International Atomic Energy Agency

INDC(NDS)-38/G

INDC-410

INDC INTERNATIONAL NUCLEAR DATA COMMITTEE

Comments on the UNISIST Study

L. Hjärne

May 1971

IAEA NUCLEAR DATA SECTION, KÄRNTNER RING 11, A-1010 VIENNA

I. Introduction

In their joint UNISIST study [1, 2], UNESCO and ICSU have addressed themselves to a very wide complex of problems. Consequently, the synopsis and the report are formulated in very general terms and have little tangible, substantial content. This makes the two documents difficult to comment upon, and my own misgivings may best be expressed by means of a few examples of what, to my mind, has received too little emphasis or has even been left out of the study.

At the outset it should be pointed out that we lack a clear definition of the concept "information system". Some people have given it a blurred and apparently devalued meaning which contrasts sharply with the definition accepted by others. As I see it, it is just as logical to call a documentation system an "information system" as it would be to call a telephone directory (including a system for its printing and updating) a "telecommunications system". I would understand an information system to be one that encompasses all aspects of information acquisition, selection and transfer. There is within some disciplines a certain risk in an increasing number of countries that bodies with which the political decisions rest might be given the impression that the funds devoted to documentation systems would contribute to the solution of the more general information problems, simply because of the name given to the activity.

Almost a decade ago, a top-level committee in the United States published a report, the so-called "Weinberg-Report" [3]. That committee made a very thorough survey of the information situation and related problems in the scientific community. It is surprising to find in the documents concerned with UNISIST that many of the very basic ideas in the Weinberg Report have been completely neglected. If we address ourselves to information problems, these aspects should be of great significance, whereas if we limit ourselves to documentation questions, the whole complex of problems is considerably reduced, so that, in fact, the justification for the planned effort on the part of UNESCO becomes slender indeed. Let me quote from the aforementioned Weinberg Report:

> "Because most of the schemes and devices for handling information are so new, their limitations are still not fully understood; in particular, it is not usually appreciated that the new systems generally retrieve documents rather than information".

It would appear that the last decade has brought the world no nearer to an understanding of these points.

II. The specialized centre

In fact, the specialized centres are the only ones whose function it is to bridge the gap between the, possibly, selected but undigested mass of documents and the fact-seeking scientist. Seen from the viewpoint of a specialized information centre, the information should produce an interaction, enabling the scientist to hold a dialogue with the rest of the community. In other words, the "feedback", which already Norbert Wiener described in his book on cybernetics [4], is lacking if the communication channels are not functioning both ways. Documents of the conventional kind provide only a one-way communication channel.

One obvious example of the disadvantages of conventional journals is the lack of an effective mechanism to communicate to readers the corrections of errors which have escaped the attention of the reviewers, and which have been discovered after the publication of a document. The data information centre, on the other hand, can remain in some contact with authors and can maintain sufficient records of queries to follow up any changes and corrections and provide the users with more correct and up-to-date information.

Another example of an information problem which the data centres are in a unique position to tackle is the proliferation of unnecessary publications. I am convinced (although many scientists' recognition of the centres' usefulness in this respect is slow indeed) that we have already been spared a certain volume of duplicate publication and of publication of smaller additions to already published work thanks to the existence of the centres. This is certainly true in the field of neutron nuclear physics.

These examples are intended to refer again to the Weinberg Report, which, in contrast to the UNISIST study discusses the topics concerned. The crucial difference between the documentation system and the data centre is that the centre works with the primary information - the information itself - rather than with the secondary information - information about information. This is the background to the data centres' use of the full spectrum of information activities, and it is the basis for a centre's "total information" system, of which the documentation is only one part. A full review of the neutron data centres' activities would cause us to digress too far, but I wish to refer to published work on the topic [5, 6, 7].

III. Management information

One sector of "total information", largely akin to the management information systems, is what the Weinberg Report called scientific intelligence. The information needed for all levels of management in Science is partly contained in the conventional documentation, and partly missing because of a lack of communication channels.

In the field of nuclear data there is a mechanism which has already been of great help to science management in the field. Requirements for new information on topics about which sufficient experimental data do not exist have been published regularly for a number of years. The individual requests, submitted together with a full justification by the requestor, have been given priorities according to certain criteria. The requests are subsequently reviewed by a competent nuclear data committee. The request list can then serve to justify an experimental effort at an institute, and thereby also to link the research effort with the areas of greatest importance; the research programmes - institutional, national, regional or international - are consequently accelerated in a natural way. The experience of this tool within the area of OECD countries represented on the European American Nuclear Data Committee (EANDC) has been quite favourable over the years. The scheme is now being extended to operate at international level, within the framework of the International Nuclear Data Committee (INDC), which will have a great responsibility in screening the individual requests for the international nuclear physics community. The system (RENDA) [8] that has been used by ENEA in the OECD countries will be modified for international use in the near future.

When such words as priorities, justification and purpose are mentioned in this context, a great many scientists will jib violently, particularly those who claim areas of Very Pure Science. Nevertheless, the means available for scientific research are necessarily limited, and the, more or less, political decisions on how to divide the resources have always been made somewhere and will inevitably continue to be made in the future. When such decisions are taken, they should be based as far as possible on available facts. It is the responsibility of the information system to make those facts available to those who need them.

IV. Data information

Much of the development experience in the field of nuclear data may be rather specific to the field and not readily applicable to other areas. Some of it, on the other hand, may be useful to others. The most important lesson in the operation of a data centre that may have an impact on information systems is the experience of frustration with existing documentation systems. Consequently the neutron data centres are simply forced to rely upon their own resources for documentation services. CINDA [9] is a reference system which can now call upon some fifteen years of experience of fully computerized operation. For the past six years it has enjoyed world-wide recognition and constitutes a truly international enterprise; it was the first international computerized documentation system.

No criticism is intended of such a publication as Nuclear Science Abstracts, which is a very useful documentation tool for an individual scientist. The system does appear, however, to be quite incompatible with the documentation sector of the data centre's operation [10]. The basic reason is simply that the centre must regard the whole field of information as the appropriate approach rather than accept a separation between documentation on one hand and data information services on the other. For the data centres to adjust their documentation systems in order to make use of so-called "information systems" of wider scope would simply be much too expensive in terms of efficiency and coverage.

A second lesson that we can draw from the international collaboration in our field is that data themselves are not very useful unless they are related to extensive factual and peripheral alphanumeric information. In particular, the classification and categorization and the unambiguous definition of the details in experimental nuclear physics required long and painful work on the part of the four centres. We are now beginning to see good results of this work, after one year of operation of the international exchange format for neutron data information, EXFOR. The international exchange of experimental neutron data has been in effect for several years, but since 1970 the exchange has been made in a commonly agreed format. This format includes an internationally agreed classification, bibliographic terms, physics terms, and a communication format.

During the development of EXFOR, a great deal of thought was devoted to the use of the information subsequent to its transmission to the user. The format is designed to serve sophisticated users who have access to display devices with man-machine-interactive facilities, as well as users with no equipment at all, and also users with more peripheral interests.

In this context, it might be worthwhile to emphasize that little mention was made in the UNISIST report and synopsis of the following essential topics which were well discussed in the Weinberg Report, to which I refer the reader for further study:

2B: The Information Process as Part of the Research Process

3E: Modern Psychological Insights into Communication should be Exploited

3K: More and Better Specialized Information Centres are Needed.

[See ANNEX II]

V. Conclusions

In my opinion, it is important that the UNISIST "movement" clarifies its objectives before taking further steps. Leaving the semantics aside, it should either restrict itself to documentation, and as a consequence, delete or redraft recommendations 9 and 10, or it should extend its view from the narrow documentation standpoint into a wider understanding of the "total information" complex of problems and take into account also information media other than documents. This would entail redrafting

recommendations 4,5,11,12,14,15 and 16 [see ANNEX I]. Only the second alternative would seem to justify the proposed size of the operation.

There is no doubt that the task at which UNISIST is aiming, is an important one. The volume of information services is greatly increasing, which will have great impact on Science at the working level, and some co-ordination of this development is commendable, necessary and certainly worth while.It is therefore a matter of urgency that UNISIST be given the best possible start in the right direction.

Bibliography

1. "UNISIST - Study Report on the Feasibility of a World Science Information System by the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions"; Paris, Unesco House, 1971.

2. "UNISIST - Synopsis of the Feasibility Study of a World Science Information System by the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions"; Paris, Unesco House, 1971.

3. A.M. Weinberg et al. "Science, Government and Information. The Responsibilities of the Technical Community and the Government in the Transfer of Information". A report of The President's Science Advisory Committee, Washington (10 Jan. 1963).

4. N. Wiener: "Cybernetics, or control and communication in the animal and the machine" Second ed., M.I.T. Press, New York (1961) (First edition published 1948).

5. J.J. Schmidt: "Neutron Nuclear Data Compilation and Evaluation - Past, Present and Future". CODATA Newsletter no. 3., ICSU CODATA, Frankfurt/Main (Dec. 1969).

6. L. Hjärne: "The International Neutron Data System". Proceedings of a Symposium "Handling of Nuclear Information" pp. 73 - 82. IAEA, Vienna 1970.

7. M.D. Goldberg: A paper presented at the 33rd annual meeting of the American Society for Information Science, 11-14 October 1970, in Philadelphia.

8. S. Schwarz: "RENDA, System Description", Report EANDC-77 "U" (Nov. 1968) and "RENDA, Compilation of EANDC Requests for Neutron Data Measurements", Report EANDC-85 "U", (April 1970)

9. "CINDA 71, An Index to the Literature on Microscopic Neutron Data", published on behalf of USAEC/DTIE, USSR/CJD , ENEA/CCDN and IAEA/NDS by the International Atomic Energy Agency, Vienna (Jan. 1971).

10. Report INDC(NDS)-30, Recommendation IV of a Consultants Group of the IAEA:

> "The Consultants group has found that, in their present form and quality, the information services from documentation systems linked to INIS, such as Nuclear Science Abstracts, are not very useful to the data centres. In view of the potential value of a collaboration, it is therefore urgently recommended that the IAEA, within the framework of INIS, venture to find appropriate ways and means for the documentation systems to establish an appropriate interface with the activities of data centres in nuclear physics."

ANNEX I

Suggested Modifications to the UNISIST Recommendations

In this annex I shall outline briefly the suggested modifications to the recommendations. The two that most urgently need to be revised are recommendations 4 and 12. For the latter one I do not propose a new formulation here. In fact, recommendation 12 is the only one describing the information responsibilities of the individual scientist, and it can therefore be regarded as the most important recommendation of them all.

Recommendation 4. Subject specification

The attention of scientists, learned societies, and information science associations should be drawn to the need for joint efforts in developing better tools for information in depth in the various disciplines and sub-disciplines in science and technology. UNISIST adherents should be invited in particular to consider the support of a few pilot projects aimed at testing new methodological or organizational devices in this respect, with a special emphasis on the interrelationship between international, interdisciplinary documentation services and the "in-depth" information services.

Comment: "...the control and conversion of natural and indexing languages ..." appears to refer only to the conditions of documentation services. I should rather suggest that the problem be tackled on the only level where this is possible: The scientists themselves in a specific disciplin must find agreements on classification and terminology. Otherwise the efforts would inevitably be reduced to fruitless arguments about semantic details.

The best sponsorship of a pilot project can be discussed, but I should think that this is not the place to indicate any preference in this respect.

This recommendation is for the specialized centre of greatest importance. The work on terminology and classification can be coordinated through the centre, which should be in a unique position with its direct contacts with leading scientists in its field. In order to do this effectively, the strong support of the interdisciplinary documentation systems would be very valuable.

Recommendation 5 Standardization: systems aspects.

Provision should be made within UNISIST for active consultations with computer and information systems experts in order to speed up the resolution of pending issues in matters of machine standards for systems interconnection, when applicable in agreement with ISO. UNISIST should encourage adhering documentation systems to develop an effective interface with specialized information centres in their respective fields.

Comment: This is not the place to discuss whether or not a "common carrier" format is feasible or even desirable for all branches of science. In fact, compatibility should certainly suffice and should be discussed where there is a need for it. Clearly, for documentation there are good reasons for internationally agreed standards.

As an example of an effective interface between specialized centres and documentation systems one could design a scheme whereby a user of the services of a data centre could - as part of the service - be provided with all the documents related to the topic of his query from the documentation service.

Recommendation 11. Responsibilities of scientific editors and referees.

Editors and referees of scientific journals should exercise a special responsibility for the maintenance of quality controls of published material. They must realize that the primary publications for which they are responsible are part of an information processing continuum and that their cooperation with other groups, particularly those who process primary information (e.g. specialized information centres and data centres) is essential.

Comment: The authors' responsibilities belong in recommendation 12 rather than in this one. The referee system should not only be encouraged; there should be made a clear distinction between primary literature which is subject to review and other scientific publication. The fact that an article has been reviewed by a referee is an important piece of evidence: Particularly in the context of interdisciplinary flow of information the referee's "screening" function is a great help for the scientist who is not a specialist on the topic on which he seeks information.

Regarding the "other groups", it is rather obvious and it needs not be spelt out in the recommendation that the cooperation with indexing and abstracting services should be encouraged, particularly if no effective specialised centre exists for a certain field. The formulation of this recommendation in the UNISIST report, however, gives me the impression of an argument in some sort of friendly quarrel between primary and secondary journals, and this should be avoided in the context of UNISIST.

Recommendation 12. Responsibilities of the scientists

Comment: The reformulation of this recommendation is rather difficult, but an urgent and important one. Already the title of the recommendation should (as indicated above) be addressed directly to the scientists, rather than to the scientific unions. The individual scientists should be addressed in their capacity as (1) producers of new information, (2) authors of articles, (3) educators of young scientists, (4) users of information services and (5) users of the information provided by the information services. This topic should perhaps have deserved several recommendations, one of which could have been directed to the scientific societies. The disproportionately small emphasis that the topic has received in the UNISIST study is rather obvious from a comparison with the Weinberg report. Consequently, I have here refrained from giving a new formulation of the recommendation. This is the weakest one of all the recommendations, and new wordings are urgently needed.

Recommendation 14. Research in information science.

The present effort of national and international organizations to conduct and to support research on many aspects of information science should be acknowledged and encouraged by UNISIST adherents as contributions of vital importance to the evolution of world-wide scientific information services. A group should be established within the framework of UNISIST to collect and evaluate information on research in this field, to assist UNISIST adherents on systems planning matters and to promote international co-operative programmes in information science analysis and research activities.

Comment: Only minor revisions of this recommendation were felt necessary.

Comment to recommendations 15 and 16:

Again, it is not clear, whether a "common carrier" format is at all feasible. This prerequisite for integration into comprehensive

systems has been assumed when describing every library or centre as an access point or "node" in a switching network in recommendation 16. The recommendations 15 - 17 might set certain less desirable limitations to the development of specialized information centres, particularly the data information centres. The organisation of a world-wide network of data centres in a given field will depend very much on the boundary conditions in that field. For example in the case of neutron data the world has been subdivided into four "service areas" where conditions vary considerably from one area to another as regards computer development, data users, kind of data produced, political and economical situation etc. However, these conditions would conceivably vary from one topic to another, and a solution for each case should be sought accordingly. The interdisciplinary linkage must be the responsibility of the documentation systems which may this way eventually evolve into subsets of true information systems. Particularly the over-emphasis on national agencies should therefore be softened. In fact these recommendations (and possibly also rec. 17) could well be replaced by appropriate recommendations to the scientists as described in my comment to recommendation 12.

ANNEX II

Sections 2B, 3B and 3K of the Weinberg Report as referred to on page 4 are reproduced here for convenience.

Carrying further the thread of argument of the previous paragraph, we come to perhaps the most essential attribute of the information process: *the information process is an integral part of research and development.* Research and development cannot be envisaged without communication of the results of research and development; moreover, such communication involves in an intimate way all segments of the technical community, not only the documentalists. The attitudes and practices toward information of all those connected with research and development must become indistinguishable from their attitudes and practices toward research and development itself. This is the central theme of our report.

We place special stress upon what seems an obvious point because, in the early days of science, the problem of communication could be managed casually. Each individual scientist could work out his own private communication system, suitable to his own needs, and, since the requirements were relatively small, the whole matter could be treated rather incidentally. But with the growth of science a casual attitude toward communication can lead only to insufficient communication. Scientists individually, technical societies, agencies supporting research and development, will have to recognize that adequate communication no longer comes free. Communication cannot be viewed merely as librarians' work; that is, as not really part of science. An appreciable and increasing fraction of science's resources, including deeply motivated, technical men as well as money, will inevitably have to go into handling the information that science creates.

Science can ultimately cope with the information expansion only if enough of its most gifted practitioners will compact, review, and interpret the literature both for their own use and for the benefit of more specialized scientists.[6] The Panel believes that such activities may eventually achieve a position in the science of the future comparable to that of theoretical physics in modern-day physics. Recognition of the importance of such scientific middlemen is discernible in the proliferation of the so-called specialized information center where information is digested and interpreted. The Panel views the specialized information center as one key to ultimate resolution of the scientific information crisis.

E. *Modern Psychological Insights Into Communication Should Be Exploited*

New information could be made easier to assimilate, and in this sense easier to retrieve, if authors wrote better. We do not understand the communication process well enough to know how our natural language can be made into an instrument for the most effective presentation of scientific and technical information, but progress is being made. Advances in our understanding of the communication process should become known to authors and to the information-handling community, and should be put to work in the improvement of our technical writing. Nor should devices other than improvements in the natural language be ignored. Recognizing the danger of creating too many highly specialized languages, we point out nevertheless that symbols or conventions to replace wordly clichés or to describe commonly used methods of instrumentation could reduce the volume of the literature and help ease its retrieval. Or judiciously used journalistic techniques, such as different type fonts, display boxes, different colors, might help to make the technical literature easier to assimilate. Many of these techniques might be repugnant to those brought up in the conservative scholarly tradition, yet if further study and experiment show them to be effective, the

K. More and Better Specialized Information Centers Are Needed

The centralized document depository is primarily a clearinghouse for documents; in general, it does not try to glean information from the documents it handles, but merely provides appropriate documents to users. But retrieval of documents is not the same as retrieval of information; a technical specialist really needs the information contained in the published literature, not the published literature itself. To retrieve information, as contrasted to documents, the technical community has devised the specialized data and information center.

A specialized information center makes it its business to know everything that is being published in a special field--such as nuclear spectroscopy or the thermophysical properties of chemical compounds; it collates and reviews the data, and provides its subscribers with regularly issued compilations, critical reviews, specialized bibliographies, and other such tools. Its input is the output of the central depository. There are now in the United States about 400 such centers; the net number is growing, though some specialized information centers can and should die because the fields of science they serve cease to be active. As originally conceived, the centers compiled data as opposed to ideas or knowhow; one of the earliest *data* centers compiled the *International Critical Tables*. Many of the data centers have evolved into *information* centers that not only compile data but also keep abreast of all developments in a field.

We believe that the specialized information center, backed by large central depositories, might well become a dominant means for transfer of technical information. It therefore behooves the technical community, at this early stage in the proliferation of specialized centers, to learn what makes a good specialized center, and to plan new centers accordingly.

Specialized information centers, to be fully effective, must be operated in closest possible contact with working scientists and engineers in the field. The activities of the most successful centers are an intrinsic part of science and technology. The centers not only disseminate and retrieve information; they create new information. Making a discriminating selection of data, as was done in preparing the *International Critical Tables*, requires scientific insight of high order, and is itself an essential scientific activity. The process of sifting through large masses of data often leads to new generalizations. The Nuclear Data Center that collects and distributes information on the static properties of nuclei contributed notably, for example, to the development of the shell model of the nucleus, one of the major theoretical underpinnings of modern nuclear physics. What is true of the Nuclear Data Center is undoubtedly true of other centers. In short, knowledgeable scientific interpreters who can collect relevant data, review a field, and distill information in a manner that goes to the heart of a technical situation are more help to the overburdened specialist than is a mere pile of relevant documents. Such knowledgeable scientific middlemen *who themselves contribute to science* are the backbone of the information center; they make an information center a technical institute rather than a technical library. The essence of a good technical information center is that it be operated by highly competent working scientists and engineers—people who see in the operation of the center an opportunity to advance and deepen their own personal contact with their science and technology. Proliferation of the specialized information centers will therefore require many such "information scientists": dedicated and knowledgeable technical men who help interpret and assimilate the literature for others working in the field.

Since the technical information center in this sense must be part of science and technology, it is natural that it be located where relevant science is flourishing. The Panel therefore urges that new information centers be established at public and private technical institutions, not as adjuncts of general libraries, or of publishing ventures, or of central depositories. Where research and development is done for the Government—at Government laboratories, national laboratories, universities, or industrial laboratories—information centers in related fields ought to find a congenial atmosphere. We note with approval that AEC has already established about a dozen such centers at its national laboratories, and we believe this