negative effect that services offered may overlap, may not be coordinated in the way the user would prefer, and activities that seem marginally profitable may be neglected. Moreover until the advent of Euronet Diane, each host had its own language and access systems so that multiple connexions where effectively excluded.

In many countries, specialized data banks had been set up as part of a national library service, helping to speed up the literature searches that where undertaken as part of the library's activities. Now if an institution is prepared to invest in a terminal and «subscribe» to one or more hosts, the individual physicist is given access to a vast index system. (Local rules may require him to make contact through a professional communicator, i.e. librarian, in which case, he still has the problem of explaining precisely the nature of the information that he is seeking.) Through his terminal, the communicator can at any time interrogate the files of the host(s) through a type writer that may or may not have an associated visual display. Having selected the information required, this is then printed out directly or very soon after. What he receives is evidently no more than is in the base. It may be only a simple reference to a published article, or may include an abstract containing much or little information that is of direct use, or it may define the specification of a particular product. Obtaining copies of the papers to which references relate is a second process that follows traditional channels.

Where then is the gain over the abstract journal? In a narrow, well-covered field, the advantage probably lies in the convenience of having a source that is regularly up-

dated and automatically re-ordered; it may also be cheaper, cost being a function of the frequency of interrogation. In unfamiliar fields, the advantage lies in the immediate availability of a wide source of information and the ability to interact with the source so that from the mass of data on call, a short-list can be prepared of the references that really should be of interest. Moreover, if the operator is skilled, his expenditure can be limited to little more than the bare cost of calling up the information he needs. According to a survey carried out by the EEC, the main factors influencing librarians to use on-line services are: speed of response (51%), cheaper than manual searches (34%), efficiency and accessibility of on-line services (21%), wider range of data available on line (16%), more work can be done by fewer staff (15%), easier to do complex searches (12%).

One of the earliest remote access systems was introduced by NASA to make available to its myriad suppliers, data on approved components that could be built into assemblies that it had ordered. The idea was subsequently taken up in Europe with space and nuclear energy leading the way and it has now become an accepted technique throughout the arts and sciences, bibliographic information predominating.

The number of data bases and their size grow continuously. By Autumn 1980, Euronet Diane expects there to be 23 hosts working with 175 data bases on its network, covering subjects ranging from «Register of Italian barristers and solicitors» through «Astronomy and physics» to «Sport and sport sciences»! Each base is known by an acronym — AB-INFORM, AFEE, AGREP, AGRIS...names which mean

nothing to the uninitiated and even the full titles give little indication of the detail stored. Of the total, 17 bases seem to be directly related to physics and three others marginally so (see panel). They are accessible through only some of the hosts that have been established in the Community countries as these generally concentrate on specific fields.

Other Networks Serving Europe

Euronet Diane is not however, the only network that is operating in Europe and for some time past, a number of organizations have been offering remote access facilities to individual users through terminals linked to the telephone system. At the European level, the Information Retrieval Service (IRS) of ESA operates an on-line interactive information retrieval system — QUEST — enabling users all over Europe to access within a few seconds some 18 bibliographic data bases and four factual data banks, stored in the computer centre at Frascati, Italy. Searchable data pertain to over 14 million references of scientific and technical documents (reports, journal articles, conference proceedings, etc). The references contain full bibliographic data as well as abstracts in most files.

The ESA service is currently available both through Euronet and its own network, ESANET (which is liable to become redundant) and is linked with networks in the U.S.A.

Very active in the field have been a number of American companies making use of satellite links to Europe at impressively economic rates, though now more expensive than Euronet Diane.

Hosts of Euronet Diane

Belgium
*CTI, Brussels	N

Denmark
Datacentralen, Copenhagen	

France
ARCDIC-CIDA, Paris	N
CATED, Paris	N
CISI, Paris	N
ERGODATA, Paris	
Institut Textile de France, Paris	N
*SPIDEL, Paris	N
*Télésystèmes-Questel, Paris	N
*Thermodata, Grenoble	N

Federal Republic of Germany
DIMDI, Cologne	N
*FIZ-Technik, Frankfurt	N
GID, Frankfurt	N
*INKA, Karlsruhe	N

Italy
CED, Rome	
CERVED, Padua	N
CILEA, Milan	
*IRC, Ispra	
*IRS, Frascati	N

Luxembourg
ECHO Service, Luxembourg	

Netherlands
European Patents Office, The Hague	

U.K.
BLAISE, London	N
*INFOLINE, London	N

N = Now available
* = Files on physics included in bases

Notable amongst these are Bibliographic Retrieval Services, Lockheed, and System Development Corp. The advantages stemming from a large integrated home market are, as in publishing generally, at once apparent. The U.S.A. is served by relatively few massive systems, while Europe is characterised by many local systems each with its own methods. Euronet goes some way to encouraging integration as five hosts have now implemented a common language developed by the EEC.

In eastern Europe, abstracts of the literature are not yet available on line, data bases being restricted for the present to lists of institutional activities. The literature in eastern European languages is nevertheless abstracted in some western European bases, as in the abstract journals, the relation with which is very close.

Costs

For the producers of data bases the present position is a little curious, as in many cases the operation is both a by-product and a competitor of traditional publication operations, Faced with a choice of updating a base or selling an abstracts journal

Data Bases for Physics

ENSDF	Evaluated Nuclear Structures Data File, including decay data for all isotopes.
EPIC	Programmes for calculation of physical properties of chemical compounds.
INIS	International Nuclear Information System.
INKA-ASTRO	Astronomy and astrophysics.
-CONF	Conference announcements in energy, nuclear science, aeronautics, astronautics, space research, physics, mathematics and astronomy.
-DATACOMP	Data compilations in energy and physics.
-HEP	High energy physics.
-MATH	Mathematics and related subjects.
-NUCLEAR	Nuclear science.
-PHYS	Physics and related fields.
-PLASMA	Plasma physics and technology.
-SPACE	Conference papers on aeronautics, astronautics, and space research.
-SURVAC	Surface and vacuum physics.
INSPEC	Physics, electronics, computing and mathematics.
PASCAL	Science and technology.
SPIN	Searchable physics information notices: solid state physics.
THERMODATA	Thermodynamic values of elements, components and alloys in minerals.
CISI-ELECNUC	Characteristics of nuclear power stations.
EDB	Energy information.
RBUPC	Register of research in British institutions of higher education.

or a specially tailored current awareness service to a customer, the producer of the basic information (typically reference, classification under various key words and independently written abstracts), would no doubt opt for either of the last two. He feels compelled for insurance reasons to enter the data base field where his return is less certain and is much slower to come in, and where comprehensiveness is financially unrewarding. It is estimated that on-line bibliographic data bases account for less than 10% of the publishers' revenue.

For the user, on the other hand, one of the most attractive aspects of the interactive data base access system is its low specific cost. Once the terminal expense has been paid — typically 6000 SwFr or an equivalent annual charge — he has a nominal fee for signing on with a host — say, 120 SwFr (that may include training experience), the call charge for each communication (that is payed to the Post Office) which is about 12—15 SwFr/h from any country to any host, and then the charge for the information extracted (payable to the host) of about 150—200 SwFr/h connexion time, plus a minimal charge for printing out references offline. One could assume that the average search (i.e. accessing one data base for the answer to one question) takes about 5—10 minutes, and would thus cost about 40—55 SwFr total. Learning the communication language is not difficult and modest practice should see the necessary skills developed in a short time.

Adequate safety procedures are built into the system to ensure that unauthorized use of the terminal is prevented, that only accepted institutions are able to open a dialogue, and that the cost of the dialogue is then charged to that institution and no other. The compiler of the file is recompensed by the host on the basis of the time during which the file is interrogated and according to the number of references printed out.

Choice

In member countries of the Communities it would seem natural for new subscribers to think first of a host in the home country or an international host if their business is directly related to a specific field such as space. If the connexion has to serve physicists only, the lists of hosts and bases shown, provides at least a starting point. If other disciplines have to be covered however, the problem of choice is much more difficult.

Regretfully *Europhysics News* knows of no single document which summarizes in any comprehensive way the services offered by the various hosts, and which include the data bases that are available and the sources from which these have been compiled. But then when the whole of science, law, medicine, politics,

industry...is involved, the compilation of such a guide is not to simple. Euronet hopes to have a reference manual available in print and on line in the Autumn but the magnitude of the task can be gauged from the fact that a full description of four data bases only, occupied 100 printed pages.

International data bases covering a wide variety of literature fields is certainly an expanding market for the present. Is it though a long term solution to the information retrieval problem? This is less clear. For institutions concerned with a wide range of subjects, remote access to big centralized data bases seems to be the only practical

solution to the ever growing problem of information output. In due time also the full text of the literature may become available on the computer, but not just yet. For those institutions with narrower interests, data banks and calculation facilities can give a further incentive to making the connexion. On the other hand, new forms of abstracts publishing such as video-disc may prove competitive eventually. In any case, the problem of browsing will still remain and for this reason alone, it would seem that the new technologies must serve as an adjunct to the traditional library and not a substitute for it.