INIS - THE APPROACH AND THE PLAN

Plans for an International Nuclear Information System (INIS) to simplify the task of scientists and engineers trying to cope with the "information explosion" were reported at the Twelfth Session of the General Conference. Below is a shortened version of a special lecture on the subject given by John E. Woolston, Director of the IAEA Division of Scientific and Technical Information.

I wish that I were speaking about something that has already been achieved – some great reactor that has been built and is now operating –or some scientific discovery whose truth is obvious from its very simplicity. My talk concerns something which is rather elusive.

For information is an elusive commodity, and, in a rigorous sense, it cannot be defined except in the context of man, his habits and his institutions. In any complete study, we have to judge both the value of each piece of information, and also what chance it has to come to the notice of those who might profit from it.

Today, however, I will use the term "information" as though it were synonymous with "recorded knowledge". A piece of information may be the record of an observation, or of a fact about the physical universe, or it may be the expression of a theory about why some part of the universe behaves the way it does.

The only information that I can store or process is information that is "written", usually on paper, but perhaps on film or magnetic tape. I therefore also speak of an "article", which may be a film or a piece of magnetic tape, but is more probably a paper in a scientific journal, the text of a talk given at a meeting, a patent, a thesis, or a mimeographed technical report.

In all of science and technology – and in that I include medicine and the industrial arts – about 2 million such articles are being published per year. This is why we often hear talk of an "information explosion". People say, quite truthfully, that no scientist now has the time to read all the articles that may be relevant to his work – and if he tries to do so, he will have no time for any other work.

However, the sheer volume is not the whole story. Most of the information a scientist needs is very new information; the work that interests him is the work of his contemporaries. As many of you know, Derek de Solla Price[1]has pointed out some remarkable differences between the population statistics for scientists and the population statistics for mankind as a whole. Thus, if you consider all the people that have lived in the world during the last two thousand years, the majority of them are now dead; on the other hand, if you consider all the scientists that have lived in the world in all of history, then close to 90 per cent of them are now alive. A scientist probably has an active life of about 45 years; it is fairly easy to show from compound-interest tables that. of all the scientific information available to a man at the end of this 45 years, 80 to 90 per cent of it was generated during his working lifetime. Thus it is that the scientist unlike the historian, cannot wait. He needs to find the most rapid means of getting access to the results of the work of his contemporaries, that is, to the articles they are publishing.

TRADITIONAL SOLUTIONS

In days gone by the quantity of articles appearing each year was not so large, but the scientists were not so specialized, and each was interested in a bigger fraction of what was published. In fact, the "information explosion" has long been with us. More than 300 years ago we already had people looking for solutions, and so we had the founding of the first great scientific journals, such as that of the PHILISOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY in London. These journals brought some order into the situation and, while they did not meet all needs, they permitted a librarian to find and catalogue man's recorded scientific knowledge, and to pack it into a relatively small space.

Then the journals began to publish indexes, and the scientists continued a very useful practice by which, when a man wrote a new article, he cited references to any related articles that had previously been published. By following these links, we could fan out from one subject to related subjects, and work backwards in time to other pieces of recorded knowledge.

Towards the end of the nineteenth century, we found that this was not enough. The journals that publish original scientific articles we call primary journals. At the turn of the century we had to create secondary journals; these treat broad fields of science, and, without publishing full texts, give identifying particulars of the articles that may have been published in hundreds of primary journals. The secondary journals are often called abstracting journals, and, generally speaking, the "unit abstracts" that they publish are made up of three parts

- (1) a "bibliographic description" showing who wrote the original article, its title and where it was published;
- (2) an "abstract" proper, that distills the information from the original article, into a few hundred words;
- (3) a group of "subject headings" or "keywords", which identify the subject of the article and which are used in indexes that are constructed to help the reader find those articles that pertain to a particular subject.

The secondary journals were mainly started on a national basis. There has been a fair measure of effective international co-operation in their preparation, and this is fostered by the Abstracting Board of the International Council for Scientific Unions.

The abstracting journals have grown to be very large and very valuable. In atomic energy, we have especially to acknowledge the work of the dedicated people who prepare NUCLEAR SCIENCE ABSTRACTS at Oak Ridge, Tennessee, the REFERATIVNYI ZHURNAL in Moscow, and the BULLETIN SIGNALETIQUE in Paris. They have set very high standards, and their work has contributed greatly to progress.

Many other devices have been developed for making it easier to search through the body of recorded knowledge. When someone writes a book or review article, or when he prepares a compilation of data, he brings together the essential information from perhaps hundreds of individual articles, evaluates it, and presents it in more convenient form. Another important device is the holding of conferences, and the printed proceedings of such conferences are sometimes accepted as a definitive record of previous work, and a starting point for new research. Striking examples of this were given by the United Nations Conferences on the Peaceful Uses of Atomic Energy, held in Geneva in 1955, 1958 and 1964.

Indeed many of our scientific societies were founded mainly to bring orderliness into the flow of information in a particular subject area.

Another link which was especially important in the nineteenth century and in the first part of this century, was the acceptance of individual scientists as "world authorities" on particular subjects. A world authority was usually a professor at a respected university, and he had earned his position through his own researches and through the researches of the young people who had gathered around him. Thus, in the nuclear field, such people as Rutherford and Bohr played an important role in evaluating and transferring information. By working in their institutions, or by correspondence, one could obtain help in finding information, in interpreting new results, and in selecting promising topics for further research.

THE SEARCH FOR NEW SOLUTIONS

With the rapid proliferation of science, we have come to see that one man cannot be a world authority, except perhaps in very new fields and on very narrow topics. We have instead, the growth of what have been called "invisible colleges" - groups of perhaps 100 to 200 scientists who may be scattered in all parts of the world, but who have a whole network of informal means of communication. They encounter one another at international conferences, they travel around on sabbatical years and work in each other's laboratories, and they exchange preprints of the articles they are about to publish. But the invisible college is hardly an adequate system in pure science, and it has no real counterparts in applied science. Can we therefore find some mechanism that will replace the individual world authority?

Dr. Alvin M. Weinberg has made some very valuable contributions to thinking on this subject. In 1962, he was chairman of a committees set up by the late President Kennedy to study scientific information questions in the United States. In the report [2] of the committee Dr. Weinberg gave special emphasis to the creation of "information analysis centres". Each such centre covers a fairly narrow scientific topic, and the centre itself is housed as an integral part of a research establishment where relevant work is in progress. The centre collects information on a worldwide basis, analyses and evaluates it, and makes it available in new and varied forms. Perhaps the prototype was the Neutron Cross-section Compilation Group at Brookhaven National Laboratory. It collected the results of measurements of cross-sections, and after evaluation and normalisation, prepared graphs that indicated for each nuclide the best values of the cross-section as a function of neutron energy. These graphs began to appear in 1952, and have been a great boon to nuclear scientists throughout the world. The work is now being continued and expanded in four linked centres at Brookhaven, at Saclay, at Obninsk, and in Vienna.

With Dr. Weinberg, I believe that such information analysis centres will grow in number and in importance, and that they will ultimately attract some of the best scientific brains. They can become the new "world authorities", not only providing information about what has already been done, but also in identifying gaps in knowledge, and recommending where new research should be carried out.

JOBS FOR COMPUTERS

And what about electronic computers? Much effort has been devoted in the last ten years to the study of the application of computers to information work.

Can these immensely powerful tools bring a unit abstract more rapidly or at less cost to particular scientists, who, at this moment, would be helped by seeing it? The abstract journals are doing a fine job, but for every article that is written, probably dozens of unit abstracts are prepared. They are prepared not only by the abstract journals, but also by librarians, for their card catalogues, and by individual scientists, who keep boxes of cards on their desks for handy reference. Is it possible to prepare the unit abstract, once, to type it once, and to have computers print it out wherever and whenever it is needed?

One pioneer has been the National Library of Medicine in the United States, which has developed a computer-based system called MEDLARS. This handles descriptions of about a quarter of a million articles per year. The computer sorts these descriptions to produce the text of a secondary journal called INDEX MEDICUS, and of the indexes to it. Cumulative issues are generated with no retyping by hand, and the computer store can be searched to provide short lists of articles that respond to particular requests from individual medical research workers. Collaborating organizations in the United States, Japan, the United Kingdom, and Sweden, have either helped to prepare the information that goes into the system - or they have received copies of the magnetic tapes on which the processed information is stored. Some of these countries now use these tapes with their own computers to provide information directly to scientists in their own countries.

INIS - HISTORY AND AIMS

Such is also the aim of the International Nuclear Information System (INIS) - a scheme that is being studied by the Agency with a view to putting it into operation in 1970. The volume of literature in the nuclear field is not as great as in the medical field - but nevertheless we still have 100 000 new articles each year.

Of course, by its Statute, the Agency has always had a strong programme in fostering the flow of information. The work has been centred in the Division of Scientific and Technical Information of the Department of Technical Operations, and it has involved a library and bibliographic service as well as extensive programmes of scientific conferences and scientific publications. For these programmes credit must go to my predecessors, Mr. J.E. Cummins of Australia and Dr. Bernhard Gross of Brazil.

INIS is quite a new addition, and its history is quite short. The first document was prepared by two consultants, Dr. Isaev of the USSR State Committee on Atomic Energy and Dr. Wakerling of the University of California, who were working for the Agency in the summer of 1966. The general lines of their proposal were endorsed by an international group of experts that met in Vienna in December 1966, and the International Atomic Energy Agency was invited to "play a leading role" in the development of such a system.

In the last year and a half, the Agency, with considerable help from outside experts, has been attempting to define a scheme that will meet a consensus of the wishes that have been expressed by its Member States.

The basic idea is that we should have a decentralized system. The unit abstract would be prepared only once and in the country in which the article was published. This means that the effort required of each country would be in proportion to the scale of its nuclear programme. The unit abstract would be sent to Vienna, and processed along with all the others - about 4 000 - that arrived in the same two-week period. The keywords and the bibliographic description would be recorded on magnetic tape for computer processing. The abstract proper and any other textual material would be recorded on a microfiche, a single sheet of photographic film containing 60-72 pages of text. The concept is that, if the original article is not readily available through commercial channels, the whole of the text would be recorded on microfiche, which can then be used for screen projection or photographic copying.

INIS would thus be the construction of a store of data in the form of unit abstracts. Copies of the complete store would be given to co-operating Member States on magnetic tape and microfiche. Each Member State would then be able to use its copy of the data store to generate information services for its own scientists. The Agency recognizes that some Member States do not yet have the computing facilities to make proper use of a data store on magnetic tape; so there are also plans to have the Agency's own computer print out the contents of the magnetic tape on to paper; and then those who prefer to receive the bibliographic descriptions and keywords in book form can do so.

Member States will generate two types of service
- current awareness services
- retrospective searching.

The first is the announcement of new information to scientists according to their special interests. Thus, if one country has a particular interest in the use of isotopes in forestry research, it can programme its computer to process each new magnetic tape (i.e. once every two weeks), and select those items that deal specifically with this subject. The computer will, in fact, set in type an announcement bulletin that can be then copied and published, and the computer will also generate indexes to such a publication.

At the level of the individual scientist, the computer could narrow down to an even smaller fraction of its store and, every two weeks, print out a very short list of items of specific interest to the one individual. Of course, the advantages of computer processing is that it can do this for many scientists all at the same time and even address the sheets of paper to the individual concerned.

Retrospective searching involves finding, not only new articles, but also the older ones that deal with some highly specific subject. Generally, these searches take more computer time, because complete files must be scanned. However, such searches can yield information of immense value, particularly at the time when new research projects are started. Such searches help ensure that previous work carried out in other laboratories is not repeated simply because a reference has been lost.

There are of course problems. Not all the centres in the world use the same computers; different computers speak different languages and use completely different methods of storing information. So compatibility of equipment is one problem. Also, the computer is very intolerant of errors and inconsistencies. And if we have many different individuals spread out all over the world, with different language backgrounds, and all are trying to prepare unit abstracts for a common system, we will have difficulties.

In particular, unless the various professional indexers prepare good and accurate sets of keywords to describe the subjects of the various articles, the computer cannot improve on their job - and the keywords are vital for any subsequent processing that involves the selection of material by subjects.

The whole business of assigning keywords becomes something of an art. All the ideas must be there, but they must also be expressed in a way that is consistent with detailed codes of practice. Differences in language and cultural background have to be covercome by training and practice. A master set of keywords has to be generated, and rules have to be written for using them. Procedures have to be developed to ensure that the master list itself evolves to take account of new ideas and concepts, and to reflect the experience obtained when the system is in operation.

Fortunately EURATOM has been operating a keyword system on a computer for several years, and has developed a suitable master list of keywords in English (the language proposed for the INIS vocabulary). This experience forms a solid basis on which INIS could build.

The Agency recently recruited a team of specialists who worked for four months on a detailed analysis of the INIS proposal. Their report advocates the establishment of INIS, and defines a method of operation. The computer-based information system at EURATOM and those that have been developed in other subject fields, have demonstrated that computers can provide a very powerful mechanism for improving information services; this gave confidence to the team.

However, in all previous systems, the preparation of the unit abstracts and the operation of the system itself has been under relatively tight central control. The particular novelty of INIS lies in its decentralized and internationalized features, particularly in the original preparation of unit abstracts.

COSTS OF THE PROJECT

This decentralization of work keeps down the amount of staff and equipment needed in Vienna. For 1970, the first year of operation, the report indicates a gross expenditure by the Agency of $500 000. However, some of this represents once-only purchases of equipment, and in 1971 the gross expenditure falls to $425 000. I would like to make three points in connection with these estimates.

(1) Because of existing programmes, the Agency is already budgeting in 1969 for many of the staff and most of the computer equipment that will be needed when INIS is in operation. A rough calculation indicates that the 1971 gross expenditure is only about $100 000 more than what has already been budgeted for 1969.

(2) The figures are for gross expenditures. Revenues from the sales of INIS products, particulalry microfiches, may eventually bring in as much as $100 000 per year.

(3) The figures are only costs to the Agency. Since the Member States will be preparing the unit abstracts in their own countries, and since they will be using INIS tapes and other products as a basis for operating

national nuclear information services, they will have to budget for local expenditures at home. The total of such expenses across all the participating countries will be many times the expenditure in Vienna.

In the course of the next two years, the Agency and its Member States have a lot of detailed work to do, writing computer programmes, testing systems and methods, writing manuals, and organizing training courses. Only time will tell whether INIS as it has now been defined for us is a workable system; as I indicated before, the main question is whether adequate consistency and accuracy can be obtained in a highly decentralized system. If difficulties are encountered, however, it should be possible to overcome them by bringing the more troublesome operations under a closer central control. And the scientists' appetite for information is so insatiable that I am sure experiments will go on until a satisfactory system is found.

The second of the special lectures given during the General Conference was by Dr. V. Sarabhai, Chairman of the Atomic Energy Commission of India, on the subject "Nuclear Power in Developing Countries". An article based on this lecture will appear in the next issue of the Bulletin.

REFERENCES

[1] PRICE, D.J. de Solla, Little Science, Big Science (New York, N.Y.: Columbia University Press 1963)

[2] U.S. President's Science Advisory Committee, Science, Government and Information (Washington, D.C.: Government Printing Office, 1963)