DOCUMENT RESUME

ED 065 139 LI 003 760

AUTHOR	Vilentchuk, Lydia, Ed.; Haimovic, Gila, Ed.
TITLE	ISLIC International Conference on Information Science: Proceedings (Tel Aviv, 29 August-3 September, 1971). Volume I.
INSTITUTION	Israel Society of Special Libraries and Information Centres, Tel Aviv.
SPONS AGENCY	National Center of Scientific and Technological Information, Tel Aviv (Israel).
PUB DATE	72
NOTE	345p.;(140 References)
EDRS PRICE	MF-$0.65 HC-$13.16
DESCRIPTORS	Conference Reports; *Economic Factors; *Information Networks; *Information Science; *Information Systems; International Programs; National Programs; Specialization
IDENTIFIERS	*Information Analysis

ABSTRACT

Information science is a young and rapidly growing field, embracing a wide range of subjects and activities. The conference, in ten technical sessions, attempted to cover various aspects of information work. The areas covered were: information analysis and information analysis centers; retrieval of information; selection, education and training of personnel; and publishing and reprography. The proceedings volumes are arranged in order of sessions. Summaries of discussions follow the papers presented at each session. Volume one includes papers delivered under the following broad topic categories: international and national information networks; information systems for specialized applications; information analysis; and economics of information systems. A table of contents for the entire proceedings is included in this volume. (Volume two of the proceedings is LI 003 761.) (Author/SJ)

Israel Society of Special Libraries and Information Centres אגודת הספריות המיוחדות ומרכזי המידע בישראל

ISLIC International Conference on Information Science

TEL AVIV, 29 AUGUST – 3 SEPTEMBER, 1971

proceedings

VOLUME 1

edited by LYDIA VILENTCHUK
assistant editor Gila Haimovic

NATIONAL CENTER OF SCIENTIFIC AND TECHNOLOGICAL INFORMATION

Library of Congress Catalog Card Number: 72-000007
UDC Number: 002:061.3(100)

Published by the:
National Center of Scientific and Technological Information
84 Hachashmonaim Street, Tel Aviv

Printed in Israel
by Multigraph Ltd

F O R E W O R D

Information Science is a young and rapidly growing field, embracing a wide range of subjects and activities. The Conference, in ten technical sessions, attempted to cover various aspects of information work. The areas covered were:

* Information Analysis and Information Analysis Centers
* Retrieval of Information
* Selection, Education and Training of Personnel
* Publishing and Reprography

In addition to the 50 papers delivered at the Conference, the Proceedings include 18 papers which for technical reasons could not be presented orally.

An open meeting of the International Federation for Documentation Study Committee "Information for Industry" (FID/II) was held at the conclusion of the Conference. A summary of this meeting is included in the Proceedings.

An FID/TM tutorial session, "System Analysis, an Approach to Information," by H. Borko, H. Schur, G.X. Amey and K. Samuelson, was offered to the participants in the framework of the ISLIC Conference. The tutorial session was originally presented in Buenos Aires on 23 September 1970 and published by the FID/TM Secretariat, Stockholm.

Thanks are due to the National Center of Scientific and Technological Information and its Director, Mr. Carl Keren, for the financial support which made publication of the Proceedings possible.

It is hoped that these Proceedings, which reflect the development of the profession, will be a useful contribution to its members.

Mrs. Lydia Vilentchuk
Chairman, Organizing Committee
International Conference
on Information Science

EDITORS' NOTE

The Proceedings are arranged in order of sessions. Summaries of discussions follow the papers presented at each session.

The discussion summaries underwent semantic editing for the sake of clarity and brevity. To avoid repetition, they include, as far as possible, only facts and ideas which have not been stated in the papers.

Papers not submitted in accordance with the prescribed format were retyped. The others are reproduced as submitted, with corrections inserted only at the request of their authors.

The Proceedings are printed in two volumes. Indexes for both volumes appear at the end of Volume Two.

Thanks are due to Mrs. Elana Covo for her patience in typing the papers and discussions.

ORGANIZING COMMITTEE

Mrs. Lydia Vilentchuk
National Center of Scientific and Technological Information, Tel Aviv, Israel

Mr. Uri Bloch
Armament Development Agency, Ministry of Defence, Haifa, Israel

Mrs. Susane Weil
Sorek Nuclear Research Center, Yavne, Israel

REVIEWERS

Area 1: Mr. Jack Mills, Northwestern Polytechnic, United Kingdom
Dr. Ch. Schaefer, CODATA, Germany

Area 2: Prof. Harold Borko, University of California, US

Area 3: Prof. W.L. Saunders, University of Sheffield, United Kingdom

Area 4: Mr. C.M. Spaulding, Council of Library Resources, US

SESSION CHAIRMEN

Session 1: Mr. W. Kenneth Lowry, Bell Telephone Laboratories, US

Session 2: Dr. Martin Cremer, Institut fuer Dokumentationswesen, Germany

Session 3: Prof. Dr. Helmut Arntz, International Federation for Documentation, Germany

Session 4: Mr. Wolfram E. Uhlmann, Research Institute of Swedish National Defence, Sweden

Session 5: Dr. A. Cockx, National Center for Scientific and Technical Documentation, Belgium

Session 6: Prof. Jacob Gross, Ministry of Commerce and Industry, Israel

Session 7: Mr. J.H. d'Olier, Centre de Documentation du CNRS, France

Session 8: Mr. Denys G. Kingwill, South African Council for Scientific and Industrial Research, South Africa

Session 9: Prof. Aron Beged-Dov, Weizmann Institute of Science, Israel

Session 10: Prof. Th. P. Loosjes, Library of the State Agricultural University, Netherlands

table of contents

Volume One

*not delivered orally

Volume Two

EVALUATION OF RETRIEVAL EFFECTIVENESS
A. Cockx, Chairman, Session Five

SELECTION, EDUCATION AND TRAINING OF PERSONNEL
J. Gross, Chairman, Session Six

addresses

OPENING ADDRESS BY CHAIRMAN OF
THE ISRAEL SOCIETY OF SPECIAL LIBRARIES AND INFORMATION CENTRES (ISLIC)
DIPL. ING. LYDIA VILENTCHUK
at the OPENING CEREMONY of the

INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE
August 30, 1971, at the Sheraton Hotel, Tel Aviv

On behalf of the Israel Society of Special Libraries and Information Centres, I have the great honour and privilege of greeting this illustrious gathering of colleagues from other countries and friends and colleagues residing in Israel.

We who are assembled here this evening represent a profession - the profession of Information Science - that is still young and vibrating with endless possibilities and far-reaching vistas. This is the case not only in Israel but in the world at large. A symptom of the youthfulness of this profession is the continual controversy regarding its exact place among other professions - more mature and with longer traditions - and the role it should assume in the cultural, technological and scientific life of modern society.

The controversial nature of the outlook on Information Science is reflected in the fact that in almost all the developed countries where it is an already recognized profession, a profession that has come to stay, the content, scope and objectives of Information Science are differently understood, and oscillate over an infinitely wide range - all the way from a more sophisticated approach to librarianship to the mathematical theory of communication. Even in discussions of problems with individuals of seemingly identical background, the same concepts are not always understood under the same terminology. However, it is this very fluidity, this absence of a rigid form, this opportunity that we - the first generation of information scientists - have to shape and mould the profession, to give it the content and direction which the most gifted among us will eventually work out, that exercises a powerful attraction for all those engaged in this field.

Today, as I have already pointed out, we are still in the pioneering stage. True, the roots of our profession first began to sprout away back at the turn of the century when Belgians Paul Otlet and Henri La Fontain first envisaged the compilation of a classified comprehensive index to published information gathered from the whole wide world. As we all know, the project proved too ambitious, too far ahead of its time, to be viable, since at that time it was not yet understood that science is truly international and does not recognize national boundaries, and it therefore follows that scientific information, as an indispensable and primary tool of scientific development, must also be international in character.

Nevertheless, out of this impractical vision there congealed as a practical residue the useful cooperation of nations within the framework of FID - International Federation of Documentation - which nurtured the spark of international cooperation generated by Otlet and La Fontain, and kept it from dying out.

Today, international cooperation in the exchange of information is recognized as indispensable to the progress of science and technology and so more fertile ground for new organizations with more and more ambitious aims has been prepared.

Whether it be UNIDO - United Nations Industrial Development Organization - with its declared aim of disseminating information on technological innovations originating in various countries and assisting in the practical application of this information, or UNISIST - Universal System for Information in Science and Technology, a flexible network based on voluntary cooperation of existing and future information services, or other organizations, we look ahead to their development into a functional tool of international cooperation in the field of information services.

Let us hope that our assembly in Israel today will prove a milestone along the highway to international cooperation in the sphere of Information Science. I look forward to a fruitful and stimulating week that will bring us closer together and pave the way to the intensified sharing of experience and ideas.

In concluding, I should like to express the hope that our visitors from abroad will truly enjoy their sojourn in our small country, and learn at least a little of our day-to-day problems - both professional and human.

ADDRESS OF H.E. YIGAL ALLON, DEPUTY PRIME MINISTER
at the OPENING CEREMONY of the

INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE
August 30, 1971, at the Sheraton Hotel, Tel Aviv

This Conference is dedicated to the communication of knowledge and I wish to welcome our guests from abroad and the Israelis who have made this important field their vocation.

Any economic and political system depends on a complex and interconnected chain of events and decisions embracing past knowledge, new research, production of goods and services, definition of national goals and the socio-economic structure which these activities support. There is not a single activity in which the availability of information and the process of its communication do not have a strong influence on the structure of national growth and the nature of the political decisions which guide it. Managers and decision-makers need information from many disciplines and from all parts of this world. Information pertaining to legal, economic, sociological, political, scientific and technological knowledge is necessary in order to make national decisions after full consideration of the many options and their possible consequences.

The government daily faces decisions which cannot be made without comprehensive, correct, timely and relevant information. We need information in science and technology because it facilitates effective progress in areas vital to the well-being of our nation, be it the introduction of an improvement in housing construction or a development vital to national defense. We need economic information because only that enables us to make decisions on how to put our limited resources to best use for the advancement of industry and trade. We need information in the social and behavioral field because a trial and error approach to human problems is fraught with danger and is time and money consuming.

We have high hopes, some of them already justified, for the role which Research and Development should play in our society. The struggle for physical security, the creation of a society in which people from many different backgrounds and cultures can live, work and produce together, building a sound economy in an inhospitable environment, all pose a formidable choice of priorities. This year we are spending about 2.5% of our GNP on Research and Development. The proportion of active scientists in the total labor force in Israel is about 40:10,000 which is amongst the highest in the world; and in the number of students in institutions of higher learning relative to the total population Israel figures third in the list of nations.

During the last few years we have thoroughly reviewed our R&D organization with the object of promoting applied research and making better use of the opportunities with which modern technology presents us. We have also examined the information services supporting R&D and the decision-making level in government and management. Here we found room for improvement - organizational, budgetary and otherwise. As a first step we decided on the creation of a national network for scientific and technological information in which a national center is to serve as a focus and coordinator of activities. Basically we have thus adopted the structure which the OECD Ministers have recommended for their member states and which are now implemented there. We have recognized the need for a science information policy and we intend to build up the necessary resources for its implementation.

We are aware that knowledge is global. A small country like Israel, where less than one half percent of all active scientists of the world live and work, needs close and efficient tools of interconnections with the international community. Knowledge knows no national borders and well-functioning systems for international information transfer are a vital necessity for us. We will therefore strongly support the UNISIST concept and any other multilateral or bilateral activity promoting free information flow and international cooperation in this vital field.

When I say that due to our small size we are heavy users of knowledge produced elsewhere, I do not intend to imply that we have nothing to contribute in return. On the contrary I am proud to say that we make freely available our technical and scientific know-how. Thus for example more than 15,000 students from developing countries have been in Israel during the last 15 years and more than 3,500 expert missions have in that time assisted other nations in solving problems in which we have accumulated expertise. Even in such a new field as information science some countries seek to learn from our experience and here - as in other disciplines - the two way flow of information has the full encouragement of Government.

For information workers in Israel this Conference is a land-mark. It is the first time that they gather in such large numbers in an international forum, demonstrating that the role of information services and their importance to Research and Development and the decision-making process of government and management are fully recognized. The Israel Society of Special Libraries and Information Centres is to be congratulated on the organization of the Conference, which for such a young organization is no mean undertaking.

In conclusion I would like to thank our guests from abroad for coming and sharing their considerable experience with our information community. I would like to assure the Israelis in this audience that we view their place in the infrastructure of our Society and their contribution to Research and Development and to managerial decision-making as a vital one, which has our fullest support.

I wish you all a fruitful exchange of ideas and a successful Conference.

KEYNOTE ADDRESS OF FID VICE-PRESIDENT
PROF. HELMUT ARNTZ
at the OPENING CEREMONY of the

INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE
August 30, 1971, at the Sheraton Hotel, Tel Aviv

Your Excellency, Mrs. Chairman, Ladies and Gentlemen,

Anyone who attempts to assign the information sciences their place within the framework of the traditional branches can only be surprised at the frequently chaotic state in which they present themselves. There is not even agreement about the terminology of their respective departments, let alone about the domains of science that are to be counted among the information sciences, or their mutual delimitation.

In this it is possible to see the expression of an evolution that is to be welcomed, an evolution in which a new branch of knowledge suddenly acquires its place within mankind's mental make-up, and the composition of its intrinsic character can hardly keep pace with the task assigned by society to this young department.

But the field is young only in the development of its methods, as was the case with sociology decades ago and with jurisprudence centuries ago, both of which made the age-old ways of human behaviour the substance of research and teaching. To a greater extent even than these, and even than all technologies, information science can go back, since it must already have played a decisive role at the moment homo sapiens came into being, and made possible his continued existence, his development and, within the course of this, culture and civilization.

For its part, this earliest period of information owes its origin to a special biological case: the enlarged capacity of certain brain centres resulting from a number of mutation processes. Parallel with this it became possible for physical attributes of the preservation of life to take on a lesser importance, or even to become superfluous, because something else, something non-animal offset the lack of dexterity, inadequate development of the senses, and physical weakness. This was the one-move-after-another game, the parallelism between the development of the brain and the utilization of information.

Only the human brain is capable of foresight which makes the planning of action and cooperation at a high level possible. Not only has it ensured the survival of mankind by the acquisition and placing at the disposal of man the information essential for existence, but by a first evaluation of this information, it created for man, even in prehistoric times, the position that singled him out from all the other living creatures.

Since then, man has constantly strengthened this position by acquiring, storing and processing, just as producing, information, and has continuously increased the distance between him and other living creatures. In the course of the development taking place in this manner, man's brain has probably crossed the "threshold of complexity" that enables a data-processing system to programme itself.

Why the brain can perform more than its fundamental function, to ensure physical survival, and why it can, in particular, accomplish tasks for which it was probably not created - all this is still largely unexplored. It would appear that, parallel with the growth in number of active functional elements and their interconnections, the capacity of the brain to accomplish more than was originally assigned to it is raised to a higher power.

If this is so, the increasing complexity of our brain leads to a constant increase in efficiency and an ability to undertake tasks which no longer have anything to do with its original functions. Through the described absorbing, processing, and producing of information, the brain is able to give rise to ideas, literary composition, and insights into the laws of nature. Culture and civilization that have developed in this way have allowed the network of communicative links between people to become ever closer, the brain ever more complex, until it has achieved an organized life of its own.

To a decisive extent, the human community, the coexistence of individuals, families and nations, is influenced by the continuous growth of information which comes streaming in upon the individual. Since the competition of inter-communicating brains is always on the increase, always better-informed brains are participating and more and more media and channels of information are being opened up, man increasingly feels that the information that comes streaming in upon him is not a vital aid but a burden.

The individual must process this information in order to be able to hold his own in the struggle for existence, no longer in the animal kingdom but in the rivalry of inter-communicating brains. The added efficiency of the brain acquired step by step in the course of biogeny, is becoming a vital necessity. Consistent with this development, the amount and quality of the absorbed information and its processing have become decisive for the professional and social performance of the individual, just as they are for the performance and international standing of States.

"Information is the beginning and the foundation of society" is the opening phrase of one of Karl Steinbuch's principal works*, and Norbert Wiener, to whom we owe cybernetics, has said: "Information is information, not matter or energy. No materialism which does not admit this can survive at the present day." This means that information is accorded the same status as matter and energy.

If this is so, there is a public interest in ensuring that the vital information achieves its object and can be made useful for the community; and this all the more since the volume of vital information has also made considerably greater demands on the brain than in former times. Not only is human knowledge as a whole rapidly increasing, but the individual must know more, and accordingly learn more, than

* Steinbuch, Karl, Zukunftsforschung. Die "informierte Gesellschaft" von morgen. Aral Journal (1966) Nr. 55, S. 8-11.

formerly. Studies are being made for this, such as the one undertaken by a research team at the University of Wisconsin, according to which, about the year 1880 a person practising a profession had to acquire approximately as much knowledge in his later professional life as he had acquired at school; today, on the other hand, the knowledge to be subsequently acquired amounts to something like ten times as much as has been learned at school.

This requirement affects a brain that is already heavily burdened by information which is not part of the knowledge referred to, but is merely a by-product of modern society; traffic noise, radio noise, visual propaganda and so on. This imposes obligations on the politicians: they have not only to guarantee the flow of information to those in need of it, but to secure the readiness and capacity to absorb, through measures to counteract the irritating optical and acoustic overburdening. Information not absorbed, even though sensibly directed, cannot achieve its object or attain its full productive value. This makes clear how closely the information problems are related to the problems of environment and why legislative measures in the sphere of protection of the environment must include information policy.

It would be fascinating to show how in the course of history people have planned their mutual relationship and have tried to cope with their problems in a hostile environment by translating the information at their disposal into action.

In doing so, information and man's natural instincts have frequently come into conflict. It seemed perfectly natural that refuse should be consigned to waters, particularly those that are flowing, which carry it away and cause it to dissolve. For centuries this practice was in keeping with the information that had been acquired and which taught that even glass and pottery would be gradually ground down, and organic refuse would nourish the marine life when administered in small doses.

Recent findings are diametrically different. The introduction of industrial waste is killing the marine life, detergents are turning the bottom of the waters into stone, plastic garbage is not being dissolved, and the over-nourishment due to organic refuse of a greatly increased human population can no longer be processed; the waters are becoming foul and "tilting over," as the technical expression says. This contamination of the waters causes disease, fouls drinking water, destroys holiday and recreation areas. At the same time the measures that have been, or are to be introduced on the strength of this improved information demonstrate impressively the obligation of the State to use the information at its disposal to conserve the health of the nation.

A particularly impressive example of overlapping information is provided by road traffic. Information Phase I was obtained at a time when the motor-car merely took the place of the horse-carriage and it was noted with pleasure that the car was quicker and left no visible excrement behind; the invisible excrement from the exhaust still formed no part of the information. This was also the case with Information Phase II, when it was seen that in the narrow, winding streets of the towns the car had lost the advantage of speed, and - again, information translated into action - it was decided to regulate the flow of more and more cars into the city centers via ever more and wider streets.

The solution of these material traffic problems was predominant for something like half a century, almost up to 1970, until Information Phase III was reached. Now, the keywords were smog, air pollution, and lead discharged from the exhaust. Hundreds of substances from the exhaust gases were found to be carcinogenous, and suspected of causing other diseases. This information was so reliable that it was made the basis of drastic measures such as the banning of motor traffic from whole city centers - once more an example of how it is the task of the State to procure the latest trustworthy information and translate it into legislative measures.

At the same time this example shows how long it took for information about the jeopardizing of the national health to become available, gain in weight and be taken seriously. It is a result of the impression left behind by the imposing achievements of industrialization and mechanization that for decades "information" formed the basis for the solution of material problems, whereas information about man was initially hardly recorded and utilized.

For decades, information has been considered from scientific and technical aspects. OECD's scientific and technical information policy was wholly dominated by this, and interest was centered on information systems in fields like chemistry, medicine, geoscience and others. This has undergone a rapid change. If OECD now calls information "the key to the wise management of the future," it is thinking - as FID and UNESCO have done for years - of information as a discipline which embraces social, economic, juridical and political aspects just as it does scientific and technical. The demand, therefore, is to ensure that, no matter by whom national or international decisions are made - supranational bodies, governments or parliaments, lower authorities, industry, or other bodies - the people who have to make the decisions have the whole spectrum of information at their disposal so that they are able to estimate what the effects of their decisions will be.

This introduces a wholly new element into the utilization of information. I have permitted myself, Ladies and Gentlemen, to cite two examples to attract your attention to the problems of environment: water contamination and air pollution. They are not intended to be technical examples, but an expression of the growing anxiety as to whether the application of technology in the form so far practiced will redound to the blessing of mankind. We can frequently estimate the technical results of the utilization of technical innovations, their social effects almost never, because the understanding of the social repercussions of this information lags behind the development of the information itself.

I hope that no one will conclude from my remarks that the scientific, technical or industrial information can be neglected in favor of the study of their social repercussions; rather, only in the interplay of all factors will a network of information embracing all fields have the greatest effect. Up to now, however, the information network is still bogged down in planning, and neither the well organized supply of social nor of technical essential data measures up to the needs of the times. Above all, however, what is lacking is the working out of a scientific basis for the discipline; indeed the lack is so great that at times the justification of the existence of a separate field of "information science" has been questioned.

This fact alone must lend great importance to a conference which has made its goal the intensification of scientific examination of information problems. We need not, however, leave it at this general allusion, since the present congress programme not only displays the exceptional capacity of the host country to organize such an event, but it is in itself a document of the services which the international scientific and technical community can contribute to this branch of knowledge.

Keynote Address

Since I have been entrusted with the honour of speaking at the opening of this conference in the name of FID, the International Federation for Documentation, which tries to serve the needs of the scientific community, may I make myself the spokesman of all who are assembled here and not least of those who are joining in this event from a distance. In the name of all, I heartily congratulate the organizers. We are convinced that ISLIC 1971 will bring us a big step forward in a field, of which a great international organization has said: "The information revolution has only just begun."

SESSION ONE

international and national information networks

Authors of Papers

WORLD SCIENCE INFORMATION SYSTEM
UNISIST

Adam Wysocki
Director
Division of Scientific Documentation and Information, UNESCO

SYNOPSIS

UNISIST is planned as a continuing, flexible programme to coordinate existing trends towards cooperation and act as catalyst for the necessary developments in scientific information. The 22 recommendations of the study report include the following 6 groups: (a) tools of system interconnection; (b) effectiveness of information services; (c) responsibilities of professional groups; (d) institutional environment; (e) international assistance to developing countries; (f) organization of UNISIST.

HISTORICAL BACKGROUND

UNISIST is the name which has been given to a study on the feasibility of a world science information system undertaken jointly by Unesco and ICSU.

In 1966, at its 11th Assembly in Bombay, the International Council of Scientific Unions (ICSU) approved the creation of a committee to examine the feasibility of a world scientific information system based upon achieving compatibility among existing and prospective programmes related to the storage and retrieval of scientific information.

At the same time, Unesco had been thinking along parallel lines and had developed preliminary plans for the holding of an international conference on scientific information. The President of ICSU proposed to the Director General of Unesco in April 1966 that the two international bodies, the one non-governmental, the other inter-governmental, jointly study the feasibility of a world science information system based on the achievement of a higher level of cooperation among the existing and the developing services.

A Working Party met at Unesco House in Paris in January 1967 to develop the guidelines for the study. It was emphasized that any system to be developed should be a flexible network based upon the voluntary cooperation of existing and future information services. It should be limited at first to the natural sciences of concern to the member unions of ICSU. It should pay special attention to the growing needs of scientists in developing countries. It should also pay attention to the problems of standardization, of selectivity and compaction, and should closely involve the participation of scientists themselves.

Under these guidelines, a Central Committee was formed to conduct the study. It was composed of specialists from different countries and chaired by Professor H. Brown, Foreign Secretary of the National Academy of Science of the U.S.A.

To explore avenues which the Central Committee considered significant, Working Groups chaired by members of the Central Committee were established in such areas as bibliographic standards (jointly with ICSU/AB), linguistic problems, communication formats, scientific information in developing countries, research needs, compaction and evaluation, and classification. The Working Groups have produced reports which have all been received as background material for the final report.

In addition, for the dual purpose of insuring practicability in its recommendations and of keeping systems operators informed, an Advisory Panel was appointed with membership from the existing information services in the sciences who has embarked on and gained experience in modern systems operation. This group was enlarged to include both the International Federation for Documentation (FID) and ICSU Abstracting Board (ICSU/AB).

The feasibility study has now been completed and the Unesco-ICSU Central Committee has reached the unanimous conclusion that a world science information system, considered as a flexible network based on the voluntary cooperation of existing information services, is both necessary and feasible.

The results of the four-year inquiry into the problem of improving the world-wide transfer of scientific and technical information are contained in the Unesco/ICSU Report on the Feasibility of a World Science Information System (UNISIST) which has also been published in an abridged version: Synopsis of the Unesco/ICSU Study Report on the Feasibility of a World Science Information System (UNISIST). +)

During the course of the study, a considerable amount of documentation was compiled. There were the reports of the Working Groups, special papers and studies voluntarily contributed or prepared under contract, the working papers of the Central Committee and of its Advisory Panel. These papers have been brought together and will be reproduced in microfiche form for deposit in the principal research libraries of the world.

What is UNISIST?

According to the Study Report, UNISIST is planned as a continuing flexible programme to coordinate existing trends towards cooperation and to act as a catalyst for the necessary developments in scientific information. The ultimate goal is the establishment of a flexible and loosely connected network of information services based on voluntary cooperation. UNISIST is to be concerned initially with the sciences, applied sciences, engineering and technology, but it will later be extended to other fields of learning.

UNISIST has at one and the same time three different aspects. First, UNISIST is a concept and a philosophy, a restatement in modern systems terms of the traditional principles of international scientific communication. Second, UNISIST is a movement and an effort to formalize, give stability and recognition to, and expedite trends towards cooperation and system compatibility which are already in existence. Third, UNISIST is viewed as the future organization which, housed in Unesco, will function to initiate and catalyse activity designed to further these trends.

+) Both publications are available from Unesco in English, French, Russian and Spanish.

I International and National Information Networks

UNISIST is not conceived as a rigid, predesigned superstructure which will provide the world scientific community with all available information in science and technology. In fact, the UNISIST Study Report does not contain a systems analysis of the present state and relationship of scientific information systems in the world. A more pragmatic approach has been taken by considering the feasibility of the system from a technical, as well as from a political and economic point of view.

UNISIST is an international effort to synthesize a diversity of philosophies, programmes and policies that relate to the free flow of scientific and technical information. It is an unprecedented attempt to stabilize and coordinate existing trends towards international cooperation in the communication of scientific information.

The broad principles on which the world science information system is based are enumerated in the preface of the UNISIST Study Report. The proposed system stands for the unimpeded exchange of published scientific information and data among scientists of the world; hospitality to the diversity of disciplines and fields of science and technology; promotion of compatibility, cooperative agreements and interchange of published information among the systems; cooperative development and maintenance of technical standards to facilitate the interchange; development of trained manpower and information resources in all countries; increased participation of the present and coming generation of scientists in the development and use of information systems; reduction of administrative and legal barriers to the flow of scientific information in the world and assistance to countries which seek access to present and future information services in the sciences.

The Study Report is a highly informative document in that it contains a great deal of specific information and is an impressive synthesis of the complex issues. It contains twenty-two recommendations for programme development, directed toward three groups of individuals: 1) government officials who are responsible for national science policy; 2) the world community of scientists; and 3) professional practitioners in the many different pursuits involved in scientific communication.

The three publics to whom the recommendations in the Study Report are addressed have common cause but different roles. The scientists are creators and users of information; information professionals organize, process and service information; governments support the development of these services in the national public interest.

UNISIST Recommendations

Twenty-two recommendations concerned with the programme development and the initial management provisions for UNISIST are presented in six groups dealing respectively with:

a) Tools of Systems Interconnexion

This group of recommendations, technically oriented, calls for the establishment of permanent records of information resources as a prerequisite to the information sharing policies. It also calls for the establishment and application of international standards for bibliographic descriptions with corresponding machine codes and formats as well as for normalized vocabularies and thesauri.

One of the recommendations of this group entails the creation of an international register of scientific periodicals to standardize the citation of journal literature in science and technology.

The attention of information specialists and scientists is drawn to the need for joint efforts in developing better tools for the control and conversion of natural and indexing languages. The initiation of a few pilot projects in this field is encouraged.

Two recommendations of the group deal with technical requirements for system compatibility and the potential use of telecommunication and teleprocessing networks for the transfer of scientific information.

b) Effectiveness of Information Services

A series of recommendations relates to the effectiveness of information services, libraries, abstracting services, information analysis centres and data evaluation centres for which various promotional programmes are suggested. These recommendations call for continued recognition of the scientific library system as an essential component of scientific information transfer in modern times, for encouragement of cooperative schemes among abstracting and indexing services in science, for the development of specialized information analysis centres, for the evaluation and synthesis of published information and for the further development of numerical data centres.

c) Responsibilities of Professional Groups

The recommendations concerning the responsibilities of professional groups are addressed to authors, editors, publishers, scientists and information specialists, with special emphasis on the need to stimulate and coordinate existing efforts in matters of information science, education and research.

These recommendations call also for improvement in the quality of scientific journals and the involvement of international and national scientific societies in the world-wide information transfer. As regards further research in information science, UNISIST does not advocate initiating another research programme but rather sponsor a small group to collect and evaluate ongoing research in information and library sciences.

d) Institutional Environment

This group of recommendations deals with policy-oriented issues and indicates guidelines which governments should follow in order to create a suitable institutional environment for the cooperative handling of scientific and technical information throughout the world.

It is proposed that each country should establish a governmental agency at the national level to guide, stimulate and conduct the development of information resources and that these agencies should adhere to the principles and goals of UNISIST.

Governments are also encouraged to interlink information services in science and technology in national networks and to provide adequate equipment and manpower for a speedy processing of scientific information.

One recommendation in this group suggests that the pricing policies of individual operating information services should be studied to observe their effects on user access to scientific information.

Another recommendation calls for reducing legal barriers, such as too restrictive national copyright laws, in information transfer.

e) International Assistance to Developing Countries

Two recommendations deal with the need of developing countries for special assistance programmes and are addressed to governments and international organizations.

The first is concerned with attaining the minimal infra-structure which a country must have in order to participate in the benefits of an international network.

The second calls on the industrialized countries to conduct, through UNISIST, a number of pilot projects to link developing countries with UNISIST.

f) Organization of UNISIST

The last recommendation sets forth the organizational requirements of UNISIST. The suggested organization should be provided with three interrelated managerial bodies: (i) an Intergovernmental Conference responsible for approving the programmes of UNISIST and reporting on their progress; (ii) an International Scientific Advisory Committee, with a strong representation of ICSU and member unions, as well as library information experts and services, responsible for assessing progress in communication practices and changes in user requirements, as a basis for and as a result of UNISIST programmes; (iii) an Executive Office, serving as the permanent secretariat of UNISIST, responsible for preparing and administering programmes and budgets. The recommendation proposes that this last body be placed in the administrative structure of Unesco, within the Science Sector.

I International and National Information Networks

Scope and Development of UNISIST

The programme recommendations briefly outlined above reflect the two broad functions of UNISIST: a) a catalyst function by which UNISIST is to stimulate international cooperative agreements among autonomous information systems, and b) an initiating function by which UNISIST will encourage new projects designed to improve the world information tools and resources. These functions will involve interaction with representatives of the governments, the scientific community and the information science specialists.

The success of the project will depend largely on the benefits which adherents may receive from UNISIST. The final part of the report is devoted to this topic. It is stressed first that the notion of benefits should not be understood in strictly financial terms: money-saving is only one of the possible motivations for establishing UNISIST; besides, it may accrue only after an initial period in which further expenditure will be needed, to reach the proposed standards of performance. Nevertheless, there are savings to be reaped in the long run from a redistribution of labour at world-wide level, leading to an improved systems economy. The resulting information network should provide a more effective tool for individual scientists, as well as for policy-makers in science, in the determination of research programmes; further benefits may thus be expected from an increased productivity of science. Finally, the principle of world-wide interconnexions, if properly understood, implies that present inequalities in the distribution of information resources over the earth may be reduced which would bring special benefits to developing countries.

Intergovernmental Conference and National Commitments for Implementation of UNISIST

The recommendations and proposed programmes of the Study Report will be submitted to an Intergovernmental Conference for the Establishment of a World Science Information System (UNISIST) to be held at Unesco House, Paris, from 4 - 9 October 1971. This Conference, convened by the Director-General of Unesco and prepared jointly with the International Council of Scientific Unions, will bring together government officials responsible for the planning and direction of science information programmes, scientists and science information specialists. It will be expected to make recommendations concerning the mechanism and procedure by which Member States and International Organizations could play an active role in the implementation of the system.

Through their endorsement of the recommendations of the UNISIST report at the Intergovernmental Conference, governments will have made a commitment to support the implementation of UNISIST. This basic commitment has two aspects: 1) a commitment to follow the courses of action proposed in the UNISIST recommendations, as endorsed by the Conference, 2) a commitment to support the proposed administrative structure within the context of the Unesco Programme and Budget.

The principal objectives and activities to which governments, in cooperation with professional organizations, could commit themselves through their endorsement of UNISIST are:

- participation in the development, promulgation, and voluntary adherence to international standards for information transfer (e.g. bibliographic descriptions, formats, codes, machine interface, telecommunications);

- participation in the cooperative development of specified features of the UNISIST programme (e.g. the World Register of Scientific Periodicals);

- the development of national scientific and technical information systems and services in such a manner as to optimize compatibility with other services and systems, thereby to work cooperatively toward an international network;

- efforts to facilitate the international transfer of scientific and technical information, including actions to reduce administrative and other impediments to the interchange;

- the pursuit of policies and courses which will lead to a sharing of the costs of processing of scientific and technical information, as well as to a sharing of the products.

COMPETITION OR COOPERATION AMONG INTERNATIONAL SCIENTIFIC AND TECHNICAL INFORMATION ACTIVITIES

Burton W. Adkinson
Director, American Geographical Society

Introduction

International treaties, conventions and agreements in practically any field of endeavor include or create an information exchange requirement. This is true of such pacts whether they deal with control of diseases, reporting on weather, control of fishing, traffic or commerce, or are international cooperative research programs such as International Geophysical Year (IGY), International Decade of Ocean Exploration (IDOE), or the International Biological Program. Yet the initiations and operations of these international activities have as their purpose something other than information exchange. Nevertheless, all of these activities generate information and require mechanisms to effect good communications for information transfer. However, most cooperative international research and other programs seldom include in the plans, procedures, and programs for the orderly dissemination of the generated information. The information is allowed to flow from these projects in a haphazard fashion, and all too frequently the individual or organization which generates the new findings is allowed to handle the information according to personal preference with little thought given to ways to make available the findings in a useful and effective manner.

Yet, on the other hand, one finds individuals and organizations that have developed national and international plans and projects to gain control of large bodies of information. For instance, history is replete with abortive attempts of man to acquire and organize the world's fund of information. They range from creation of the Great Library at Alexandria, through efforts to develop a bibliography of the world's literature in Belgium prior to World War II, to the suggestions by a Pug Wash Conference, or Hans Roeter's present activity to have United Nations create a center which would have the world's scientific information controlled by a computer with a sophisticated dissemination system.

Development of International Information Systems

This paper will describe a few ongoing efforts to develop, through international cooperation, more effective communication and information systems and to suggest some points that should be considered as the national and international information systems evolve.

The activities of OECD and the ICSU-UNESCO Joint Committee (UNISIST) have been described in many publications. Therefore, this paper will not outline these organizations' programs. In addition, there are many other international cooperative programs for scientific and technical communications, a few of these will be briefly described.

World Meteorological Organization (WMO) has a program for rapid exchange of Meteorological Information which includes:

(a) The standardization of observation methods, instrumentation and methods of reporting.

(b) A network of meteorological and weather stations that allows for regional, national and continental collection and dissemination of information.

(c) A system known as the World Weather Watch which includes use of satellites and telecommunication.

The International Atomic Energy Agency (IAEA) is developing an International Information System (INIS) based on:

Input responsibility by participating member countries; processing responsibility of the international agency; and output to participating member countries in magnetic tape, micro-image or conventional forms.

COMECON. Eight Eastern European countries, including the U.S.S.R., have established an International Center for Scientific and Technical Information in Moscow. Regional branch offices may be established in member countries. Input is by member countries to the Center which will service countries. These countries have agreed on standards of input and output. At present the system is based on manual methods. The International Center also is attempting to collect materials from non-member countries.

ICIREPAT (International Committee for Information Retrieval for Patents). Major patent offices of the Berne and Universal Copyright conventions have come to agreement on minimum uniform standards which include information exchange on tape and microfilm, as well as adoption of a standard classification for patents. The United International Bureau for the Protection of Intellectual Property is the Secretariat for this endeavor.

ICSU/AB (International Council of Scientific Unions Abstracting Board) has developed a plan for cooperative exchange of abstracts among some of its members, and is developing a plan to rationalize coverage of journals peripheral to the core journals for each service.

In addition, in cooperation with the UNISIST Central Committee, ICSU/AB is developing agreements on handling citations, common formats for exchange of information on tapes, and in some fields of science is developing multilingual thesaurus and rationalizing the classification of materials in data banks.

All the above steps are considered necessary for an evolution toward an international network of information services.

UNISIST-International Serials Data Program. One of the outgrowths of UNISIST consideration has been progress toward compiling a world inventory of scientific and technical serials. The UNISIST Central Committee commissioned a feasibility study and preliminary plan for inventorying and numbering all serials. As a result of the report of this study an ad hoc group was established to consider steps to implement an international serials data program. Also, the French Government, through the Ministry of Education, offered several hundred thousand francs to support the initiation of the program.

The ad hoc group has recommended to the French Government and UNESCO that:

1. There be an International Governing Board of 12 representatives from 8 countries and 4 international organizations.

2. There be an operating center under the control of a Board of Governors.

3. There be an international technical advisory committee.

The ad hoc group urged that priority be given to identifying and numbering active serials, preferrably in the sciences and technology. It is expected that this international serials program will get under way this year.

International Book Numbering System. In contrast to the preliminary planning for an international serials data system, there is now in being an international system for numbering books. An international office for this activity is located at R. R. Bowker Company in New York. The United States, United Kingdom, France and most of the major book publishing countries are now participating in this endeavor.

CODATA is a cooperative program for obtaining and disseminating critically evaluated data in the physical sciences. This international program was initiated by ICSU. Its secretariat is located in Frankfurt, Germany.

One would be remiss in not mentioning the well-known efforts of the International Federation for Documentation's activities in coordinating the Universal Decimal Classification and Building Research Documentation, or the International Federation of Library Associations' work in developing an international descriptive cataloging code and its catalyzing efforts to develop national bibliographies.

The above are but a few of the developing international programs that are being fostered by international organizations. In addition, a number of national organizations have initiated systems development programs with an international character. Among these are:

The American Chemical Society has agreements with the United Kingdom, Germany, Sweden, The Netherlands, and Denmark whereby these countries are participating as output or input centers or both. The Society plans to expand this international cooperative program.

The American Institute of Physics is negotiating with United Kingdom organizations for a joint effort in physics and astronomy. Also, the American Society for Metals has a cooperative agreement with the United Kingdom Iron and Steel Institute for the production of an abstract bulletin.

Among the United States Federal agencies, are the National Library of Medicine's (NLM) efforts with other countries for the internationalization of Medlars; The Library of Congress' Marc II System is being adopted by other countries, and NASA and ESRO (European Space Research Organization) have a cooperative agreement for the use of Star and Recon Systems.

One should also mention the British effort called "Inspec" whereby the Institute of Electrical Engineers has created a bibliographic file on magnetic tape of citations to physics and electrical engineering publications, and has offered to cooperate with national organizations of other countries in the improvement and use of this file.

The efforts of the Commonwealth Agriculture Bureau in producing abstracts and indexes on agricultural science and technology are one of the older international cooperative ventures. Marked technical changes are now being introduced into this service. West Germany, the United Kingdom, and the United States have cooperated to initiate an information on food science and technology.

These are illustrative of the evolving information systems which are becoming international in character.

Conclusion

What are the implications of these national and international efforts?

First, it is apparent that effective scientific and technical information systems must have an international base. Development of new science and technology is no longer the prerogative of one or even a few countries. Therefore, information must be communicated across national boundaries. Also, no country today can afford to collect, organize, process and disseminate the total world fund of scientific and technical information. Economics, as well as other factors, dictate cooperative efforts which are essential.

Second, there is a growing realization that there is need for rationalization of techniques for handling information elements such as a common form of bibliographic citations, and recording data elements on tape or microfilms, etc., for exchange purposes.

Third, there is need for an improvement in mechanisms for fostering coordination and cooperation within and among information systems organizations. Often the most difficult part of systems development is the adjustment of organizations to new requirements. This leads to the next point.

Fourth, and one of the most important considerations, is that increased effort must be applied to the evolutionary development of organizational structures that can foster, develop and operate large international information systems.

For example, it is pretty well agreed that an international serials control program would be of real benefit to every country. How can the various organizational components such as authors, publishers, abstracting and indexing services, libraries and user groups, be welded into an effective international serials program? Each of these is an important element of the program, but each must adjust to new requirements and be willing to forego some freedom of action.

Fifth, there is need for increased effort toward facilitating the participation of less technically advanced countries. Even if these countries could financially afford the newer systems, they frequently lack the intellectual, technical or organizational bases for effective use of the large complex information systems.

Sixth, how can effective educational programs be developed so that users, operators and producers can effectively use these complicated and expensive tools?

Finally, no realistic economic plan has been devised to develop and support these new systems that will allow user access to information products and services at costs that do not necessarily discriminate against the less affluent users.

It appears to this author that there is an increasing need to develop national and international focal points that can guide the development of these various efforts toward a realistic network of national and international information systems.

It is suggested that the endeavors of such organizations as OECD, UNISIST, COMECON, ICSU, ISO, FID, IFLA, and UNESCO must be encouraged to redouble their efforts to foster cooperation and coordination. Also national efforts, both governmental and private, must give serious attention to assisting in the furtherance of the evolution toward an international integrated information network.

If the pace toward integration isn't accelerated, I predict the information systems of the future would make the Tower of Babel sound like a melodious chorus.

INFORMAL INFORMATION

Rolf Gezelius
Chief Research Engineer, Research Institute of the Swedish National Defence, Stockholm

SYNOPSIS

The total scientific information system in for instance a research organization is composed of two main parts, formal and informal information. The formal part is connected to libraries and documentation centres, the informal part is mainly dependent on interaction and contacts between the scientific users of information. In the paper examples are given of informal information, a comparison is made between formal and informal information and also a survey of the origin of informal information and its conversion to formal information.

Background

The aspects and ideas given in this paper are based on experiences from the Electronics Department of the Research Institute of the Swedish National Defence, FOA. This research institute has several hundreds of active scientists and the research is mainly directed towards applied research and system development but a considerable amount of basic research is also performed. Talks and interviews with scientists mainly in the field of electronics give the background of this paper.

As in most research organizations the scientific work and the scientific results are to a great extent depending on a positive information environment and on an effective total scientific information system. We have found that this total information system is composed by two main parts.

1. The formal information disseminated from libraries and documentation centres.

2. The informal information which most scientists acquire, often by personal contacts.

A very interesting thing is that on request many scientists after long cogitation say that they do consider the informal information to be of the same order of importance as the formal one. They say that they very roughly get half their scientific information by informal channels. In science and technology this matter has hitherto hardly been observed. In psychology, however, studies have been made (1) which indicate that in this field informal channels are used to a great extent. Otherwise informal information is seldom dealt with in literature (2) and such an organization as FID seems to be very little interested in this problem.

The purpose of this paper is to direct attention to this important part of the total information system. It must be pointed out that the informal information dissemination almost completely is to be found outside the conventional documentation field. It is still not much influenced of the very large efforts made, for instance by computers, in order to improve the formal part of the information system.

Examples of Formal Information Dissemination

The formal scientific information field could also be termed library-connected information or documentation. In the electronics department of FOA the operative dissemination is handled by a special documentation centre, and it is divided in about twenty channels. The most important of these channels can be summarized in the following way.

Primary Literature

Journals
Books
Reports (public)
Special translations

Secondary Literature

Abstract journals
Computerized dissemination of internal literature (under development)
Computerized dissemination of external literature (three channels)
Special abstracts (internal production)
Shortened articles (internal production)

Our system is mainly used for current dissemination of information but also for retrospective literature search.

Secondary literature is not so much used in electronics as for instance in chemistry. If one set a figure for the importance of secondary literature relatively to all literature, one could estimate this figure at 5 - 10 % in electronics and perhaps at 20 - 30 % in chemistry. These figures are more a guess than a proved fact. However, because computerized dissemination of information is a kind of secondary literature, the interest of such information is not very large in our organization. Perhaps it may change in the future. By far most popular of secondary literature is the shortened article service, which is rather close to primary literature. This kind of scientific information is indeed very appreciated by the users for its quality and selectivity.

As to scientific journals we have the very simple principle that every scientist or group of scientists are subscribing directly to the 5 - 15 journals most important for them. The journals go directly from the publishers to the scientists without passing the documentation centre and the scientists can keep the journals for the future and consider them as their property. We have found this uncomplicated system to be much cheaper and much more effective than any journal circulation system. In our country journals are cheap and salaries for librarians are rather high. Only for journals of secondary interest we use a limited circulation system.

Informal Information Channels

As mentioned earlier the informal information channels seem to be very important for the scientists in our research organization. Some typical examples are
Secret information (not dealt with in this paper)
Internal reports
Reports with limited dissemination
Non-formalized information (e g notes and memoranda)
Fragmentary information
Personal information (e g conversation, visits)
Closed information systems (e g information exchange groups)
"Paperless" information

It is very difficult to make a complete and reasonable systematization of this field. The conceptions are not commensurable and the limit between formal and informal information may sometimes be vague.

A preliminary inquiry has been made with about ten scientists and it will be further intensified and widened. Preliminary and on the average it seems that very roughly half the information need will be satisfied by formal channels and the other half by informal channels. This is in good agreement with the statements in (1).

Sociological Aspects

When speaking to people about informal information it becomes obvious that analysing this field is not mainly a problem of information science but to a great extent involves sociological and psychological problems. Of special interest are the contact channels between the scientists, because these contacts very often deal with informal scientific information. A sociological study of among other things the contacts of the FOA scientists has been performed by Swedish sociologists (3, 4). This study shows that young scientists commonly have fewer contacts than their elder colleagues. A more unexpected result is that the shorter time a scientists has been a head of research, the larger number of contacts does he have inside his own organization. The longer time he has been a head of research, the more his contact pattern is developed from a large number of sporadic contacts to a smaller number of frequent contacts. A high degree of education will lead to rather few contacts inside the own research group but to many contacts with scientists outside the group.

When speaking about sociology we have experienced that many people expect or demand information but not so many ask for it. This somewhat paradoxical statement means that all agree that information as a conception in general is useful and necessary but not even scientists ask very often by themselves for information. They must sometimes be initiated in some way. In connection with this, one can notice that retrospective literature search is much less used than current dissemination of information.

Very interesting are also some statements from the U.S. that during the last twenty years a decrease has been observed in the read/write ratio of authors of technical articles and papers. This ratio is defined as the quotient of the number of pages, which an author is reading to the number he is writing during the same time, e g one year. Thus the authors are said to know less now than before about the subjects they write about. This conclusion, however, could be wrong. It might be possible that the deficit of formal information is compensated by an increasing informal information, but in fact this is only a guess.

Of a great interest would be to make a sociological study of the motives of scientists to publish their results. Documentalists are often of the opinion that only the valuable results are published in journals or disseminated by other formal channels. There are in fact reasons to doubt that this is always the case. It is clear that many scientists are publishing articles more in order to qualify themselves than to give valuable information to other people.

Comparison between Formal and Informal Information

Sometimes it is difficult to decide whether scientific information should be considered as formal or informal but several typical differences can be observed.

1. Formal information can be stored and more or less easily retrieved. Informal information often will be lost in a short time.

2. Formal information is in principle easy to survey and check, informal information could be difficult in this respect.

3. Formal information is often not very concentrated but mixed with irrelevant information. The noise level can be high. Informal information is often more concentrated.

4. Informal information is often newer and more actual than formal information.

5. The formal channels in the total information system are rather few but cheap to use and accessible to everyone. They have often a large audience. Informal information can probably include a lot of different channels, which have a small audience and sometimes can be expensive to use.

6. Informal channels generally give more selective information than formal channels.

7. Formal channels are controlled of documentalists, in many cases, however, after feedback from the scientific users. Informal channels are more or less directly controlled of the users and in most cases they are out of control of documentalists.

8. Informal information is by nature often exclusive, formal information can in principle be disseminated in an unlimited way.

Both kinds of channels are usually indispensable in a complete information system. Normally they seem to balance each other. If the balance is disturbed, the users normally take necessary steps. Informal channels are developed and controlled mainly of the scientists, who function both as producers, distributors and users of information. Informal channels are often characterized by direct contact by correspodance or face-to-face.

It is often difficult to estimate the relevancy of formal information for one's own work. Such computer retrieved information, which gives only the titles of the articles, is often insignificant and gives lots of noise. Better are lists with titles and abstracts, but much better as to the question of relevancy is the informal or personal information, which makes it easy for a scientists to see if he and his colleagues are working on the same problems. Scientists, who act informally, speak more freely about their work and plans, progress and mistakes. The informal information is to a large extent connected to interaction between scientists.

An interesting feature of the informal information is that it is often more or less exclusive. For society and mankind as a whole it is probably most suitable if information is disseminated to everyone who can use it, but for the private scientists it could often be of interest, especially from an economical point of view, to keep valuable information inside a small team. By patent protection it has in technology been possible, at least in principle, to join the exclusive right of an invention with the demand of society on information dissemination.

The Origin of Informal Information and its Conversion to Formal Information

A common reason to establish informal information is that the formal channels do not correspond to the actual demands of the scientists. The ground of the discontent can be too long publishing time for articles or reports, bad selectivity, bad quality and too much noise in the formal channels. They may also be unable to follow the front of some special scientific field under rapid development.

From various causes informal information channels are established, usually by scientists themselves. Such channels may be shaped as closed information exchange groups between scientists in the same organization, in the same country or internationally. Such closed groups may be of value for the members but of annoyance for the outsiders. Another common way to establish an informal channel is when a research team in an advanced field starts a small informal bulletin for actual information in that special field. This bulletin will be disseminated just to a few scientists and will not be abstracted in the conventional or the computerized secondary literature. A third type of informal channels may arise when a research team is arranging an informal symposium in a new field with potentially promising development possibilities. The symposium becomes a success, will be repeated some times and after some years it may have developed to a formal periodical congress with hundreds of participants.

This gives an idea of how the formal channels can arise, especially if they give information in a field, which is developing rapidly. Informal channels can increase in importance and by and by develop to formal channels with their advantages and disadvantages. The closed information exchange group can develop into a scientific association, the small bulletin can grow to a large journal and the informal symposium can change to a periodical congress. In the total scientific information system there is always a permanent renewing of information channels and this renewing goes almost ever in the same direction, from the formation of small channels through their growth and conversion to large formal channels. Further ideas about the conversion of informal to formal information are presented in (5).

Are Informal Channels Worth Studying?

Many documentalists may consider the aspects and opinions presented in this paper to be very simple and self-evident. This is quite right. The importance of personal contacts and other informal channels as a part of the total information system is for most scientists so habituated and self-evident that they seldom, until thinking over the matter carefully, realize how important it is. To the scientists in our organization it has been a new experience to realize that the informal part of the total information system is about as important as the formal one. In another scientific field, the psychological, an estimate has given roughly the same result. An interesting sociological object would be to make similar investigations in other parts of science in order to elucidate the relative importance of informal versus formal information.

When finding the informal information to be important in science in common, the next question would be if there is anything at all to do for information science in this field, because everything seems to be quite all right with informal dissemination of information. Special subventions are seldom used in this field, compared to the very large government and other subventions used to disseminate formal information, e g by computers. The informal dissemination seems to manage itself very successfully. In some way this is the triumph of uncomplicated relations over a development of increasing complexity. One might believe that the informal dissemination is functioning so very good just because it never has been the object of any external measures.

A cautious investigation without disturbing interference of the informal flow of information could perhaps give interesting results how information really is transferred in some typical cases. It would also be of value to see how the flow of informal information will be influenced by an increasing computerization of information dissemination. If in such a case the informal information increases, it must be explained as a negative attitude of the scientists to computerized service. If the informal flow will decrease, one must consider that the computerization makes the formal channels more attractive. In general one might use the amount of informal information in an information system as a measure of the effectivity of the formal part of the system.

Another thing we might learn of the study of informal information is that good contacts give valuable information and this in turn initiates progressive research on high level. Investigations in our organization have shown that young scientists do not generally have so good contacts as the elder ones. Every young scientist must understand how important it is for him to build up a personal contact network as rapidly as possible. This declaration will certainly be valid also in a future of highly mechanized information dissemination.

References

1. Garvey W D, Griffith B C: Scientific Communication as a Social System. Science 157 (1 sept 1967), 1011 - 1016.

2. Graham W et al: Exploration of Oral/Informal Technical Communications Behavior. American Institutes for Research, Silver Spring, Maryland, USA, 1967. Report AD 669 586

3. Swedner H, Persson R: Forskarmiljöer på FOA (Scientist Environments in FOA) (In Swedish) FOA P Report C 8170 - 10. Research Institute of National Defence, Stockholm 1967.

4. Ljugh B: Forskningsledare och forskningsorganisation (Heads of Research and Research Organization) (In Swedish). Memorandum no 28. Committee of Research Organization and Research Economy, Stockholm 1967.

5. Garvey W D, Lin N, Nelson C E: Communication in the Physical and Social Sciences. Science 170 (11 Dec 1970), 1166 - 1173.

INFORMATION NETWORKS AND THEIR RELATION TO NATIONAL POLICY

G.X. Amey
Head, Information System Development
Defence Research Board, Ottawa, Ontario, Canada

National Policy and Science Policy

The decade of the sixties has seen a groundswell of opinion rise to challenge the conventional wisdom of previous decades which enshrined technology in its own unchallengeable niche. Informed critics are questioning the effectiveness of governmental structures currently used in the formation of public policy, especially in the exploitation of scientific discovery. In addition, attempts are being made to relate national objectives more directly to the allocation of resources among competing strata of research, from basic science to hardware development.

Efforts are being made to find explicit relationships between research carried out and that desired in a given country, in order to verify scientists' claims that the conduct of fundamental research is a necessary underpinning to development of advanced technology in that country. The evidence seems to be overwhelming that technological innovation is not usually generated by the scientists doing basic research, but arises from the elaboration of existing technology, supported by the body of generally available information, usually not very novel (1).

Reliance on a fuzzy association between fundamental research and improvement of the economic resources of a nation as a justification for funding research is being replaced by formal analysis of the relation between the scientific effort expended and the quality of life of the economy that results. Policy-making itself has been found susceptible to analysis (2).

A prime consequence of this soul-searching is the rediscovery that scientific research is not an end in itself but simply a method, a very expensive method of obtaining information. In recent decades, the well-known 'flood' of information has often appeared to be considered by scientists as a nuisance, although it is in fact the only permanent product of research and the only long-term justification for the high public investment in this area.

Inevitably the genesis of a new body of knowledge, 'information science' has become necessary, to organize this accumulation, in order to realize its potentialities for tackling current problems. Most scientists are interested only in generating information. The information scientist however, is not interested in how information is generated except from the point of view of cost-effectiveness. He seeks only the quickest and cheapest way to obtain information, by searching existing cumulations to determine whether new research is necessary. In 1959 the state of retrieval techniques was at such a low ebb that it was considered that if the research required would cost only $100,000 it was cheaper in certain documented cases to repeat the research rather than to attempt to retrieve data from the existing literature (3). In certain of the social sciences, retrieval of all the facts relating to a given area is dignified by the name of research, owing to the essential inaccessibility of such information as stored and catalogued in conventional archives.

One of the latest of a spate of reports generated by national and international organs on this problem is the Canadian 'Report on Science Policy', which I shall draw on heavily in this paper. As Omond Solandt puts it:

"We are now in transition from a period in which we thought of science primarily for science (but nowadays) science policy at the national level must be concerned more with how we use science to solve social and economic problems rather than with how to develop science" (4).

The area which has come most under attack is that of fundamental science, and in particular, big science. There was once the "myth of invincible research based on wartime experience with nuclear weapons .... if you take a great doctor and surround him with scientists in a laboratory, and if you have a clear-cut objective and put essentially unlimited amounts of money into that system, you can do anything" (5). Disenchantment with the premisses of science has set in after the launching of a series of unrealizable technological dreams.

On the other hand, where new technology has been introduced with unseemly haste, it has sometimes taken decades to perceive the cumulative ill-effects, as in the case of certain chemical products.

In many of the well-tilled fields of science, the need now seems to be, not to seek out newer and more esoteric facts, but to digest what has already been recorded and relate the various specialist disciplines so as to allow technological assessment of the consequences of exploiting the existing vast stores of knowledge. This calls for a marriage of system analysis and information science to solve the formidable technical problems involved in drawing the right inferences from seemingly unrelated bodies of facts.

In recent years, as research spending has been reduced in many countries, a surplus of people with technical or scientific training, has arisen. The influx of young graduates into mature research establishments has been drastically cut, thus raising the average age of staff in these bodies. This would appear to call for a reappraisal of the utilization of human as well as material resources to ensure that there is a sound distribution of people between the tasks of generating new information and of correlating and assessing the social relevance of data already generated. A conceptual analysis of certain aspects of these problems follows.

ALLOCATION OF RESOURCES BETWEEN RESEARCH AND INFORMATION RETRIEVAL*

In this paper I shall be speaking primarily of scientific and technical information (STI) and not of the personal and statistical information, storage of which in highly organized data banks could indeed be influential in developing national policy. The problem of privacy of data required in furthering sociopolicical developments is a very important one but not the subject of my paper. Only information from which the personal element has been eradicated will be considered, except for the obvious exceptions, - e.g., authorship of scientific papers.

In most studies of Scientific and Technical Information (STI), an attempt is made to determine the value of information in economic terms. These are often abortive since they are usually based on data from a sprinkling of cases in which research was duplicated. Allocation of funds to information retrieval is customarily expressed as an arbitrary percentage of the research budget. A more rational basis could be formulated by considering research and information retrieval as alternate channels for obtaining the same information. The choice of which route to follow could then be based on the relative cost effectiveness of the two methods (Fig. 1).

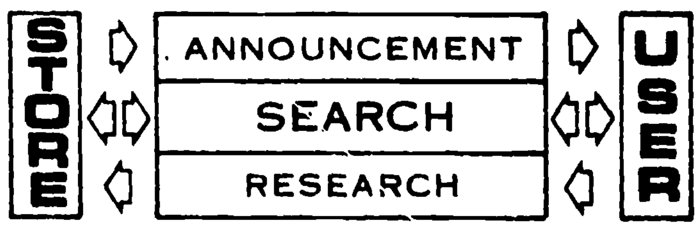


Figure 1 Channels for acquiring information

BIG SCIENCE ---- LITTLE SCIENCE

Research efforts can be divided into three classes according to the resources of money and people (team size) devoted to them, and thus according to their ability to achieve significant results. (Figure 2).

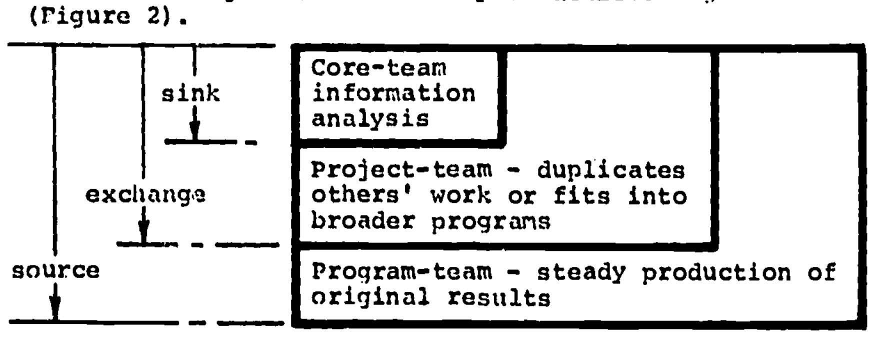


Figure 2 Categories of Research Teams

We may use a thermodynamic analogy to express the directions of flow of information between entities in each category.

<u>Core-team</u> This group can provide a core of experts capable of analyzing past and current literature so as to be aware of new discoveries, but is unable to contribute new results through research. Thus it serves essentially as a sink of information. That is, it can absorb any information it receives, but can make no sizable contribution in return.

The project-team is large enough to conduct non-trivial investigation in cooperation with other such teams with which they exchange information on a parity basis. Since such a team can make a significant contribution to research being carried out in a given field it is in good bargaining position for exchanging such information, either informally through the "invisible college" or through more formal channels, where proprietary or priveleged information is concerned with other sources of information. This it can do on equal terms with other like teams. Excellent concepts may be generated in such an environment, but funding may be insufficient to enable the ideas to be fully validated and so lead to further development of concepts.

The programme-team is large enough to ensure that whatever research it undertakes can reach the "takeoff" stage, as Rostoff has put it in referring to the parallel case of the economies of developing countries (6). At this level of effort, fundamental research can be conducted with a good chance of producing visible results in a relatively short time span. After a certain degree of success has been achieved, the success of the programme leads to generation of new ideas requiring further exploration and attracts resources needed to further the research. As shown by the "Hindsight" project, the payoff from such research may not appear until twenty years later (7).

Strategies for obtaining information

We shall concentrate on the mission-oriented branch since it is usually impossible to obtain a rapid economic payoff by seeking fundamental information.

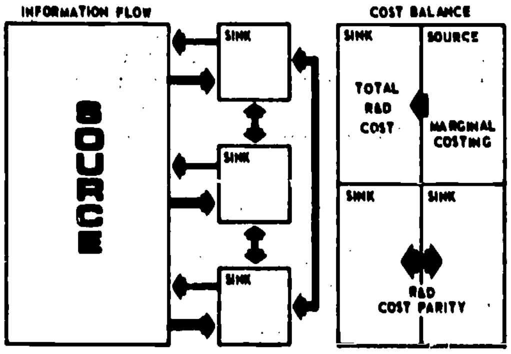


Figure 3

As Wedgewood-Benn put it, "In the early fifties, the promises of the national prestige and a technical capability equal to the 'super powers' was irresistible, and high technology entered the world of the fabulous. The large, prestigious projects... an expression of amour propre, are seen to be passing in favour of earthbound ones.. prompted by latent feelings of social responsibility" (5).

Supplies of information are often restricted for proprietary or other reasons. Thus it may be necessary to conduct research in a given area so as to gain entree to restricted sources of related information. Access may be attained by affiliation with international companies that possess large amounts of proprietary technology, according to Helge Berg, (14). The net flow of information between sources and sinks, to return to our thermodynamic analogy, can be schematically represented as shown in Figure 3.

Most nations contribute only one or two percent per annum to the world output of information. Since the accumulated store of all information may be taken as equivalent to about twenty years production at the current rate, the annual contribution by individual sources to the total store is no more than a tenth of one percent of the whole. Therefore the bulk of their information requirements must be satisfied by drawing on the cumulated stores of the rest of the world.

Only the United States and the Soviet Union qualify as net sources of information. They can justify costs of conducting research as parts of specific projects. The exchange of STI with other countries can be justified by marginal costing of the information supplied, since the prime costs are assigned to the projects supporting the research. On the other hand, in dealings between nations which are net sinks of information, an attempt will be made to equate the cost of information received to that given. This discussion is of course rather simplistic since it treats only the asymptotic cases. There are in reality, many combinations of circumstances which negate these extreme positions. Obvious examples are the status of Denmark, a small country, in theoretical physics, and of Isreal in a number of fields.

Relative costs of alternative channels

In view of this we shall consider the conceptual basis for selecting the options of research and information retrieval, strictly on a cost-effectiveness basis as alternative channels for acquisition of mission-oriented information. We shall assume that the total cost of a search and the probability of success are known, that accurate cost estimates for research projects exist; and that statistics are available on exchange of information between organizations.

If the estimated cost of research is greater than (cost of a search) divided by (the probability of non-success), then the route to be chosen should be that of information retrieval. There is often a premium for speed, as a function of time. This can be added to the cost of the research in the above inequality. A situation can be visualized where the premium could be a step-function where the information becomes valueless after a certain date, as in the case of facts required to support litigation. More typical is a continuous decrease with time, of the value of information as in the case of industrial information the value of which relates to the life of a patent.

Costs of Retrieval

Because of the overlapping of interests and the possibility of applying concepts developed in one field to another field, it is generally true that the probability of finding an answer is highly dependent on the size of the file. The cost of achieving success increases even more rapidly than the size of the file, because of the increasing difficulty of indexing material uniquely and then of extracting relevant data. It is customary in computerized retrieval systems to maintain statistics on successful search strategies. Thus the probability of a successful search with a given file can be estimated in advance and can be used to determine whether it is worth searching that file. Since there are now many computerized data banks in existence statistical data of this kind is necessary if money is not to be wasted in searching banks which offer little probability of success.

Human resources

It is generally believed that the judgment of a scientist increases with age while creativity declines after peaking at an age which is typical for the discipline (8). The analytic ability of a scientist no longer actively engaged in research is usually considered to be inferior to that of an active research worker. Creativity is traditionally measured by the production of scientific papers or by the assessment of his peers. The latter method is the only possible one in the case of consultants, who are essentially an elite class of information analysts. It is evident that for any given scientist, a time may come when he can be used with better cost-benefit in a function related rather to retrieval and evaluation of stored information than to the generation of new knowledge. When a number of such information analysts are using the same sources, it is possible to evaluate their relative skills in retrieving information, and thus provide an economic measure of their value in performing this function. The ratio of the cost of supporting such a scientist doing research, to his known output of scientific papers, could provide a measure of his value in the research function.

The notions considered above cannot give precise criteria for allocation of human and material resources between research and information retrieval functions. They may however indicate that a shift in emphasis is required in the use of human resources that have accumulated in recent decades. The age distribution of scientists in many organizations now peaks at an age well above the creativity peak for most disciplines. This indicates that a different philosophy in utilizing available resources should perhaps be considered.

Organization of knowledge to allow flexibility in use

The accessibility of information embedded in masses of data is very dependent on the methods used in indexing it. All indexing is subjective to some degree. While it may be hoped that the basis of classification is such that is will endure for all time, it is unlikely that categories selected, say a hundred years ago, have retained their relevance today. The current swing in interest away from traditional disciplines to interdisciplinary studies calls for a new approach to indexing. The abstraction of concepts developed in one discipline for application in another discipline has long been a problem. Since titles and indexing usually relate to concrete rather than conceptual data contained in the entities indexed, it is difficult to retrieve conceptual data from material indexed in the traditional problem oriented manner.

Retrieval of information thus falls into two main categories: one in which the information required refers to explicit data, about an entity already measured, recorded and labelled according to the same viewpoint as the searcher. In the second category, conceptual data is sought, that is, relations between entities, which have certain formal characteristics in common. The properties of such entities are not likely to be of interest beyond the discipline for which they were derived. The mathematical concepts of interrelation however are often transferable and can lead to significant insights when applied in a field other than that in which they were developed.

The problem of efficient retrieval of the first kind, from large bodies of data, has been solved using computer aids. Subject analysis by competent specialists enables standard terms to be selected from a thesaurus to characterize the stored material. It can then be recovered if the terms chosen include those by which retrieval will later be sought. Rapid obsolescence of terminology or the appearance of new applications of the information which were not forecast when it was stored will frustrate subsequent attempts at retrieval.

Interdisciplinary fields, formed by associating formerly unrelated fields, give rise to special problems. These however can be solved in some measure by machine-processing of sufficient textual matter, to produce index terms, chosen by algorithms which can in some cases add a time-evolutionary quality to indexing (9). The burden of retrieval is no longer equally shared with the original indexer, but falls largely on the searcher.

Retrieval of the second kind - of concepts or relationships developed in one subject-area and hopefully to be mapped over into another subject area, requires the serendipity, even genius of a searcher who himself requires this information, since only he can recognize it when it appears as through a glass darkly. This absolutely requires direct access to all possible data, however marginal at first appearance, in machine readable form, for manipulation in novel and personal ways to reveal possible hidden information patterns.

With a reorientation of aims in relation to exploitation of science and technology, networking systems that allow the creative user to coordinate at one location, the data from various sources, could play an essential role in moulding intellectual resources to meet new problems.

The traditional system imposes too great a burden of prediction on the indexer. An alternative method which has been extensively explored in recent years calls for storage and ordering of as much text as possible, either of original documents or of surrogates to the main body of recorded information. For short-run purposes, and selective dissemination of current information, traditional mission-oriented indexing, controlled by thesauri is very effective. Retrospective searching can however be greatly aided by concordances drawn from stored texts. These ensure that the searcher will not be strait jacketed by the indexers' conception of the intended application of the information.

Networks and end uses of stored information

The dual goal of modern nations can usually be summed up as that of improving both the standard of living and the quality of life. The paradoxical relationship existing between these goals has caused an abrupt discontinuity to occur in planning cycles of many nations. Reconciliation of the demands of rampant technology and the deep-seated distrust engendered by the perversions of nature which have arisen from undisciplined economic growth in the past, call for a fuller understanding of the whole information process... of education, research and the utilization of knowledge arising from these.

Utilization of research calls for more effective dissemination (and conceptual translation) of information than has typically been the case. Traditional systems are geared to moving information within one stratum of the hierarchy of users (10). Since fundamental researchers are not usually inventors or technological innovators, translation from the language of academe to that of the inventor is necessary. Certain systems utilizing much human input (11) are already helping in this direction.

To tackle the problem of technological assessment and verify that specific innovations will not import more harm than good into the ecosystem calls for new groupings of social science and other disciplines. 'The socio-technological institute, problem-oriented rather than discipline-oriented, interdisciplinary rather than disciplinary' as envisioned by Alvin Weinberg (12) does not really require the breaking up of existing disciplinary groups to form aggregations in multidisciplinary groups. Modern communications technology can enable men to swap ideas without requiring they be at the same location. Libraries do not need to be uprouted and reorganized to meet the fashions of the moment, if the original machine-language cataloguing contains enough data to allow new mappings of old coordinate indexes to chart new areas of involvement. This process has long been used by social scientists who employ statistical and textual data gathered for one purpose (land registration, birth notices etc.) to form new insights into social processes. The automatic storage of large quantities of textual data by libraries can make such uses much easier. The principal characteristics of such networks would appear to be as follows:

1. Access to stored information must be convenient and flexible, erecting minimal barriers between the user and the data in which he wishes to browse (13).

2. Standardization of data storage and of access languages should cut across barriers between disciplines. The tendency to treat bibliographic or any other homogeneous data such a geocoded data as unique should be resisted, since any user may need access to data of many kinds.

3. There should be a common query language, available in various subsets of elaboration for use with data banks associated with autonomous centres of information analysis.

4. Natural language retrieval should be available in addition to searching with thesaurus controlled terms.

Bibliography

1. Price, D. de Solla The Difference between Science and Technology. T.A. Edison Foundation, Detroit (1968)

2. Dror, Y. Public Policymaking Reexamined. Chandler Pub. Co. San Francisco (1968)

3. Sisco, F.T. Metal Progress, 70, No. 4. p. 122-124 (1959)

4. Senate Special Committee on Science Policy. A Science Policy for Canada. Vol I Critical Review: Past and Present. Queen's Printer, Ottawa (1970)

5. Wedgwood-Benn, A. Science Journal, 5A, No. 4, p. 3 (1969)

6. Rostoff, W.W. The Stages of Economic Growth. Cambridge University Press, U.S. (1970)

7. Sherwin, C.W. and Isensen, R.S. First Interim Report on Project Hindsight (Summary). Office of Defense Development, Research and Evaluation (1964)

8. Deutsch, K.W., Platt, J. and Senghaas, D. Conditions Favoring Major Advances in Social Science. Science, 171, p. 456. Feb. 5, 1971

9. Hoyle, W.G. Automatic Retrieval and Dissemination. Cranfield Conference, (1971) (to be published)

10. Amey, G.X. Channel Hierarchies for Matching Information Sources to Users' Needs. Proceedings of the American Society for Information Science, Vol. 6 (1968)

11. Scott, R. Communicating Information to Users. FID International Congress of Documentation, Buenos Aires (1970)

12. Weinberg, A. In Defense of Science. Science, 167, p. 141-145 (1970)

13. Amey, G.X. The Forgiving System: Human Aspects of Information System Design. FID International Congress of Documentation, Buenos Aires (1970)

14. Berge, H. Business Week Nov. 7, p. 29 (1970)

LONG TERM PLANNING OF I&D IN NORWAY

A. Disch
Director, Norwegian Industries
Development Association

SYNOPSIS

Long term planning (LTP) for the I&D-sector, as part of an overall LTP for scientific and technical research in Norway. I&D perspective analysis for the 1970-80 period, economic planning frames and I&D programs and projects for the 1972-75 period. Results and experience from the study.

<u>Activities of the Norwegian Council for Scientific and Industrial Research, NTNF</u>

The Royal Norwegian Council for Scientific and Industrial Research (NTNF) was established in 1946 as a semi-governmental organization with the object of promoting scientific and industrial research in Norway. The activities of the organization include the following sectors:

manufacturing industries
energy and electricity supply
shipping
building and construction
transportation
pollution
technical information, etc.

The Research Council endeavours to develop a national science policy within its terms of reference as a basis for its activities which are comprised of the following programs:

- recruitment of scientific and technical personnel, fellowship grants, etc.
- research grant programs, both for university institutes and applied research institutes
- establishment and administration of research institutes in fields of national interest (the NTNF institutes employ at present approximately 1800 people).
- research contracts relating to applied research institutes and to industry.

In 1964 the Council presented a comprehensive report to the Government with an evaluation of the need for research in all of its fields. The priorities established have so far been a guideline for the activities of NTNF.

Introducing Long Term Planning
In 1969 the Council decided to make a new study with the aim of introducing formal long term planning all through its system. The study started in late 1969 - early 1970 and ended with a comprehensive report late in April 1971.

The central planning committee for this study is the Executive Committee of the Council, in conjunction with the head of the Economic Planning Section of the Ministry of Finance. The whole area covered by the Council has been divided into 22 fields of activity, each of which is covered by a planning committee, including one on Scientific and Technical Information, the "I&D-Committee", which was established in February 1970.

The Five Phases of the Study
The long term planning study has been divided into 5 different phases for each of the sectors to be analyzed. In the following paragraphs the general guide lines set up for all the sectors will be briefly described and the special procedure carried out for the I&D-sector treated in some detail.

		Finished by:
Phase 1:	Develop 10 year perspective analysis for each sector	June 1970
Phase 2:	Determine the R&D role in attaining national goals, establish economic planning frames	Feb. 1971
Phase 3:	Prepare 1972-75 programmes and priorities in each sector	April 1971
Phase 4:	Plan detailed action programmes and projects	June 1971
Phase 5:	Sum up, prepare comprehensive report	Oct. 1971

Figure 1

Time Schedule for Long Term Study

Developing Perspective Analyses
Phase 1: By means of methodical studies each sector committee was to try to identify major opportunities and threats confronting the sector in the next 10-year period. This included analyzing the technological changes needed to maintain cost and quality at internationally competitive levels, the technological opportunities to obtain new growth from new products, processes or material changes, technological threats from competitive materials or processes, identification of significant breakthroughs and the impact of these on industry if they actually do occur.

In carrying out these studies, the various study committees were supposed to use different types of technological forecasting techniques, such as Delphi-analysis, relevance-tree analysis, etc. but first of all to collect reports on similar studies carried out in other countries.

The I&D-sector

As an inter-sectoral activity with no specific ties to any one particular industrial sector, the long range planning scheme visualized for the industrial sectors could not be used directly for the I&D-sector, and as far as could be seen from available literature, no similar studies had been carried out elsewhere. The committee thus had to try and work out a method that could be adapted to the national I&D-situation.

The first step in the analysis was to structure the problems as systematically as possible. To ensure a systematic survey of all the I&D-activities it was found that the use of matrices was very helpful in identifying the impact of the various activities on the different user groups and vice versa. This technique was used partly as an aide-memoire, partly as a tool in the structuring process. (Figure 2).

Based on this exercise, which was carried out as a kind of committee brain-storming, but constrained and aided by the matrix method, a structure of the ideas on the future development of the I&D-sector evolved. This structure was further developed into a final plan for the perspective analysis report. This presented the committee with the problem of deciding whether to use a "horizontal" or "vertical" structuring according to the information tools used, according to the missions or subjects covered, according to the users' different needs etc. As happens in so many similar cases a compromise had to be made.

USERS / ACTIVITIES		Politicians	Industrial Managers	R & D Specialists	Practitioners			School sector		Societies	Others	Æ	Ø	Å
					Engineers	Foremen	Workers	Teachers	Students					
Primary Literature	Trad.													
	New													
Secondary Literature	Trad.													
	New													
Libraries	Trad.													
	New													
SDI	Trad.													
	New													
Retrieval	Trad.													
	New													
X														
Y														
Z														

Figure 2

Example of Type of Matrix Used to Check Cross Effects

<u>Developing the I&D Perspective Analysis Report</u>
Each chapter was assigned to an author or a team of authors to be treated according to guide-lines set up by the committee. The committee realized that this mode of operation had certain disadvantages. It is, of course, impossible to get a number of people to think and write uniformly. There would undoubtedly be misunderstandings and misinterpretations of the guide-lines, resulting in considerable variations in form and content of the contributions. The committee felt, nevertheless, that this method had one big advantage compensating for the disadvantages: a great number of people became interested in and personally involved in the perspective analysis. It is important in a small country such as ours that the people who eventually will be involved in the actions resulting from the long term planning, not only feel that they have been consulted, but have personally taken part in evolving and expressing the ideas that are parts of the total plan.

One interesting experience was gained during this stage: the specialists who are deeply involved in practical information work often have difficulties vizualizing future developments beyond the next few years. It appears that the every-day practical problems that they have to deal with put a constraint on their imagination. Their time-perspective seemed in many cases to be limited to a maximum of 3-5 years, while the committee was thinking in terms of a 10-year perspective. This meant that the committee in a number of cases had to extend the perspective of some of the sections by 5 years or so.

After having received the various contributions, the rather tedious task of moulding them into a uniform report followed. This was done parallel with a number of hearings where the people involved were presented with the preliminary, condensed report and asked to express their views and criticisms.

<u>Overall National Goals and the Role of the R&D-Sectors</u>
The perspective analysis report from this committee, together with the reports from all the other committees formed the basis for the <u>Phase 2</u> activities.

In this phase the central planning committee of the Council condensed and evaluated all the reports and made a summary of its findings. This was presented and discussed with representatives from the Parliament, the Government, the ministries and the labour and industry organizations. From these discussions, the part that R&D should play to help reach the national goals was established, and the overall allocation from Government funds to R&D was estimated for the four-year period 1972-1975, within a lower and an upper frame. The lower frame represents an extrapolation of the present trend in Government allocations to R&D, the upper frame a more optimistic view, as a result of the present exercise.

The planning committee then had to tackle the formidable task of evaluating the importance of each of the 22 sectors in relation to each other and determine high and low financial planning frames for each sector. The basis for establishing the "starting point" (1972) was in part earlier activity in the sector, partly the importance of the sector in contributing to the national goals.

The expected <u>growth</u> for each of the sectors for the four-year-period, was decided on the basis of a matrix, where the interaction of each sector with all other sectors was estimated on a 1-2-3 scale (Figure 3). The I&D-sector was found to have the highest total ranking and consequently given the highest growth factor.

	Chemical Industry	Electronic Industry	Wood and Wood Processing	Building and Construction			Technical Information
Chemical Industry		2	2	1			3
Electronic Industry	1		1	1			3
Wood and Wood Processing	2	1		2			2
Building and Construction	1	1					3
Technical Information	1	3	1	1			

<u>Figure 3</u>

Interaction of the Sectors

<u>Developing I&D Programs 1972-75</u>

In <u>Phase 3</u> each sector committee was given economic planning frames expressed in million Norwegian Kroner for each of the years 1972-75, in both an upper and a lower planning frame. The committees were then asked to draw up well defined R&D programs within these two frames for the four-year period (Figure 4).

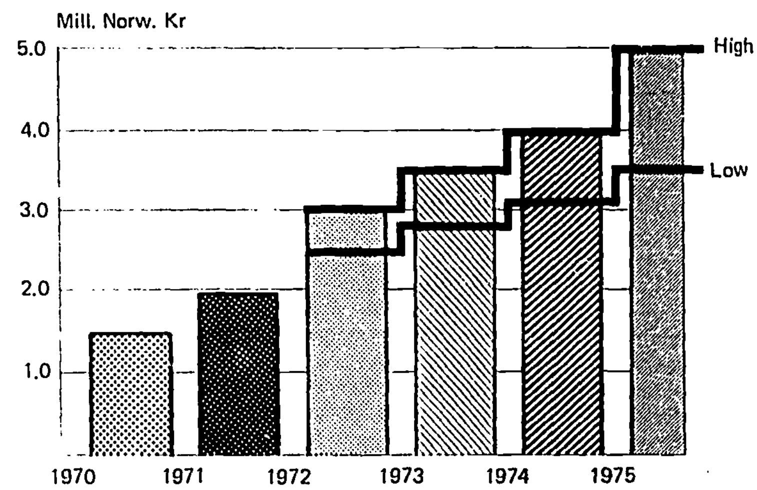


Figure 4

High and Low Four-year Planning Frames for the I&D Area

In the I&D sector, this was done through a series of processes.

National goals, sector goals, operative goals, etc. were defined as illustrated in Figure 5. From this, three main fields of activity were identified and, within each field, the I&D-programs defined which in turn were sub-divided into I&D-projects. For each program and each project, the necessary resources for the years 1972-75 within the two frames, were indicated.

Obtaining the Users' Reactions

Before arriving at this stage, however, the committee had arranged a number of meetings, involving some 250 people in the three largest cities in Norway. At these meetings the participants were presented with the perspective analysis report and a framework for the proposed action program. The participants were asked to supply the committee with proposals for I&D programs and projects within each of the three fields and, insofar as possible, rank the various programs. To ensure maximum efficiency of this particular excercise, the participants were split up into small work groups of 6-8 people, with members of the planning committee present in each group to answer quastions on procedure and intent of the study.

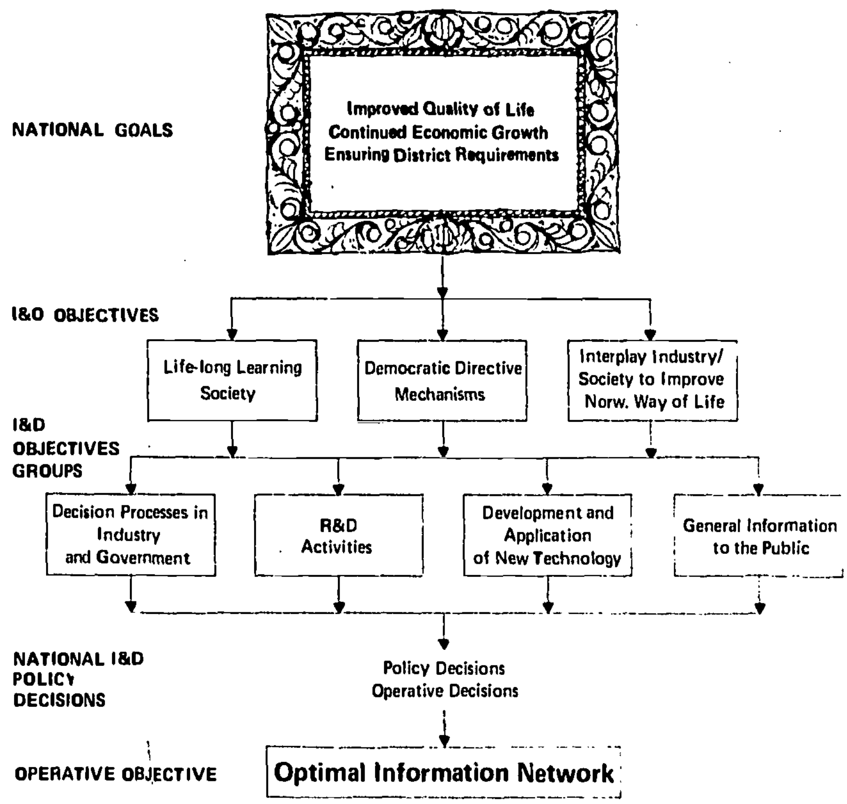


Figure 5

From National Goals to I&D Operative Objective

Each work-group presented its findings in a final plenary session, where they were discussed generally. From these discussions and sessions evolved a set of I&D-programs and projects which were structured, ranked and put together by the committee.

This, then, forms the basis for the Council's central I&D-activities for the four-year period 1972-75.

In Phase 4, the various programs and projects were treated in some detail and decisions made as to where the projects are to be carried out, by whom and how.

Phase 5 consists in summing up the results of the activities in all the committees and presenting these, together with the central committee's conclusions, in a final official report.

Figure 6 gives a concentration of the I&D-programs. It should be noted that this represents only the Council's own central I&D-activities which are a supplement to those activities already taking place in the Council's various R&D institutes, in industry and in other organizations.

Field:	Programme:
1. User's Use of Information	1.1 Personal communication 1.2 Information adaptrd to Norw. conditions 1.3 Motivating the user
2. Education and Training	2.1 Fellowships 2.2 Course and seminar activities 2.3 Integrating I&D in Education
3. Systems, Technical Aids	3.1 Systems Development 3.2 Increasing Library Efficiency 3.3 Presentation Modes of Information 3.4 New Media and Technical Aids (Computers etc.)

Figure 6

The Three Main Fields and Main I&D Programs

The I&D area has been divided into three main fields, each field into a number of programs and each program into a number of projects. Before granting approval to any project, a detailed plan, including a budget, a milestone-chart and users' interests had to be presented.

This could result in an extremely rigorous and inflexible system without opportunities for that private initiative and personal engagement which is so important for the healthy development of any activity. This was fully realized by the committee, and the necessary flexibility has been built into the system. The overall program will thus be updated every year, the emphasis can be shifted freely from one information program or project to another, entirely new projects can be included and the course of the old ones can be changed.

Field:	Programme:	Project:	1972 High/Low	1973 High/Low	1974 High/Low	1975 High/Low
3. Systems, Technical Aids	3.1 Systems Development	3.1.1 Development of Integrated Information Systems	50/35	150/100	300/100	350/200
		3.1.2 Studies of IAC in Special Areas	0	50/25	50/25	100/50
		3.1.3 Establishment of IC for I & D	75/0	100/75	100/75	150/150
		3.1.4 National Referral Center for R & D	0	50/0	50/25	100/50
		3.1.5 Development of Technical and Other Thesauri	20/20	20/20	20/20	20/20

Figure 7

Example of Field, Divided into Programs and Projects with High and Low Allocation

General Conclusions

The I&D committee decided to try and build a fairly broad I&D platform, not concentrating only on one or a few advanced areas. It was felt that since the I&D environment in Norway is so limited it was necessary to create an awareness of the importance of information in to-day's society both among the users and the authorities, and to build up information know-how in industry and in the academic institutions.

On the "hardware"-side the national initiative will be limited to an alert-service and to creating the necessary know-how within the country in order to make use of advanced information technology on a national basis, when the need arises.

Experience Gained from the Long Term Study

Looking at the conclusions for the I&D sector these might seem to be a rather meager result after this hectic activity, involving many people and costing considerable amounts of man-hours and money. It would appear that it should have been possible to draw these same, seemingly obvious conclusions with considerably less effort.

The most important result of this excercise lies, however, on a different level. Through the involvement of so many people in thinking constructively about long term planning in the I&D-sector, the ground has been thoroughly prepared for future activities. Specialists who are engaged in I&D-work had to stop short in their daily routines and get themselves involved in the problems of to-morrow and consider what the impact of to-morrow's development would have on today's activities. An awareness of the necessity for thinking and planning ahead was created. A favourable information climate in the different R&D and industrial trial milieus seems now to have evolved as a result of this study. This in itself is extremely important.

But besides this, there is of course the more tangible effect: The creation of a concrete four-year program with proposed allocation frames is a powerful tool in steering the I&D-activities in the right direction. With the limited resources that we have available for I&D-activities, it is important that these resources be used as constructively as possible to pursue the objectives of the national I&D-activities. And it is also very important to bear in mind at all times that I&D is no goal in itself but only one of several tools that can be used in trying to reach the overall national objectives that we, as a people, have set.

CLOSING THE INFORMATION GAP - TOWARDS A NATIONAL NETWORK IN ISRAEL

Carl Keren
National Center of Scientific and Technological Information
Tel Aviv, Israel

SYNOPSIS

Existing information facilities in Israel are related to the users' community and the economic developments of the country. Deficiencies of the system are identified. The creation of a national network is proposed, and the role of its various components are outlined to provide a framework within which information services can be improved.

Past development of information services in Israel has produced some significant milestones:

* A National Library, now the Jewish National and University Library, which with more than 1 1/2 million volumes is the largest library in the country.

* The library of the Technion (Israel Institute of Technology) containing a comprehensive collection of technical books and periodicals.

* A Center for Scientific and Technological Information (COSTI) which is becoming a national focal point for information activities.

* The Israel Society for Special Libraries and Information Centers (ISLIC), an association which in the short span of five years has grown to 170 active members.

* The Graduate Library School of the Hebrew University in Jerusalem and the postgraduate courses in Information Science organized by COSTI and the Weizmann Institute of Science.

* The preparation of a comprehensive survey of information facilities in Israel and the submission of a report to the government proposing the adoption of a policy designed to foster information services in the country.[1]

Along the road marked by these milestones are library and information services developed in the country, which today comprise:

* Seven university libraries

* Some three hundred special libraries[2]

* Several mission-oriented information Centers

* Several discipline-oriented information Centers

* A national focus for information activities (COSTI) within which some nationwide services are executed and which maintains international relations and representation in this field

At first glance, these developments can be considered rather satisfactory for a small country just entering its 23rd year of independence and only recently attempting to create an industrial society. They certainly form a sound basis for future development and for the formulation of a policy for science information.

Additional investigation, however, has shown that the quality and number of users to be served has grown more rapidly than the services available. Much information pertaining to new technologies has not been collected in the past, partly perhaps due to the emphasis on basic research which was prevalent in Israel. Non-book information media (reports, films, tapes, etc.) had been neglected by most libraries. There was little or no knowledge of advanced methods, i.e. automation, computerized type-setting, micro-recording, storage and retrieval, data analysis and manipulation. But most important, there was little awareness among information workers of the new trends and of the fact that responses to the needs of a technologist or an applied scientist differ from interaction with a scientist in an academic environment where time and cost benefits are not as important.

Israel today has some 4000 active scientists, approximately 15,000 engineers, a fast developing high-technology industry, an average 9% per annum increase (over 20 years) in the G.N.P., and a per capita income reaching US $ 1550 in 1969.[3] Its expenditures on R&D will amount to IL 450 million in 1971/72 or more than 2% of the G.N.P. These very few figures highlight a very accelerated rate of development to be served at each and every stage by adequate information services.

Member states of the OECD have found that, on the average, a developed society spends about 5% of its R&D expenditure on information. This figure varies in different countries. It is far from being unambiguous and uncontested, but in the absence of any other measurement of the expenditure on information which can be better defined, it is more or less accepted. Israel spends on the same kind of services less than 1 1/2 % of its R&D expenditures; and if the availability of a service can be measured by the price paid for it by the community, then this figure points to the fact that there is much room for improvement.

Availability of information by assessing the money spent on it relative to money spent on activities in which information is essential input is one way of describing the situation. There are additional factors which play a role and in view of their applicability to developing countries in general, and those which are geographically remote from Centers of Science and Technology in particular, they seem worthy of mention.

* Information, as many other resources, requires a "critical mass" in order to be fully utilized. For example, the same basic library stock is needed in a large research establishment as it is in a smaller one. The basic price for an SDI system is the same, whether 50 users or 300 are served.

* Information is transferred by various media of which the transfer by word-of-mouth is the most prominent. Within R&D Centers, meetings, conferences, telecommunication and other means of oral transfer are common. Users who are geographically remote are unable to participate freely in this exchange and are more dependent on the written word.

* Colleagues active in the same or similar professional fields supplement each others' need for specific or general items of information. R&D group members consult freely within the group or with similar groups elsewhere. In a country the size of Israel R&D groups are by necessity small and in some fields only one or two specialists are active. Opportunities for information exchange and consultation with colleagues are thus severely restricted.

* Information stores improve with time; the longer they exist, the more comprehensive their holdings, the larger the data base on which they draw. Young countries with relatively recent research establishments and libraries have not had the opportunity to gather the collections they need in order to respond fully to their users actual and potential requirements.

* Information stores, and especially libraries, are chronically underfunded everywhere. Wherever rapid development is the rule this is even more apparent because means are often appropriated according to priorities which rarely include such invisibles as stocks of books and the personnel to manage them.

* Finally, information services thrive on cooperation and exchange. Services in developing countries have little to offer and much to ask for, another factor retarding their growth.

Information is a national resource, even more so in a country like Israel where more than 80% of the national R&D effort is government financed and where the government provides most of the infra-structure of the national economy. After having recognized in our report that information services in Israel have not kept pace with the development of that infra-structure, and after having identified some of the more significant reasons for this lack, a policy was proposed - and accepted by the authorities - aiming to provide a framework and the means to correct existing deficiencies. In our plans we drew heavily on the experience accumulated elsewhere and especially in OECD and CEMA countries.

Within the next few years we hope for the following developments in information services in Israel:

* An improved network of university libraries to be developed by the Standing Committee of University Librarians.

* A significant improvement of the special libraries and information centers network, especially in industrial and other mission-oriented enterprises. Those services are to be partly initiated and partly aided by a number of newly trained science information specialists graduating from courses held by the Center for Scientific and Technological Information and other institutions.

* A network of discipline-oriented information Centers, to be established according to users needs and with users participation. Information Centers for Chemistry, Computer Sciences, Ecology, Water and Desalination Research, Electronics and Management Sciences are presently under discussion.

* The Center for Scientific and Technological Information has recently been designated to be in name, as well as in fact, the national focus for science information activities in Israel. No detailed plan for its future role has been authorized as yet. It is, however, anticipated that the Center will devote much effort to the education of information users and suppliers, including continuous training of workers in the field; the initiation and development of discipline-oriented information centers; the introduction, and sometimes operation, of computerized services including SDI services and data collections; initiation of investigations into the improvement of technical and economic information services, especially those required by industry; the introduction of common standards and procedures facilitating cooperation between the various nodes of the national network.

* Increasing participation in international activities such as UNISIST, CODATA, FID, INIS, and other organizations. Close cooperation with international as well as national organizations aiming at free and unhampered exchange of knowledge and material must be considered an absolute necessity. We are well aware that participation in the international network must work in both directions and steps have already been taken to share some of our experience with others.

This is the goal we have set ourselves and the organization we are building for its achievement. We hope that through them we can close the information gap and through better information contribute towards raising the standard of living.

BIBLIOGRAPHY

1. Report of the Committee for the Development of Scientific and Technological Information Networks in Israel. The National Council for Research and Development, Jerusalem, 1970.

2. Directory of Special Libraries in Israel, 3rd Ed. Center of Scientific and Technological Information, Tel-Aviv, 1969.

3. Scientific Research in Israel, Center of Scientific and Technological Information, Tel-Aviv, 1971.

DOCUMENTING CURRENT MATERIAL ON AFRICA & ASIA

Dr. K. K. ROY
Editor, Bibliographia Africana & Bibliographia Asiatica, Calcutta, India.

SYNOPSIS

The ever increasing number of newspapers and periodicals published in Africa and Asia present various problems of access and understanding. An effective solution could perhaps be a monthly monolithic encyclopedia produced at a single centre in each continent with the help of all countries within these two continents. Such a centre should operate internationally with computer and photo duplication facilities.

The past decade has witnessed a meteoric rise in the number of journals newspapers and periodicals in African and Asian countries, the output of which is so voluminous and widespread that no researcher could master it without adequate bibliographical and documentation tools. Moreover the character of the material being currently published in these two continents is also rapidly changing. It has spread over a vast media in a number of different languages, giving rise to problems of access and understanding.

This situation has resulted in several bibliographic and documentation projects and services both within these two continents and also in America and Europe. In reviewing these bibliographical and documentation services, we find that about 16 countries each in Africa and Asia publish national bibliographies. These national bibliographies are invariably behind schedule and list mostly books, government documents and reports. Very little indexing and itemising of newspaper and periodical articles is done which is the subject of this study. In both these continents, the various procurement centres of the U. S. Library of Congress issue Accessions Lists which are published on time, but these Lists are very selective and do not index any newspaper or periodical articles which in reality are outside its scope. India and Japan are probably the only two Asian countries in these two continents who publish bibliographies and documentation lists on current materials that are published in Africa and Asia. Mention here may be made of the Institute of Asian Economic Affairs and the Kinsei Monjo Bunrui-Hyo both of whom are in Tokyo and the Bibliographia Africana and Bibliographia Asiatica both of whom are in Calcutta with which I am associated. Apart from these two centres in Tokyo and Calcutta there are no other centres throughout Africa and Asia who document or in any way index current articles in newspapers and journals relating to countries other than their own. But both the Indian and Japanese centres also face limitations in as much as only current material in English is being documented and indexed as translation difficulties prevent this service from being extended to African and Asian languages media. The Japanese centre has another disadvantage because there the documentation is being done only in the Japanese language and is thus not easily accessible to those researchers who do not know this language.

The position outside Africa and Asia i.e. in America and Europe falls into three main groups (1) those serving the whole of Africa and Asia in respect of most subjects in the humanites and social sciences (2) those which cover the whole of Africa or Asia or an extensive region of either of these two continents but which are restricted to a single subject or a relatively restricted range and (3) those which are concerned only with a single region or a country in Africa or Asia but which include writings on practically every topic. Moreover, both in America and Europe, we find that there are numerous centres and projects for Asia but relatively few for Africa. This is probably because the scope of Asia is compellingly wider both in content and complexity. The history and cultural heritage of Asia goes back several thousand years whereas the interest of the outside world in Africa stretches at the most to the past hundred years or so. American and European bibliographic and documentation projects are practically confined to the articles, books and theses published primarily in the West. Very little material indigenously published in Afro-Asian countries finds place in the indexing services of these projects. This observation is being made on the basis of a study made on four current bibliographic services produced in card form in America and Europe and about sixty periodical publications of Western countries which regularly listed titles or gave resumes of articles relating to Africa and Asia. It was seen that these publications cover about 400 periodicals in Canada, Europe and U.S.A.

While all these projects without a single exception are commendable almost all of them have given rise to a duplication of efforts despite the fact that none of the existing projects guarantee exhaustive coverage. To take an example, a careful analysis would reveal that most American and European services covering most of Africa and a wide range of subjects list anything from 2000 to 3000 titles in a year. Whereas a bibliography produced for a single subject ''Bibliography of Agriculture'' by the U. S. National Agricultural Library, claims that about 3500 subject listings annually were specifically concerned with Africa. The situation regarding Asia is also not much different.

The inadequacy of the present bibliographical and documentation services is therefore apparent and unless compilers of these services work to-gether to determine the strong points and lacunae of existing services not much co-ordination is possible. Compilers must first of all try and understand the present needs of the research worker who has to basically use such services. Most researchers are very ignorant of current bibliographical and documentation processes. They search through hundreds of periodicals and indexes for a particular topic which costs them an inordinate amount of time achieving slender results. A recent study showed that on an average seventy five percent of a researcher's time was spent on bibliographical search.

It is not suggested that the present services are not useful. Certainly these indexes to current literature render yeoman's service in some fields like the political and economical sciences, but my point is, that they do not offer a thoroughly satisfactory solution to the problem. Each of these services uses a different system of compilation so that to compile a corpus of information on a particular subject by utilising these various resources is a lengthy operation by any standards. The multiplicity of such current bibliographical services poses the following two questions :—

1. Is it possible to reduce the amount of time actually required to search for a newspaper or periodical article relating to a particular subject ?
2. Is it necessary for a library to subscribe to several card services and thirty or more bibliographical or documentation services in list form in order to be able to furnish a useful index to current literature ?

To the first question I would answer yes. In view of the increasing complexity of scientific knowledge, the importance of compiling bibliographies in different disciplines of different regions should be undertaken by specialists in various subjects who should work under the direction of Bibliographical Editors so that indexing and classification may be done according to a standardized system. This work should be shared among various documentation centres, each centre being responsible for the exhaustive coverage of current materials in its own specified field. This should eliminate a lot of duplication and possibilities of spending time in searching a particular article would thus also be considerably reduced, since time is usually spent in referring to a number of bibliographical sources to see if the particular article one is looking for is listed therein or not.

My answer to the second question is in the negative. It is my considered opinion that a library need not subscribe to so many services, if there is a single comprehensive service available which can meet its needs.

This brings me to the ideal African and Asian Bibliography. How would we all like to have at our disposal a monthly monolithic encyclopedia, unrestricted as to subject, type of material, language and country of issue. Such a bibliography would record all current material being produced all over the world relating to each of these continents, with suitable annotations, according to an accepted system of classification along with an indication of where the documents might be found. This monthly encyclopedia would be further divided for the sake of convenience into three main sub divisions :—(a) Subject (b) National and (c) Area or regional.

Interestingly enough a number of inter-continental library organisations both in Africa and Asia have been debating such a proposal, but the problems of skilled professional bibliographical effort and the funding to see it through have led to such an ambitious project being dropped for the time being. This type of a proposal was first mooted for Africa at the Tropical African Studies Conference at Ibadan in 1964, where a broad framework had been reached. Thereafter the International African Institute and The International Congress of Africanists at their respective international seminars decided to sponsor co-operative bibliographic projects so as to exchange technical knowhow and minimize duplication of bibliographical and documentation work relating to African studies. Some headway has undoubtedly been made in this case but it again relates to books and retrospective material. Current material has not so far been approached by such projects.

The position in Asia has been somewhat different than that in Africa. As a result of the three conferences on Southeast Asia Documentation, Southeast Asia Research Materials and the Southeast Asian Librarians held in 1969 and 1970, three committees have been formed namely—The Committee on Research Materials in Southeast Asia (CORMOSEA), The Southeast Asia Library Group (SEALG) and the Conference on Southeast Asian Librarians (CONSAL).

Other Regions in Asia were covered earlier in 1957 at the Tokyo conference of the Asian Federation of Library Associations. The only significant contribution made by these meetings was that a need was felt all around to make a sincere effort to control, expand and standardize existing bibliographic and docmmentation resources. Not much headway however was made in these directions because of various difficulties mostly related to finance, dearth of skilled personnel, proper procurement procedures etc.

Western Europe, Canada and U.S.A. have also hoisted such meetings from time to time, but again due to a lack of coordination such conferences have had little or no impact upon the basic problem which face African and Asian bibliographic and documentation work.

To prepare a current continental bibliography and documentation service we need to index all current newspaper and magazine articles as well as research papers and documents, economic and industrial reports, conference proceedings and commercial statistics publiehed by government and international organisations located in the particular continent. It should be annotated according to an internationally agreed classification syetem with an indication of where the documents might be found. Such a volume can only be produced in time if the material to be listed is broken down in three main divisions namely: (1) subject (2) national (3) area, country or regional. The acquisition of material to be indexed is a basic factor to be considered first. To solve this problem, I would suggest that atleast one library in a country should acquire all current publications issued locally either by legal deposit or informational exchange. These should be listed promptly. This institution should also be equipped to develop photographic facilities so that current material which soon goes out of print, could be copied and dietributed at a nominal price. Facilities for the acquieition and recording of current local publications should be offered to thie chosen institution, perhaps largely in the form of provision for travel and expenses of an itinerant bibliographer. Two obvious needs here would be (1) to encourage the local production of acceptable national bibliographiee and (2) to satisfy the acquisition requirements of libraries elsewhere.

This limited national bibliography could then be the national contribution to a larger project for the continental bibliography in general, for it seems clear that when basic data so acquired is manipulated rapidly, it can provide a variety of bibliographies, with the help of a computer. I agree that bibliography is not yet in the computer age and mechanized information retrieval is presently considered more suitable for the natural sciences and technological subjects, but computer possibilities are the only ones to be taken into account when we consider speed of production an important factor. Another factor in favour of a computerized bibliography would be that all countries contributing frequent material listings would naturally expect to receive equally regularly the more compreheneive and cumulated bibliographies of their countriee from this continental centre.

The coet of auch a computerized bibliography might, of course, be prohibitive, and collaboration in such a project would quite naturally entail sacrifices on the part of organisations already profilic in the production of African and Asian bibliographies, but in my opinion, such a continentally produced computerized bibliography is worth all such sacrifices — probably many more, when weighed against the possibility of finding all published listings in the continent within its pages.

To get such a project going it is considered appropriate to grant financial support following inquiries into the real needs of the various countries and to set up a co-ordinating body whose task will be to transform the national contributions into effective bibliographies of various kinds.

A careful parusal of this problem perhaps could help us to initially start on the project on the undernoted basis :

I International and National Information Networks

1) Every African and Asian country irrespective of its national language should establish or designate a national bibliography co-ordination centre. This centre would be responsible for gathering to-gether all the bibliographic information produced in that country.

2) The level of work done by each national centre would depend on the facilities of the particular country.

   a) This centre should have key punching and computerization equipment and it should ensure the processing of bibliographic information on to tapes.

   b) For those countries who do not possess computerization equipment, the national centre there should be responsible for ensuring that bibliographic information would be processed at least to the stage where it would be ready for key punching. This data could then be sent to a centre of any neighbouring country where the concerned centre has computerization facilities.

   c) The amount and type of work required at any national centre would depend upon the capabilities and number of the institutions within that country.

3) The highest priority should be given to the establishment of atleast one documentation centre in each of the two comtinents with key punching and computerization equiqment. The location of this centre should be decided upon by the respective countries within the two continents and it is suggested that its governing body consist of one representative from all the countries within the continent.

4) Each national centre within the two continents would be entitled to receive from ite continental documentation centre a complete set of punched cards, and at least until it obtained equipment equivalent to an I.B.M.-870 document writer, a print out of the bibliographic information in whatever physical format might be agreed upon.

5) All bibliographic information should be provided by the national centres in English or French. This would mean that all articles in other languages should be translated into either English or French and that all annotation work be done in one of these two languages.

6) Two working parties should be establiahed :

   a) An international co-operation committee to plan for the establishment of two continent wide documentation centres—(The establiahment of the national co-ordination centre should be left to the individual countriea). Once the national centres and the continintal centre were established, the composition of the international committee should then consist of a representative from each national centre and its function would be that of ensuring co-ordination between the centres.

   b) An international standardization committee for establishing a common bibliographic style and annotation scheme, meeting both scholarly and computerization requirements. It is recommended that the H.R.A.F. (Human Relations Area Files, Yale University, U.S.A.) computerized bibliographic system be used as the basis, at least initially, for the deliberations concerning standardization. This system in my opinion is sufficiently oriented to both the human and technological sciences and much of the necessary programming has already been done. It would be very expensive to repeat this work and the system as it stands is open to truly collaborative decisions regarding style and the specifics of any annotation scheme.

The broad lines of the project proposed above is wholly based on computerization. There is a school of thought however which argues that Bibliography is not yet in the computer age and that mechanized information retrieval is more suitable for the natural and technological sciences. To my mind it is a seeming misconception on the part of some bibliographers regarding the implications of computerization. A computerized bibliography would in no way dispense with bibliographers or take the 'heart' out of bibliographic research. Quite the contrary ; if anything, more bibliographers would be needed and a greater precision and higher quality of bibliographic work would be required.

The basic bibliographic needs from the standpoint of the researcher are :

a) A survey of all relevant current bibliography concerning his research problem.

b) The evaluation and selection of 'critical' sources, which in turn involves the following considerations :

1) Assessing the relevancy of a source in terms of data content (especially subject coverage).

2) Judging the compatability of data. This includes such factors as the type of unit analysed, the time period covered, the data level, the disciplinary orientation of the author, and so forth

3) Assessing the reliability of a source. In addition to mechanical errors, the problem of judging reliability requires the following kinds of information :—(a) the background of the author and of other members of a research team (professional, theoretical etc.) (b) the accessibility of the subject investigated, (c) nature of sources consulted (d) research conditions (e) research methods and research or other objectives.

4) Determining the physical availability of a source. This requires information on the physical location of a source and its circulation status.

5) Predicting the volume and nature of the research task. Among

those factors which must be considered are the number and size of relevant sources, the languages in which they are written, their accessibility and so forth.

It will be seen from the above points that if any substantial portion of this desired information is to be provided in readily usable form for a large volume of bibliographic items, computerization is essential. Furthermore, both the selective retrieval of bibliographic items and effective international co-operation also require a standardized computerization system. Also, the updating of bibliographies would be greatly facilitated, since no re-typing or retrospective proof reading would be required.

Although it is generally felt that computerization is more suitable for the national sciences and technological subjects, I feel that the H.R.A.F. (Human Relations Area Files) system devised at the Yale University U.S.A. can effectively cover the arts and humanities subjects also. The basic principle of this system is the use of ''fixed fields'', requiring that information entered on work sheets should follow a particular arrangement. Two types of information have to be distinguished : (a) descriptive bibliographic information—author, title, imprint, collation etc. and (b) analytical information i.e. the annotation of a source. The first category of information must be standardized. The second category must be standardized in certain aspects, whereas other aspects might be left open and capable of retrieval by use of a K.W.I.C. (Key Word in Context) system.

The H.R.A.F. worksheets to be filled out by bibliographers could be viewed as consisting of three broad sections :

1) The leader area—This consists of a series of eighty 'holes' or items to be punched, providing analytical information on the document at hand. This treats the national or ethnic identification of the people dealt with, the material type of the document, language of the title, country of publication and has spaces for a topical code, a geographical code, a temporal code and so forth. This area provides for a rapid and cheap retrieval of documents by these key items.
2) The second area deals with descriptive bibliographic information.
3) The third area dealt with analytical information. Here the descriptors are written out and a discursive annotation is given if possible and desirable.

In other words the H.R.A.F. system provides for a hierarchy of analytical categories from quite general to very specific descriptions, and a multi-dimensional annotation to provide for highly flexible retrieval. Another important point is that this type of worksheet allows for cumulative entries. Any given bibliographer needs only to enter the information which he feels competent to deal with. Finally a variety of types of physical print outs are possible, from a straight listing arranged under any desired headings to library card catalogue forms.

It is clear, I think, to all, that we stand on the brink of a much needed revolution in bibliographical work heralded by the arrival of a computer. Admittedly, much of the work that has been so far done by employing the computer for bibliography and ratrieval has been disappointing, and many difficulties have been faced by pioneers. At the same time major projects have now been firmly established. The largest of these are perhaps the Defence Documentation Centre and the Department of Commerce Clearing House for Scientific and Technical Information in Washington D.C. U.S.A. These have devised systems for retrieval, computerized bibliography and reproduction which are now working well, though not without certain difficulties. In the field of Asian bibliography, the U.S. National Service Foundation is sponsoring the transformation of the Bibliography of Asian Studies from a manual to a computerized system. The necessary hardware, in the shape of computers are now becoming part and parcel of the scene of even the developing countries in Africa and Asia and while installation and initial costs are high, long run potential benefits for researchers in African and Asian countries can be of great significance.

What is now needed is a standardized and understandable system through which bibliographical information may be compiled and fed into a central pool, with the possibility of dispensing of bibliographical tapes to all participants in such a continental programme. This first and vital step seems to be still lacking and has yet to be developed. Until that time, the present wastages of individual and un-cordinated compilation of bibliography and documentation will continue.

COMPUTERISED INFORMATION STORAGE AND RETRIEVAL SYSTEMS IN INDIA WITH SPECIAL REFERENCE TO THE ACTIVITIES OF THE BHABHA ATOMIC RESEARCH CENTRE

V. A. Kamath and N. M. Malwad
Library & Technical Information Section
Bhabha Atomic Research Centre, Bombay-85, India

SYNOPSIS

A survey of computers in various organisations and institutions in India is made and their present use for information handling, documentation and library procedures as well as the future plans of some of the institutions in this regard are briefly touched upon.

1 Introduction

The spectacular progress of science and technology during this century has inevitably resulted in the generation of tremendous amount of scientific and technical information in the entire spectrum of science. The pace of growth of this information, which is embodied in papers published in journals and presented at conferences, symposia etc., books, reports, theses, patents, specifications etc. has been such that it has become nigh impossible to cope up with it. Various approaches have been suggested from time to time, and methods and techniques adopted to handle this information effectively and expeditiously for the benefit of the scientific community. However, none of them can be said to be really satisfactory. At the moment, computer processing of information appears to be the most effective answer to this problem.

2 Scientific and Technological Developments in India

The scientific and technological progress in India, the largest democracy and the second most populous developing country in the world, since she achieved independence in 1947 has been extremely rapid. Table 1 shows the growth of universities and higher educational institutions from 695 in 1950-51 to 2, 899 in 1968-69. The number of scientists and technicians employed in the R & D establishments in the country rose from 18, 043 in 1958-59 to 62, 349 in 1968-69 as

Number of Universities, Institutions Deemed to be Universities under Section 3 of the UGC Act, 1956, and Constituent and Affiliated Colleges

Institution	1950-51	1955-56	1960-61	1965-66	1967-68
1. Universities	27	33	45	64	71
2. Institutions deemed to be universities	-	-	2	9	10
3. Colleges by type:					
(a) Arts, Science and Commerce	548	783	1,161	2,002	2,303*
(b) Engineering/ Technology	31	49	76	103	106
(c) Medicine	34	51	80	123	141**
(d) Agriculture	16	24	37	54	54
(e) Veterinary Science	7	14	18	20	21
(f) Law	22	27	40	70	66
(g) Teachers' Training	36	75	125	193	202
(h) Physical Education	1	2	5	7	7
Total (Colleges)	695	1,025	1,542	2,572	2,899

* Includes colleges of oriental learning (179) and music and Fine Arts (69)

** Includes colleges of modern medicine (100), ayurvedic and unani (21), pharmacy (4), dentistry (8) and nursing (8).

Source: India, Pocket Book of University Education. University Grants Commission, 1969.

Table 1

Total Number of Scientific/Technical Personnel Employed
in R & D Establishments

	1958-59	1968-69
(a) Major Organisations under the Central Government		
i) C. S. I. R.	3,512	8,848
ii) D. A. E.	1,067	7,209
iii) D. R. D. O.	1,500	4,747
iv) I. C. A. R.	1,500	7,820
v) I. C. M. R.	1,001	1,221
	8,580	29,845
(b) Other Central Ministries	5,663	15,593
Total Central Government	14,243	45,438
(c) Universities	2,600	7,778
(d) State Governments	1,000	6,900
(e) Private Sector	200	2,233
Grand Total	18,043	62,349

Note: (1) Data relates to information so far received in the CoST Secretariat and in some cases the figures are incomplete.

(2) The manpower for 1968-69 under universities, item (c) has been computed from U.G.C. data on the assumption that entire number of university professors and readers and 50% of the lecturers in universities and 10% of the senior teachers and 5% of lecturers in the affiliated colleges are engaged in the R & D work. The manpower for 1958-59 has been computed, assuming the same rate of growth approximately, as under the Central Sector.

(3) Manpower under the State Governments has been estimated at an annual expenditure of Rs. 10,000 per R & D worker, on ad hoc basis.

Source: India, Committee on Science and Technology : Report on Science and Technology 1969.

Table 2

indicated in Table 2. Among the leading government scientific organisations and institutions are the Department of Atomic Energy (DAE) in the field of nuclear science and technology; the Council of Scientific and Industrial Research (CSIR), with its 34 national laboratories devoted to research and development work in various fields such as physics, chemistry, mechanical engineering, etc.; 34 R & D organisations under the Defence Ministry and 26 institutes under the Indian Council of Agricultural Research. Some 900 scientific and technical periodicals are published in the country, besides a large number of books, theses, technical reports, patents, etc. Most of these organisations and institutions have library and documentation facilities for collecting, evaluating, organising and disseminating information in specific fields. The Indian National Scientific Documentation Centre (INSDOC), New Delhi, provides a full range of documentation services at the national level.

3 Computers in India

The computer age in India dawned with the establishment of computer centres at the Indian Institute of Technology (IIT), Kanpur, and the Tata Institute of Fundamental Research (TIFR), Bombay, during 1963-64. At present, there are some 210 computers installed in 110 organisations and institutions in 24 urban centres all over the country. The universities and research institutions have about 20 computers. The computers range from small size Honeywell-400 to large size CDC-3600. Most of these computers are of medium size, business-oriented and belong to the second generation, operating in off-line and indirectly coupled mode of operation. The CDC-3600 at the TIFR and the IBM-7044 at the IIT, Kanpur, are the only two large size second generation computers in the country. There are a number of plans by various organisations and institutions, both government and private, to install several third generation computers during the next few years.

4 Use of Computers for Information Processing

A brief account is given below of the result of a survey conducted by the authors on the application of computers to various library routines and documentation work in the country. For this purpose, a questionnaire, as per Appendix I, was sent to librarians and documentalists of institutions and organisations having computer facilities of their own or access to them. The results of the survey are summarised below.

4.1 Indian National Scientific Documentation Centre (INSDOC), New Delhi

The Indian National Scientific Documentation Centre was established in 1952 by the Council of Scientific and Industrial Research, Government of India, with technical assistance from the UNESCO to provide a full range of documentation services to Indian scientists and technologists, including reprographic, translation and bibliographic service. The INSDOC gives advanced training in documentation and reprography and devotes a full term to the subject of computer utilisation for information processing. Basic facilities for punching, verifying and sorting punched cards are available at the Centre and the IBM 1620 computer facility at the Delhi School of Economics is hired for processing the data. With these facilities, the INSDOC has developed programmes for the preparation of "Union Catalogue" of periodical publications available in Indian libraries, preparation of a roaster of scientific and technical translators in the country, preparation of monthly author and subject indexes and cumulated annual author and subject indexes for its monthly abstracting periodical, "Indian Science Abstracts", organising a SDI service in the field of electronics and for data processing.

4.2 Documentation Research and Training Centre (DRTC), Bangalore

The Documentation Research and Training Centre was established in 1962 by the Indian Statistical Institute, Calcutta, under the able and inspiring guidance of Dr. S.R. Ranganathan to give intensive training to documentalists in research methods in documentation, and to give a course of training to documentalists for service in India and also to extend the training facilities to documentalists from other countries, particularly the developing countries of Asia and Africa. As one of the research activities, the DRTC has developed non-conventional methods in document finding, with emphasis on incorporating faceted scheme of classification such as the Colon Scheme of Classification of subjects. The results of the developments are summarised from time to time in its official periodical, "Library Science with Slant to Documentation". The Centre has developed a programme-package in Autocoder language for IBM-1401 computer system which consists of 15 programmes for use in the design, development and operation of a system for document finding and integrates within it a freely faceted scheme of classification of subjects. The system aims at serving with SDI, retrospective searches, subject bibliographies and so on.

4.3 Physical Research Laboratory (PRL), Ahmedabad

The Library of the Physical Research Laboratory, with the co-operation of the computer programmers of the Laboratory, has developed a system for mechanical indexing which selects, matches with insignificant words and prepares a key-word out-of-context (KWOC) index, with titles and abstracts of documents, along with the bibliographic citations as input. Programmes have been developed in SPS for IBM-1620 computer. The PRL is likely to instal an IBM-360/64 computer by the end of 1971 and its library has planned to develop a comprehensive system for information storage and retrieval using this computer.

4.4 Tata Institute of Fundamental Research (TIFR), Bombay

The Tata Institute of Fundamental Research is the national centre for advanced studies and fundamental research in nuclear science and mathematics. The Institute has three computers in-housed, namely, CDC-3600, CDC-360A and a recently installed PDP-11/20 on-line digital computer with a 16K 16 bit word memory. Its library uses the computer system CDC-3600/360A for preparing its monthly addition lists of books and documents, and to control the periodical publications it receives. The programmes have been developed in FORTRAN IV.

4.5 Indian Institute of Technology (IIT), Kanpur

The Indian Institute of Technology, Kanpur, has one of the most developed computer facilities in the country which is also the oldest. It has two computers, namely, IBM-1620 with 40K characters memory and IBM-7044 with 32K words memory. Its library, which has a collection of 1,12,000 books and bound volumes of journals, utilises the facilities for indexing and preparing addition lists. Programmes have been developed in FORTRAN IV and are being made available to interested organisations and institutions.

4.6 Bhabha Atomic Research Centre (BARC), Trombay

The Bhabha Atomic Research Centre is the national centre for research in and development of atomic energy for peaceful purposes. The information needs of about 6,500 scientists, engineers and technicians of the Centre are met by its Library and Technical Information Section (L & TIS). The Library services are extended also to the sister institutions under the Department of Atomic Energy (DAE) and other institutions and organisations in the country so far as their requirements in the field of nuclear science and technology are concerned. The L & TIS comprises various units formed on functional basis as illustrated in the organisation chart vide Appendix 2. The Library, which has an area of 22,000 sq. ft., has some 63,000 books and bound volumes of journals in the nuclear field. It subscribes to some 1,200 scientific and technical journals. Its Depository Unit has a collection of about 190.000 scientific and technical reports of the various atomic energy organisations in the world, procured mostly on the basis of bilateral publications exchange agreements, and annually some 16,000 new reports are added to it. The reports are both in full size and microform i.e. microfiche, microcard and microfilm. The micro-reading facility of the L & TIS has several micro-readers and reader-printers. The reprographic and photographic facilities include Rank Xerox 1824 Printer and Rank Xerox 1385 Master Maker. Besides providing full size copies of required micro-documents to scientists, the facilities, in conjunction with the printery, are also used for bringing out the technical reports of the BARC and the proceedings of symposia, conferences, seminars etc. held under the auspices of the DAE. The documentation activities include the current awareness service entitled "Bibliography of Current Reports", list of additions to the Library, retrospective bibliographies on request and in anticipation, SDI etc.

5 Computerisation Programme of the Library and Technical Information Section

5.1 Need for Computerisation

So far only conventional methods have been used by the L & TIS. However, in view of the rapidly increasing amount of nuclear information available, the diversity of research and development programmes of the Indian Atomic Energy Commission, India's participation in the International Nuclear Information System (INIS) sponsored by the International Atomic Energy Agency (IAEA) and the contemplated computerised national information service in the field of electronics, it has been decided to embark on a phased programme of computerisation of some of the documentation activities and the library procedures.

5.2 Facilities

The L & TIS uses two computer facilities, namely, the in-house Honeywell-400 with 4K memory and CDC-3600 at the TIFR with 32K memory. There are plans to instal a large size general purpose third generation computer in the BARC which will be used by the L & TIS. A detailed study is now underway. in consultation with computer experts of BARC and TIFR, to asses the feasibility of various contemplated computer-based applications for information processing.

5.3 Work done

So far computer programmes have been developed for preparing a sound general data base on magnetic tape for the holdings of the Library, which comprises technical reports, journal articles, books, standards, specifications, proceedings of symposia, conferences etc., preparing KWOC indexes, author and report number indexes, preparing overdue reminders to be sent to the borrowers of books/reports and preparation of monthly list of additions to the Library. The programmes have been developed in FORTRAN IV and COBOL languages for the hardware configuration of the available facilities. The work done in this field by the L & TIS cannot of course be compared with the already developed IR systems on large size computers in advanced countries. However, it is a modest effort to provide a base for the future growth.

5.4 Future Plans

The future plans envisage the following: i) Preparation of a master file of all documents in the L & TIS with full bibliographic descriptions ii) Up-dating of the master file from time to time iii) Compilation of retrospective bibliographies on request and in anticipation on various scientific topics using the master file iv) Development of selective dissemination of information (SDI) and current awareness services (CAS) v) Computerisation of the following library routines and services, namely, (a) registration and control of periodicals; further, a variety of indexes by subject, title, frequency, language, publisher, country of

origin etc. will be produced as print-outs (b) procurement of documents etc.
vi) Processing of tapes from INIS, NSA and other computer-based information systems in the world.

5.5 INIS Activities

The L & TIS is solely responsible for discharging, both nationally and internationally, all the obligations accruing to India's participation in INIS, the computer-based information system sponsored by the IAEA. Its principal tasks in this regard are: i) the procurement of all available scientific and technical literature produced in India falling within the initial (later on, the final) subject-scope of INIS, ii) thoroughly processing the documents thus collected and preparing the inputs for INIS in accordance with its specifications and transmitting them to INIS as per the time schedule, iii) speedy processing of INIS outputs when received and disseminating them, in various forms, to the maximum possible extent, to universities, national laboratories, scientific institutions and organisations, and active scientific workers in the country. The input to INIS has to be preferably on computer readable input media such as magnetic or paper tapes and the output on the same media. However, to enable the maximum utility of the INIS to the countries wherein computer facilities do not exist or are not the same as are available at INIS, Vienna, at present INIS accepts input on full-size work-sheets and sends its output in full-size conventional form as INIS Atomindex. Due to the limited hardware configuration, the L & TIS has not been able to send its input on computer readable media and also could not process the output on magnetic tape so far. This difficulty has been brought to the notice of INIS authorities. The same difficulty has also been encountered by a few other countries. It has now been decided that the INIS authorities will give due consideration to this problem and send the INIS output tapes in six-bit-code on seven-track magnetic tapes instead of the eight-bit-code on nine-track magnetic tapes. This will help the L & TIS to use the output tapes by developing a matrix in computer memory to recognise the INIS 120 character set to produce INIS Atomindex at local level, to compile retrospective subject bibliographies and produce documentation lists. It will also be possible to send Indian input on magnetic tape using the available facilities.

6 Assessment of Computerisation in India

Apart from major institutional libraries already mentioned, computers have been installed in several other institutions and organisations where the librarians and documentalists have not yet utilised the computers. The Indian Standards Institution, which has about 300,000 specifications, codes of practices etc., has not yet utilised the computer for its IR system. It should be noted that any plan for computerisation in India has to take into consideration the availability of vast manpower, including technically qualified one, and the obvious need for its utilisation. It is imperative, both from the economic and political considerations, that computerisation should not result in any substantial reduction in the number of personnel already employed or required to be employed in the future. This is

Computerisation of Information Processing in India

Procedures	Number of Libraries	Computor/s used	Software
Library Routines			
a) Procurement	1	CDC-3600	FORTRAN IV
b) Charging and Discharging	1	ICL-1909	--
c) Cataloguing	2	CDC-3600, H-400, IBM-1401	FORTRAN IV, COBOL, AUTOCODER
d) Addition lists	4	IBM-1620, IBM-7044, CDC-3600, H-400, IBM-1401	SPS II D, FORTRAN IV, AUTOCODER
e) Periodical holdings	2	CDC-3600, H-400	FORTRAN IV, COBOL
f) Union catalogue	1	IBM-1620	SPS II D
Documentation Activities			
a) Indexing	5	IBM-1620, IBM-7044, CDC-3600, H-400, IBM-1401	FORTRAN IV, COBOL, SPS II D, AUTOCODER
b) Bibliographies	1	IBM-1401	AUTOCODER
c) SDI	1 (2 systems are under development)	IBM-1401, CDC-3600	AUTOCODER, FORTRAN IV
d) Directory (Roaster of technical translators in India)	1	IBM-1620	SPS II D

Table 3

perhaps one of the reasons for not computerising information processing at several organisations in the country. It is clear from the Table 3 summarising the various computerised activities in the country that there is only one library having computerised procedures for procurement, one for charging and discharging, two for cataloguing, four for preparing additional lists and one for preparing union catalogue of periodicals. Among documentation activities, indexing seems to be more common though the criteria for indexing are different. There exists one general computer system for information storage and retrieval integrating within it a freely faceted scheme of classification such as the Colon Scheme of Classification developed by Dr. S. R. Ranganathan. The software techniques have been developed in FORTRAN IV, SPS, AUTOCODER, and in certain cases COBOL. The programmes are meant mainly for the configuration available at the institutions. The libraries make use of the spare capacity of many computer facilities which are suitable for the type of work on off-line use.

7 Educating the Librarians and Documentalists

The INSDOC and the DRTC devote a full term of their advanced training courses for computer applications in libraries and documentation centres to study, develop designs and evolve mechanised storage and retrieval systems suitable for local conditions. The INSDOC organises refresher courses on subjects of interest in fields such as file organisation, programming languages, systems analysis etc. To asses existing computerisation of documentation activities and library procedures, the INSDOC organised a 'Seminar on Automation Problems in University Libraries and Special Libraries' from May 5-7, 1969 at Delhi. The IAEA held a 'Regional Seminar for Asia and the Far East on INPUT Preparation for INIS' at BARC at the invitation of the Department of Atomic Energy from November 23 to December 11, 1970, with the object of giving the necessary training to personnel engaged in INIS work in the preparation of their national inputs. These training courses and seminars will go a long way in building a computer-based national IR system in not too distant future.

8 Computer-based National IR System

There is, in fact, duplication of efforts due to non-existence of a central body for co-ordinating the software developments in the country and disseminating the information on software techniques already developed in other countries. The need for one large computer-based national system, though felt for a long time, could not be fulfilled so far on account of hardware limitations of the existing computers in the country. However, it is expected that within a period of two years, large size IBM-360 series, CDC, and ICL-1900 series computers will be available in the country, making it possible to have a national IR system.

The INSDOC being the national documentation centre should perhaps take the lead: i) To identify the areas for computerisation and orient its documentation services towards designing and developing a computer-based national IR system

ii) To assess its feasibility and applicability in other institutions in the country taking into consideration (a) large variations in the hardware configurations and software techniques as the institutional libraries make use of the spare capacities of the computer facilities (b) cost involved as compared to the additional services to be offered (c) availability of manpower iii) To standardize the bibliographic description and its format on computer acceptable input/output media such as magnetic tapes, paper tapes etc. in the light of the recent progress in this field achieved by international organisations such as ISO, INIS etc. iv) To co-ordinate the indigenous software techniques developed for information processing and library routines at institutional levels v) To collect and disseminate information on the current developments in computer technology and their applications in the field and vi) To educate the documentalists and librarians for active participation in the national IR system and necessary developments required at the institutional level.

9 Conclusion

India's steadily accelerating pace of industrialisation, rapidly increasing literacy and the national urge to catch up with the advanced countries have given a great fillip to the adoption of modern techniques and methods in several fields of enterprise. Computers are gradually assuming importance as one of the most indispensable modern tools for enhancing efficiency and speed in various operations and their number in the country has already exceeded two hundred. Besides, computers are being produced indigenously and it is expected that within a decade the most sophisticated ones will not only be manufactured but also exported.

While computers will undoubtedly find ever increasing applications in various fields in India, a good beginning has already been made in their use for information processing and documentation activities as enumerated in this paper and it can be confidently predicted that with the increasing tempo of the scientific and technological progress of the country and the consequent need for the expeditious dissemination of scientific and technical information generated in India and all over the world, the computers will play a very crucial role in this field.

Use of Computer for Scientific Documentation and Library Routine Jobs : Survey

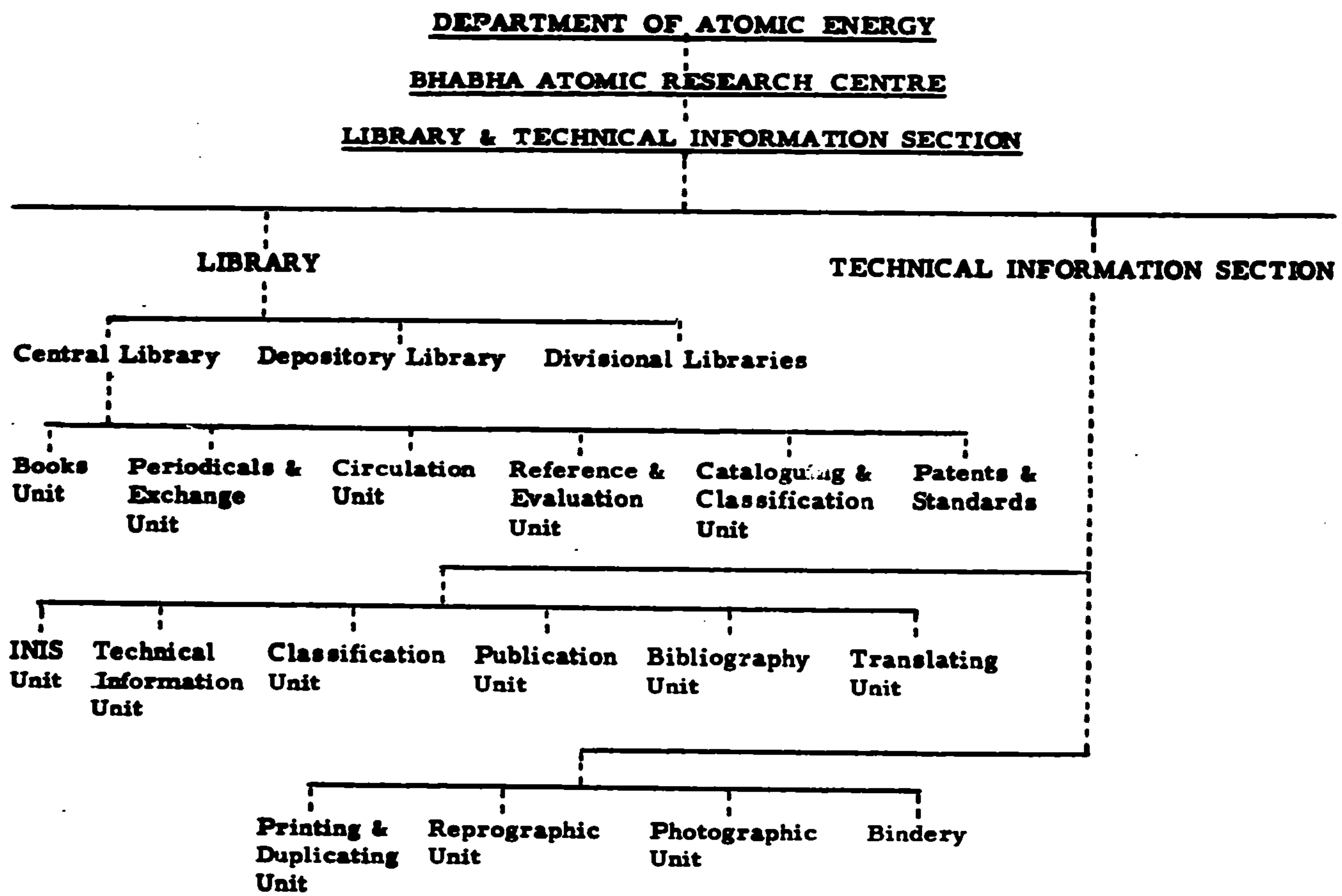
(Please Print or Type)

1. **Name of the institution :**
2. **Address :**
3. **Library collection :**
4. **Is the library using computer for mechanisation of its procedures?** **YES/NO**
5. **If yes, computer used :**

   Configuration
   - i) **Memory capacity :**
   - ii) **Tape drives :**
   - iii) **Discs :**
   - iv) **Reader/Punch :**
   - v) **Mechanical sorter :**
   - vi) **Other :**
6. **Documentation activities that have been computerised**
   - i) **Indexing :**
   - ii) **Bibliographies :**
   - iii) **SDI :**
   - iv) **Other :**
7. **Library routine jobs that have been computerised**
   - i) **Procurement :**
   - ii) **Charging/Discharging :**
   - iii) **Cataloguing :**
   - iv) **Addition lists :**
   - v) **Other :**
8. **Computer time used per month :**
9. **Software details :**
10. **Are the programmes available for those interested in applications ?** **YES/NO**

Appendix I

Organisation Chart of L & TIS



Appendix 2

FORSCHUNG IN DEN INFORMATIONSWISSENSCHAFTEN: INTERNATIONALE VERFÜGBARKEIT VON FORSCHUNGS-VORHABEN UND FORSCHUNGSBERICHTEN.

Prof.Dr.Helmut Arntz
Vice-Président, Fédération Internationale de Documentation.

SYNOPSIS

Recommendations of the UNESCO Consultation Groupe on the Promotion of Research in Documentation (1970 and 1971). UNESCO's Computerized Documentation Service (CDS). FID Research Referral Service (FID-RRS). International Information System on Research in Documentation (System ISORID) worked out by FID under UNESCO contract. Development in UNESCO of a computerized system for collecting, processing and disseminating information on research projects and research reports in the information sciences. Building-up in the UNESCO library of a research reports collection. Joint action of UNESCO and FID.

1. Ein nationales Beispiel: VNTITS.

Da die internationalen Planungen durch das sowjetische "All-Unions-Zentrum für Wissenschaftliche und Technische Information" vielfach angeregt wurden, soll VNTITS, 1966 gegründet und dem Staatskomitee für Wissenschaft und Technik des Ministerrats der UdSSR unterstellt, kurz geschildert werden.

Das Institut beschränkt sich nicht auf Forschung in den Informationswissenschaften. Alle Institutionen, in denen reine oder angewandte Forschung[1] durchgeführt wird, sind meldepflichtig; das betrifft z.B. die Ministerräte der Republiken, Akademien der Wissenschaften, Höhere Lehranstalten ebenso wie Betriebe. Die Meldungen werden auf dafür entwickelten Formblättern direkt, das heißt unter Nichtbeachtung eines sonst vorgeschriebenen Dienstwegs, an VNTITS gegeben. Wenn eine Arbeit mehrere Organisationen oder Einrichtungen betrifft, sendet jede ein Blatt über ihren Anteil und der Federführende ein Generalblatt über das Vorhaben ein. Wird ein Vorhaben unterbrochen oder abgeändert, ist erneute Meldung erforderlich. Wesentliche Quellen sind die zentralen Informationsinstitute der einzelnen Industriezweige.

1) Forschung hier wie im folgenden = F & E.

VNTITS veranlaßt die Eintragung in das staatliche Register und sendet innerhalb von fünf Tagen eine Kopie des Themenblatts zurück; innerhalb eines Monats dann zusätzlich alle Angaben, die dort bereits über ähnliche Forschungen oder Vorhaben vorliegen. Nach Abschluß der Forschungsarbeit ist der Forschungsbericht an VNTITS zu senden; falls kein Bericht erscheint, eine Zusammenfassung der Ergebnisse.

VNTITS ist genau genommen nur buchführende Stelle; es nimmt keine Informationsbearbeitung (z.B. Referierung) vor. Die Registrierung verleiht auch keine Prioritätsansprüche hinsichtlich Erfindungen usw., sondern dafür ist der übliche Weg einzuhalten. Schließlich bewahrt VNTITS die Forschungsberichte nur in Form von Mikrokopien auf. Die Originale werden an die Informationsinstitute der betreffenden Industriesparten bzw. Akademien geschickt. Aus dem Material werden aber Berichte über F & E auf bestimmten Gebieten zusammengestellt und veröffentlicht, statistische Informationen geliefert und Anfragen von Institutionen (aber nicht von Einzelpersonen) beantwortet.

Es ist nicht bekannt, wie weit der Auftrag erfüllt werden kann, alle Nachrichten über F & E-Projekte im Ausland zu sammeln; aber die Forschungsvorhaben und Forschungsberichte der Sowjetunion dürften lückenlos erfaßt werden. Ein voll mechanisiertes Informationssystem ist in Vorbereitung.

2. UNESCO-Beratergruppe für Dokumentationsforschung 197o.

UNESCO-DBA (Department of Documentation, Libraries and Archives; Direktor O.A.Mikhailov) hat im Mai 197o eine Sachverständigengruppe zur Frage der "Förderung der Forschung in den Informationswissenschaften[2]" zusammengerufen.

An erster Stelle stand die Empfehlung an die 125 UNESCO-Mitgliedstaaten und die internationalen Organisationen, regelmäßig Listen und Kursbeschreibungen aller laufenden und für die Zukunft geplanten Forschungsvorhaben sowie Listen der (veröffentlichten oder unveröffentlichten[3]) Forschungsberichte mit Referaten in einer oder mehreren der offiziellen UNESCO-Sprachen (E, F, R, S) einzureichen; außerdem eine Kopie jedes Forschungsberichts mit der Erlaubnis, sie auf Anforderung Dritten zugänglich zu machen. Zustandsberichte über ein Gebiet und Forschungsübersichten sollten Forschungsberichten gleichgestellt sein.

UNESCO (in Zusammenarbeit mit FID, soweit sachdienlich) sollte drei internationale Register schaffen: Forschungsvorhaben; Forschungsberichte; Liste der in Kopie vorhandenen Berichte. Durch laufende Neuheitsberichte und Nachsuche auf Auftrag sollte die hinter den drei Registern stehende Information weiter aufgeschlossen und Interessierten zugänglich gemacht werden.

2) In der UNESCO-Terminologie "Dokumentation": "Documentation is used here as shorthand for documentation, librarianship and archives as well.-"

3) "unpublished" meint alle nicht über den Buchhandel beziehbaren Berichte.

3. UNESCO-CDS.

Die Expertengruppe stellte 1970 schon in Rechnung, daß UNESCO über ein maschinelles Dokumentationssystem verfügen würde; denn die UNESCO General Conference hatte auf ihrer 42.Vollsitzung am 20.11.1968 die Entschließung angenommen, daß "there should be greater attention to the mechanization and automation of information processing to the point where UNESCO's own documentation services become a 'pilot project' for the demonstration of, and training in, modern equipment and procedures; with due attention to audio-visual materials" (15 C/Res.10, Teil IV D). Demgemäß empfiehlt das IACOD[4]-Protokoll vom 19. - 22.8.1969: "computerization, at the earliest opportunity, of UNESCO's own documentation and publication activities, extending subsequently to all holdings and acquisitions of documents and publications, irrespective of origin" (COM/IACOD/69/11, S.6.). Die Vorarbeiten beruhten hauptsächlich auf Erfahrungen anderer Organisationen der United Nations Family (ILO, FAO, IAEA, UNIDO).

Anfang 1970 wurde entschieden, daß mit den Vorarbeiten zur Systemanalyse, Systemplanung und Rechnerprogrammierung mit dem Ziel begonnen werden sollte, den CDS (Computerized Documentation Service) in der zweiten Hälfte 1971 arbeitsbereit zu haben. Man entschied sich für ein verbessertes Marc II aus Gründen der Flexibilität in der Anwendung und der Verträglichkeit mit bestehenden maschinellen Systemen (Internat.Arbeitsorganisation, Kongreßbibliothek, Britische Nationalbibliographie). Systemanalyse und Systemplanung wurden bei UNESCO durchgeführt; die Rechnerprogramme wurden im Kontrakt von einer Programmierfirma in London geschrieben.

Noch nicht gelöst ist das Ordnungsproblem; denn die auf der Aligned List of Descriptors[5] aufgebaute UNESCO List of Descriptors entspricht zwar weitgehend den Bedürfnissen im Erziehungs-, Wissenschafts- und Kultursektor, wie ein 1969 durchgeführter Versuch an 500 eigenen UNESCO-Dokumenten bewies; Dokumentation und Informationswissenschaften insgesamt sind aber in dieser Liste so gut wie unberücksichtigt. Die Empfehlungen der Expertengruppe vom Juni 1971 (siehe unten) sehen daher vertiefte Arbeit an den Deskriptoren mit dem Ziel vor, am Ende zu einem international annehmbaren Thesaurus der Informationswissenschaften zu gelangen (unter Einbeziehung des für Ende dieses Jahres erwarteten VINITI-Thesaurus, des ebenfalls bis Ende 1971 im UNESCO-Kontrakt zu erstellenden Vocabularium Documentationis usw.).

Planstellen für Indexierer sind inzwischen bewilligt worden. Eine andere Gruppe wird außerhalb von UNESCO unter Vertrag die Dokumente, Monographien und Zeitschriften, die von CDS verarbeitet werden, bibliographisch erfassen

4) International Advisory Committee on Documentation, Libraries and Archives.

5) Gemeinsame Deskriptorenliste von OECD, FAO, ILO und DSE (Deutsche Stiftung für Entwicklungsländer).

und indexieren (soweit dies nach den derzeitigen terminologischen Unterlagen möglich ist). Der Plan geht dahin, in dem ersten Betriebs-Biennium 30.000 Dokumente und Veröffentlichungen zu verarbeiten; das wären alle laufenden UNESCO-Dokumente und -Publikationen zuzüglich einer Auswahl von etwa 6.000 "retrospektiven" aus dem Bestand.

Es lag nahe, die Erfassung durch den CDS so früh wie möglich auf die Erfassung aller Forschungsprojekte und -berichte im Bereich der Informationswissenschaften auszudehnen; doch war es den Experten unwahrscheinlich, daß dies unter fachlichen Gesichtspunkten in UNESCO geleistet werden könne.

4. FID-RRS.

Die Tatsache, daß FID bereits seit 1967 an der Einrichtung eines Informationsdienstes über Forschung im Bereich der Informationswissenschaften arbeitete, war ein wesentlicher Faktor der frühzeitig ins Auge gefaßten Partnerschaft beider Organisationen. Der Research Referral Service der FID (RRS) nahm seine Tätigkeit, das heißt die Sammlung und Verbreitung von Informationen über einschlägige Forschungsvorhaben, 1970 auf; im einzelnen bis Juni 1971:

4.1. Anschriftenliste von 600 Einrichtungen (Bibliotheken, Dokumentations- und Informationszentren usw.), an denen F & E-Vorhaben durchgeführt werden.

4.2. Umfassende Durchsicht der Literatur (Monographien, Primär- und Sekundärzeitschriften, Bibliographien, Listen, unpublizierte Quellen), um Hinweise auf F & E zu finden.

4.3. Fragebogen an Einrichtungen und Personen, soweit sie mit F & E-Projekten befaßt sind; durch die Fragebogen wurden Daten über 250 Projekte erfaßt.

4.4. Daten über rund 750 Projekte aus anderen Quellen.

4.5. Kartei von Nachweisdaten über rund 1.000 Projekte aus mehr als 30 Ländern. (Die Kartei enthält Vorhaben beginnend mit Anfang 1969; das Verhältnis der laufenden zu den inzwischen fertiggestellten ist etwa 60 : 40).

4.6. Veröffentlichung eines monatlichen Dienstes "R & D Projects in Documentation and Librarianship".

4.7. Ankündigung laufender Auskunftsdienste auf Anfrage seitens der Fachinteressenten.

Ob die angegebenen Zahlen repräsentativ sind, wird weitgehend davon abhängen, mit wievielen Forschungsvorhaben im Bereich der Informationswissenschaften gerechnet wird. Die Sachverständigen fanden sich vor gänzlich abweichenden Zahlen, die mutmaßlich großenteils zu hoch gegriffen sind. Nachprüfbare Angaben sagen folgendes aus:

"Current Research and Development in Scientific Documentation", herausgegeben von der National Science Foundation - Office of Science Information Services (1969 mit Nr. 15 eingestellt), brachte jedes zweite Jahr Unterlagen über 600 bis 800 Forschungsprojekte.

Die "List 1971" der Science Associates Inc., Maryland University, enthält 820 Vorhaben. (Diese Liste, die ab 1971 jährlich veröffentlicht werden soll, deckt wesentlich das Jahr 1970 ab).

Das Yearbook der Library Association of the U.K. enthält (nur für das Vereinigte Königreich) 300 Forschungsprojekte und -berichte.

"Research in Education" (Reihe LI), herausgegeben vom ERIC Clearinghouse "Information and Library Sciences", monatlich erscheinend, enthält etwa 360 bis 430 Forschungsberichte.

Die Referateorgane des Gebiets (Bulletin Signalétique Nr.101; Information Science Abstracts; Library and Information Science Abstracts; Referativnij Žurnal "Informatika") referieren jährlich ein paar hundert Forschungsberichte bei insgesamt 3.000 bis 4.000 Referaten je Zeitschrift und Jahr.

Daraus scheint sich zu ergeben, daß die hoch in die Tausende sowohl für Forschungsvorhaben wie Forschungsberichte reichenden Schätzungen überhöht sind - so wünschenswert es wäre, dem hohen finanziellen Aufwand adäquate Forschungstätigkeit zu sehen. Es wird geschätzt, daß 5 bis 8 % der in der Welt für F & E insgesamt ausgegebenen Mittel auf F & E in den Informationswissenschaften entfallen, und daß in den USA von 700 Millionen Dollar, die 1966 für wissenschaftliche und technische Informations- und Datendienste ausgegeben wurden, 60 Millionen Dollar für die Verbesserung dieser Dienste aufgewendet wurden.

5. ISORID.

Im August 1970 schloß UNESCO mit FID einen Vertrag, der FID folgende Aufgaben stellte:

"1. Detailed instructions for national documentation centres or other institutions designated for this work:
(a) on the selection of reports, specifying the field which is covered by the term "documentation" including library and archives activity (e.g., problems of classification, indexing, as well as mechanization and automation of documentation processes, etc.);
(b) on the form of report abstracts (with descriptors) to be submitted to UNESCO's CDS with a view to ensuring uniformity.

2. Guidelines for CDS operations using the above mentioned reports, including the preparation of lists of descriptors as well as mechanised lists of titles and abstracts."

In der folgenden Korrespondenz wurde der Auftrag auf research projects (completed, in progress, or planned) erweitert.

FID lieferte im April 1971 die Studie "Project ISORID" (International Information System on Research in Documentation; Study on the application of mechanized methods for the retrieval and dissemination of research information and documents in the field of documentation)[6] mit folgender Unterteilung:

5.1. Allgemeine Beschreibung des Informationssystems über Informationsforschung (einschließlich der Aufgabenverteilung zwischen UNESCO, FID und IFLA).

5.2. Elemente der UNESCO - Verlautbarungen an die Mitgliedstaaten (z.B. Umfang der Erfassung; Definition der in der Systemplanung benutzten Termini; Datenerfassung für die Eingabe; Zugang zu Forschungsberichten).

5.3. Die Tätigkeit der nationalen Institutionen, die für die Förderung und Koordination der Dokumentationsforschung verantwortlich sind (z.B. nationale Register; Auswahl laufender Vorschungsvorhaben und von Forschungsberichten für ISORID; Berichte über laufende F & E; Aufnahme und Referierung von Berichten; Identifizierungsnummern (nationale, die später durch einheitliche ISORID-Nummern auf der Grundlage der vorausgehenden Projekt-Nummern ersetzt werden).

5.4. Systembetrieb (genormte Formblätter zur Übermittlung von Information über Forschungsvorhaben; Erklärung der aufgenommenen Angaben; Indexierung und Klassifikation der Projekte; manuelle und maschinelle Bearbeitung der Daten für die Eingabe; der CDS von UNESCO).

5.5. Ausgabedienste - Informationsverbreitung (Internationale Register von Forschungsvorhaben und Forschungsberichten; Nachweis- und Referatedienste; Schaubild der Arbeitsabläufe in den ISORID-Zentren; Beschaffung von Kopien von Forschungsberichten; die potentiellen Benutzer von ISORID.)

Die im Juni 1971 bei UNESCO tagende Expertengruppe hatte nicht die Zeit, das Dokument in allen Einzelheiten zu studieren. Sie billigte aber im Prinzip ISORID als Rahmen für das künftige Tätigwerden von UNESCO für den Aufbau des Systems (unten 6.). Diese Billigung schließt auch die Empfehlungen an die nationalen Institutionen ein, die im ISORID - Projekt vorgesehen sind, auf nationaler Ebene Informationen über F & E in den Informationswissenschaften zu sammeln und zu liefern; ferner die Technik der Informationsbearbeitung und die Zusammenarbeit mit FID und IFLA.

Gleichzeitig empfahlen die Experten, die Kriterien der Informationsauswahl für das geplante System flexibel zu halten, bis von seinen Benutzern genügend Erfahrungen vorliegen. Auswahlprobleme werden erwartet z.B. hinsichtlich der Aufnahme von Grenzgebieten, der Abgrenzung zwischen Praxis und F & E sowie der Festlegung von Prioritäten.

6) Als Manuskript vervielfältigt; 33 S. und 6 Anhänge.

Die Beratergruppe billigte auch, dass UNESCO alle Information, die sie über Forschungsvorhaben und -berichte erhält, in Kopie an FID gibt, damit FID den Referral Service (FID-RRS, oben 4.) sowohl über Projekte wie über Berichte betreiben kann. Daraufhin hat FID sich verpflichtet, UNESCO mit allen Informationen zu beliefern, die als zusätzliche Materialien für UNESCO-CDS notwendig sind. Weitere Einzelheiten der Zusammenarbeit von UNESCO und FID sollen noch ausgehandelt werden.

6. UNESCO-Beratergruppe für Dokumentationsforschung 1971.

Die vom 22. bis 24. Juni 1971 bei UNESCO tagende Beratergruppe (Consultation Group on the Promotion of Research in Documentation) gab zwölf Empfehlungen ab. Die ersten drei enthalten die grundsätzliche Aufforderung, UNESCO möge in enger Zusammenarbeit mit den internationalen und nationalen Fachorganisationen unverzüglich das geplante System für die Sammlung, Bearbeitung und Verbreitung von Information über Forschung und Entwicklung im Gebiet der Informationswissenschaften aufbauen. Die Experten sehen darin zugleich einen wichtigen Baustein zum UNISIST-Programm. UNESCO soll die Mitarbeit der 125 Mitgliedstaaten sowohl für den Aufbau und Ausbau des Systems wie bei der Beschaffung des Eingabematerials (Information über Projekte und Berichte sowie Kopien dieser Berichte) erbitten und eine Spezialsammlung aller erhältlichen Forschungsberichte in der UNESCO-Bibliothek aufbauen.

Alles bei UNESCO gesammelte Material soll auf Anforderung Mitgliedstaaten, interessierten Institutionen und Spezialisten zugänglich sein. Dabei wurde nicht an das Rohmaterial (Meldebogen der Mitgliedstaaten) sondern die Ergebnisse der maschinellen Bearbeitung gedacht; insbesondere an laufende Register über Projekte und Berichte und Verzeichnisse der in Kopie vorhandenen Reports.

Das gesammelte Material soll zugleich UNESCO selbst die Möglichkeit geben, sich auf Prioritätsgebieten durch Vergabe von Kontrakten steuernd zu beteiligen, wobei immer stärker auch die eigene Sammlung der Forschungsberichte als Arbeitsgrundlage dienen wird. Die Untersuchungen sollen sich in einem weit gespannten Rahmen vollziehen, der zum Beispiel Entscheidungsmodelle, Terminologie, Relevanzkontrolle, Rentabilität, Bewertung, Benutzerprobleme usw. umspannt. Die Expertengruppe hat gemeinsam mit UNESCO-DBA auch die Thematik der drei weiteren Konsultationen festgelegt, die 1971/72 noch stattfinden und das schrittweise Anlaufen von UNESCO-CDS begleiten sollen.

Session One - Discussions

INTERNATIONAL AND NATIONAL INFORMATION NETWORKS
Chairman: Mr. W.K. Lowry (US)

CHAIRMAN (Mr. Lowry): In a way this particular session serves as the backdrop to all the other sessions of this Conference. The establishment of national and international networks depends greatly on subjects to be discussed at this Conference. Our ultimate goal is that in time and with good will we shall indeed have national and international information networks.

There are some rather new, but at the same time, old factors that influence the development of networks of any kind, including information networks. Economics is certainly one of the most prominent, and the whole idea of establishing networks is at least in part directed to avoiding redundant efforts.

Similarly, in the past twenty or thirty years, the advances made possible by technology are a factor. I refer particularly to technology associated with computers and with communication, which is especially pertinent to the information systems we are developing today. We now have new research tools available; not only library science, which of course is still very essential, but tools such as systems analysis, which is becoming more and more important in all fields of endeavor.

Another factor that is becoming more important to all of us is the emergence of new countries. Information networks can prove beneficial to their advancement.

I think of information networks from my particular standpoint as analogous to telephone networks. There are some basic requirements for network design that any information network must pay heed to. Among the participants in network operations there must be common agreement on what the objectives are, what technical design standards and what financial support will be necessary. Not the least important is the level of competency to make that worthwhile.

There must be compatibility and interchangeability in the components of any network. If it is going to be continually useful, there must be elements of control to keep the network on the track. To do this feedback mechanisms must be part of it.

We are emerging into a new international effort called UNISIST, which if successful, will be made up of modules which might well be national systems and it is only reasonable in the development of information networks that you develop your national systems with some regard for the development of other national systems. Unless this is done it is not easy for any international system to superimpose order over disorder. The aim of this Conference is to encourage cooperation on a national level which may in time result in an international network.

DR. B.W. ADKINSON (US): Since the exchange of information is such a large and complex problem, international efforts should be concentrated in two areas to start with. First, there should be the development and acceptance of a few basic technical standards. One of the most important concerns bibliographic citations where because everybody now uses his own, there is complete chaos today. Handling them in a routine manner, manually or with a computer, is a big problem. Common agreement on international exchange of information through electronic means has to be evolved. We find today that computers will not talk to one another, since formats are different. If we could get an agreement on just a few standards in this area we could exchange information more rapidly and cheaply.

The numbering system for monographs is an area in which I have a particular interest and have worked for ten years to reach some sort of an international agreement. If we do away with long citations and use numbers instead, we could save hours of time and pages of writing and achieve better identification.

The second major effect should be to develop international organizational networks which can effectively implement the above technical standards and plan realistic organizational, operational and financial patterns which will facilitate the exchange of scientific and technical information.

Let me give you an illustration. In the area where everyone has an interest - "everybody talks about it and nobody does much about it" - the weather. Today with international agreement on a few basic technical standards and with an international organizational setup that is financially underwritten by various governments, we have a very effective weather information service which tells us what the weather is. It does not necessarily tell us what the weather is going to be. We have now set up a network on world weather watching that is international in terms of organization, economics and technical standards.

On the other hand, in the field of chemistry, which so many people say is a leader in the handling of its bibliographic and other information, we have no realistic coordination on an international level. The American Chemical Society thinks they have an international program, but if you look around you will discover that many others claim the same. They are not very far apart, but they cannot agree on the economic basis, on a few simple technical standards or on an organizational set-up that will really work.

If the present trend in the international information field continues, I predict that information flow may, in a few years, be beaten by our so-called systems, rather than expedited by them. Today, there are no realistic international agreements on which organization is responsible for which area. The result is competition. If we look at things critically, we see that there are too few effective plans for cooperation and coordination among the scientific information processors and distributors.

Governments, non-profit and private organizations pay little attention to the very effective role that commercial information processors and distributors play. Too little attention is paid to them at international or national meetings where the planning of information networks take place.

Basic to all of information activities is the realistic appraisal of the value and cost of information. Information is an expensive commodity to process, organize and distribute. The governments who underwrite much of the costs do not realize its the real worth and expense. In most countries, information has been pedalled at low cost - it is said, so that the people can use it; but often the low cost has meant inadequate information systems; when we move to the international arena we find this economic problem even greater.

As former President of FID, I know that the people who preceeded me and those who succeeded me know what a terrible time we had to try to underwrite a simple little problem we had. It wasn't simple to us and it was not little to us; but it is in the total aspect of information - we could not find the money and the manpower to get the UDC working efficiently and effectively. UDC is something on which numerous organizations and numerous countries depended and we could not get the few hundred thousand dollars that were necessary. Financial planning has been left out of many information activities, both on a national and an international level whereas it has to be an important part of it.

UNESCO and ICSU through UNISIST should concentrate on improving the organizational, political, social and financial planning in the international information exchange. That does not mean they should not foster and encourage the development of technical standards and systems design of international scientific and technical organizations, both private and public. However the problem is that without the organizational and the financial plan to back them, even the best designed systems in the world will fail.

DR. R.J. GEZELIUS (Sweden): Informal information manages itself very successfully, but there still are means to improve scientific contacts: technical, educational, social, economic and political means. Technical means, like telephone, television, computers, and other equipment, are not especially well adapted today for personal contacts between scientists, or for documentation services in general. Information people were never asked how they would like those means to function from their special point of view. They just had to accept them as they were.

However, documentalists in Sweden now have a good chance to improve on telephone equipment to serve documentation needs. The time has come for information scientists to specify their demands for technical means to be developed. They must try to influence the technical development in fields which are relevant to their special work. How to demolish barriers which prevent contacts between scientists through education, economics, politics, etc., is a big problem not only in the information field, but for science in general.

Scientific information disseminated by libraries and documentation centers is often very vague. Perhaps the information explosion we all are talking about is in fact an information inflation.

DR. G.X. AMEY (Canada): With a reorientation of aims in the exploitation of science and technology, networks that allow creative use of data from various sources coordinated in one location could play an essential role in molding intellectual resources to meet new problems. The traditional system imposes too great a burden of prediction on the indexer. Even with new and seemingly more comprehensive systems of indexing, the indexer can never really look after future needs.

If material you index now is to be correctly indexed for a period of 20 years, you have to know how people will want to use information in 20 years time and this is quite impossible.

Futurologists in the past have been shown to be wrong time and again in predicting what the future will do. If you attempt to index for all possible cases, your indexing becomes an absurdity. I suggest that the complete text needs to be stored so that it can be searched by the searcher and the responsibility for finding the material lies with him.

This, of course, is only necessary in the case of long term considerations or interdisciplinary work. When it comes to short term considerations like SDI, we have found in our system that indexing from thesauri is the most efficient method of indexing and setting up profiles, and so in our own system we allow the possibility of access to our material by several alternative approaches, either using natural language or using controlled terms.

I International and National Information Networks

MR. A. DISCH (Norway): An information program is, of course, not a goal in itself. The national goal in our case was defined as improved quality of life, continued economic growth, and ensuring district requirements. From these national goals a program for their fulfillment evolved. One of the prerequisites for this program was the development of an active, broadly based information service aiming at a lifelong learning society, democratic directive mechanisms and interplay between industry and society to conserve what we call our way of life.

This information service is expected to serve the following four interest areas: the decision processes in industry and government, R&D activities, development and application of new technology, and general information to the public. Both policy and operative decisions will have to be made in order to reach both the information objectives and the national goals that have been formulated.

A four-year information program was developed based upon two concepts: (1) overall R&D economic planning frames for the years 1972-1975; and (2) a ten-year prospective analysis for R&D. These were determined by a central planning committee, based on previous activities in the field, and the capacity and impact the information program will have on the R&D activities.

These planning frames were given as high and low four-year frames. The low frames are an extension of present activities, and the high frames, increased activity as a result of this.

The first phase of the long term planning program was based on literature studies, consultations, simplified technological forecasting techniques and intuitive thinking. Forecasts of the probable development of information services in the next eight to ten years were made and presented as a prospective R&D report.

The question was then asked, if what this report says is reasonable, what should one do on the national level to meet this development? This question was presented to groups of experts representing users, leading information specialists and librarians.

Based on these consultations and on a statement of analysis, a four-year information program was drawn up within the high and low planning frames. The program was divided into three main fields, each with a number of programs consisting of a number of projects. Before granting approval to any project a detailed plan including budget, marginal charge and users interests had to be presented.

This may sound like an extremely rigorous and inflexible system without opportunities for much private initiative and personal engagement which are so important for the development of any activity. This was fully realized by the planning committee and the necessary flexibility has been built into the system. The overall program will be updated every year, and the emphasis can be shifted from one program or project to another. Entirely new projects can be included or the course of the old ones can be changed.

Benefits derived from the long-term program are twofold. We have formulated national goals and overall information plans as an aid to reaching these goals. Due to the involvement of so many people thinking constructively about long term planning of information centers, the ground has been prepared for future activities. An awareness of the necessity for thinking and planning ahead has been created and a more favorable information climate has evolved as a result of this study.

MR. C. KEREN (Israel): Israel's system evolved on a pragmatic, and not a very systematic, basis because as it evolved it had to fight for every facet which came into existence.

The organizational setup of information services are crucial and not enough attention has been given to this area. Most professionals in the field of science information devote a great deal of attention to the methods by which information is handled. Very few touch upon the managerial aspect of the problem. Ours is a small country, geographically remote from the mainstream of developments and, therefore, I quite admit that our view may be biased. We think that information must be physically close to the user. The closer an information center interfaces with the daily requirements of the user, the more useful and essential it becomes.

No single information unit is able to collect and disseminate all the knowledge which is actually or potentially required. Information services must, and should, be recognized as networks. They must establish channels of communication permitting them to have free and privileged access to each other's capabilities.

Networks take various forms and perform on various levels: mission oriented, discipline oriented, organization oriented, subject oriented, national and international. They will probably perform on all levels concurrently. However, a network, as any chain, is as strong as its weakest link.

Much attention has been given to the material from which this chain is forged; indexing methods, abstracting, mechanized systems developments, reprography, etc. Especially in international bodies like FID, UNISIST, CODATA, INIS and others, much has been done to strengthen the bonds between the various links of the chain. But we have left to private and rather uncoordinated initiative the creation of the link itself.

As a result of living and working in a small country geographically and politically remote from the large centers where most of the knowledge is produced, we have identified some factors which we consider crucial for information systems development which might not be as easily identifiable in large R&D oriented environments.

First, users' groups are generally small; therefore, the price to each user of information is high.

Second, dependence on published information is high, as objective restrictions limit oral information transfer. In the US oral information accounts for 80% of relevant information. In Britain, the percentage is much lower, and the figure for Poland is 30 to 35%, the rest coming from published literature. I do not know the exact figure for Israel, but the percentage of published information, either formal or informal, is very high.

Another point is that a relatively large number of users here, or in any country like ours, have interdisciplinary problems because they must cover broader areas and therefore require access to broader information.

In addition, few users can satisfy their needs in the holdings provided by their organizations. For instance, our Chairman, who is connected with the Bell Telephone Labs, most of the time can find what he needs within his own organization. Or somebody working with ICI can find it there. There is nothing like ICI or Bell Telephone Labs anywhere outside the Western industrialized world or the Eastern large industrial combines. In the industrialized countries anybody who does not find what he needs in his own organization can go to another well-organized facility which is not far or with which he can communicate by telephone or telex. These alternatives are not available to somebody in Israel.

Information is expensive because basic stocks are required to satisfy the needs of small groups. When we attempt to provide an SDI service, we have to run the same tape in Israel for fifty users which in the United Kingdom serves 500 or more. The same is true for any other source from which information is derived.

I International and National Information Networks

Trained manpower and suitable machines are scarce and information is almost never the first priority. Financial and operational support without which we cannot work is hard to get because the government has a long list of urgent priorities where such invisibles as libraries and information centers and personnel to operate them rate rather low.

The solutions we attempt in Israel must be viewed with an understanding of these factors. These are the reasons why we view so favorably and support so strongly any international undertaking and investigation which could assist us in eventually getting better access to 99.5% of the knowledge that we must import in order to remain competitive. What might not be vitally important for a network operating in the USA, the USSR or Western Europe is of vital importance to us and I believe to all the 100 or so countries who have not yet developed vast centers of knowledge of their own.

May I take this opportunity to ask our guests from abroad to bear these points in mind when they advise their delegations to UNISIST or to any other international organization. I would also like to ask them to make a greater effort than in the past towards standardization and compatability in pricing and other managerial aspects of our work. The development of individual links such as national networks, is too important a matter to be left to individual inclinations, capabilities and financial organizational facilities. The need exists to formalize some of the methods, including semantics, which are one of the links forming the chain on which we all depend.

MR. LOWRY: Most of the efforts in information transfer as pointed out by Mr. Gezelius are concerned with formal channels. In informal information the sense of hearing, inflection, intonation carries information that you never get from a written page and even more so, the visual aspects of seeing the raised eyebrow, the smile, the frown. Information so transferred is of tremendous value, if you think of information as imparting knowledge.

PROF. V. SESSIONS (US): I was particularly impressed by Dr. Gezelius' trying to divide information into library connected and non-library connected, formal and informal. Obviously libraries can never substitute for some of the means of informal communication. On the other hand, organizations often deliberately keep libraries out of contact with policy makers. For example, in the New York Town Planning Department a staff meeting evolved a new concept on town planning. I was not present but ten minutes after the end of the meeting the participants were down in the library asking, "What do you have on this subject?" Had I been at the meeting where the concept was developed I would at least know what its origins had been. Not having been at the meeting I was at a total loss.

Librarians thus have a double responsibility. They have to keep up not only with developments in the library field, but also with developments in their subject field.

MR. E. SCHAFER (US): There is a tendency by national and international groups to ignore completely the opinions of people who are not on committees and contributions from people who are outside the big cities. If we aim at national and international cooperation, we shall have to work with people within our own communities who are able to work with national and international committees.

DR. M. CREMER (Germany): There is some good experience with bilateral and multilateral systems. One or two partners have a good system and invite other partners to join it. Dr. Adkinson has mentioned cooperation between the US, UK, the Netherlands and Germany in the field of food science and technology. We have also had some successful experiences in the field of management science and I think the task of large governmental and non-governmental organizations is to promote further development of bilateral or multilateral cooperation. There are nationally organized systems of international importance, like INSPEC, which are universally accepted, and other nations are intereste in joining such systems. I would like to ask such national organizations to enlarge the partnership and distribute responsibility, so that other countries will not feel so dependent on their own national systems.

We see a lot of difficulties from a political and economic point of view in starting international organizations like the European Community or FID. UNISIST should play the role of a kind of club to bring people together on a bilateral or multilateral basis, and to make national assistance more internationally organized.

MR. *J.H.* D'OLIER *(France)*: My statement is about informal channels of communication. C'est tres important parce que c'est comme cela que la science progresse. C'est une observation et c'est l'ensemble d'une serie d'observations qui sont fichees par chaque observateur, en particulier en pedologie, en medicine, et dans tous les domaines de science d'observation et c'est du rapprochement de ces observations que naissant finalement les decouvertes, et si l'on saisit cette information a la base, on gagne non seulement du temps, mais aussi on fait progresser la science.

DR. A. COCKX *(Belgium)*: *To Mr. Disch.* What variables do you use to define the quality of life and how does the information network connect with this?

MR. DISCH: *To Dr. Cockx.* What we define as the quality of life has to do with environment, housing, having enough food and in our case also being able to take off on skis and getting away from everybody for a couple of hours. Your national attitude defines the kind of life you want to live. From national goals we try to develop sub-goals and to reach these sub-goals we need information.

MRS. E. EHRLICH-DEVRIES *(Israel)*: *To Mr. Keren.* I think it might be possible at first not to talk about the top level, but to start from the bottom. If the special libraries who very often compile their own lists of information would start to exchange information, and if in each country there were a place to pool the information, for example, in Israel, the National Center for Scientific and Technological Information, which would again disseminate the acquired information, this would be a good start. It should not be so complicated to start at the bottom and exchange information first in the country itself.

MR. KEREN: *To Mrs. Ehrlich.* We have to consider that if we apply a too rigid system on a national scale or within an organization, we would probably wind up with something too formalized to be useful. I think we have to find a way to combine the informal cooperation between the information supplier and the information user, with a rational use of the total means which we have available for that purpose. I personally am not much in favor of creating strict rules which would superimpose a central empire on individual systems.

DR. GEZELIUS: *To Mr. Disch and Mr. Keren.* Have you ever considered making a study in your countries in order to have a basis for the national planning of information and documentation? In Sweden we did such a study very recently. It was published this spring. I do not know what it will be used for, but it could be used for national planning in this field. I would like to ask your opinion on the value of such a study.

MR. DISCH: To Dr. Gezelius. In working out the prospective analysis, we used very simplified programing techniques, using the relevant tree technique. We know that in Sweden they used the Delphi technique. We used some of these facts from Sweden in our own prospective analysis. It is a very useful aid in reports and in making a system of how to plan your information activities. There is the risk that you get so involved in the technique that it becomes a goal in itself, which one should be very careful to avoid.

DR. W.E. BATTEN (UK): To Dr. Amey. You seemed to be regretting if not asking to correct the lack of eternity or the lack of ability to be eternal in some of the information systems that are designed. I am sure this is a very worthwhile objective, but I do hope that we are not to be inhibited because we are unable to make our products any more eternal than those of the airplane or ship designers or even the city planners.

DR. AMEY: To Dr. Batten. Information systems have been in existence for four or five millenia. The attitude of the librarian is basically that of an archivist. This in fact is one of the main objections of documentalists to the traditional library methods where the librarian feels that what he is doing is sacred work, and when he catalogues something it is for eternity. Because of this he may allow backlogs to develop even though the material might be very urgently needed.

Some information will be used in 20 years time. Our own system is about 35 years old and it has about 500,000 documents in it and some of them are still useful.

There is a need for flexibility in indexing. It is impossible to index an item so that somebody in five years time can retrieve it if he wants to use it for a purpose different from that for which it was originally indexed. Material is tied to the indexing done according to the fashion of the moment. When there is a change of national policy, as happened in Canada, previously accumulated material might be relevant to new uses, but if it wasn't indexed for them, it is difficult to retrieve.

MR. LOWRY: Perhaps you have heard something about librarians and documentalists not being the same kind of people. I detect, however, that there is a new wave emerging which makes librarians and documentalists more like each other and there are at least some institutions where there seems to be nothing wrong with having the source of information associated with the exploitation of it and I think we will find more and more of this as time goes on. Close proximity to the source is important and the old dichotomy may one day not far off become less of a dichotomy.

MR. A. LEBOWITZ (Israel): Questions have been raised pertaining to cooperation among Israeli libraries. I serve as Chairman of the Standing Committee of the National and Universities Libraries. Through a series of sub-committees the committee has undertaken various activities intended to make the seven university libraries in the country function as one coordinated network.

Membership in the committee and participation in its activities is purely voluntary, and we haven't had anybody who hasn't cooperated completely with what we are trying to do. We started with several relatively simple projects which have come into being very effectively. The first was the establishment of a telex network linking the libraries and a system of delivery of photocopies and/or books for inter-library loan by means of a special contractual arrangement with the bus company.

As a result, any item requested from any of the seven university libraries which is not actually in use by someone else, is now available to the user within 48-72 hrs.

The number of requests by the telex is increasing. The curve follows very much the curve that the National Library of Medicine (US) published showing the increases in its telex activities.

We have another sub-committee which is working on coordination of acquisitions between the libraries. This is a very touchy subject because in effect it says to the libraries, you must give up some of your sovereignty over how you spend your money. On the other hand, lately there have been so many results of research on the utilization of materials, on new techniques that can be used taking into account the number of users of the materials, the rate of obsolescence or the "half life" of material, that by using this kind of data, we can plan for the logical growth and development of the individual libraries on a cooperative, collective basis.

The union catalogue of books is being worked on, primarily at the National Library, but under the general direction and mostly with the funding of the Standing Committee. The first stage input of the union catalogue is going to be the catalogue of the National Library, supplemented by the catalogues of other university and some special libraries.

The first mechanized edition of the union list will allow the individual holding lists for the member libraries to be produced from one data base, so that instead of duplicating the information and having every library do its own holding statement, this will come as a by-product of the national union list.

The Standing Committee does have funds of its own and it is able to fund and to operate projects such as the union list and the national catalogue. This gives it more leverage than some analogous committees in other countries.

Finally, the committee is trying to do some long-range planning in terms of the automation and mechanization activities of the libraries and trying to see that uniform systems develop.

DR. CREMER: One of the forms of non-formal information is the preprint. In Germany we have very controversial discussions about their usefulness. Some of the users say they are not interested in preprints and prefer to wait for the final version. Others, physicists for instance, say that they are interested in preprints only because the time lag between the invention in the laboratory and the preprint may be as little as one year; whereas the time lag between the preprint and the final publication is another year. That is true for all so-called "new technologies" where the scientific development is very rapid.

Our physicists asked us to prepare a special information store of all the world's preprints to have the fastest possible access to new inventions, and so avoid duplication of work and to make use of the newest inventions for their own work. Are there similar experiences in other institutions or countries?

MR. LOWRY: To Dr. Cremer. I am sure that the problems that you mentioned are indeed problems that you would find in a number of countries. The idea of early claiming of information is well known. You referred to physicists. In the US, Physics Letters comes out rather promptly. There used to be great competition to get things published in certain journals and now there is great competition to get into Physics Letters as early as possible. The more we change, the more we seem to be the same.

SESSION TWO

information systems for specialized applications

Authors of Papers

SOME RESULTS OF AN INQUIRY INTO AGRICULTURAL DOCUMENTATION SERVICES

H. BUNTROCK
European Communities, Center for Information and Documentation (CID),
Luxemburg

SYNOPSIS

In 1970 a worldwide inquiry into agricultural documentation services was carried out. More than 500 services together producing 1.5 million references a year were analysed. Statistics are presented according to various viewpoints as for example publishing, country, language, data handling methods, output data, etc.

Agriculture is one of the oldest human activities - cultivating the land being the main occupation of a people when they became stationary. Compared with this early application, agricultural research and agricultural documentation are rather recent activities. Albrecht THAER, by some historians called the "father of agricultural science", lived and worked some 150 years ago. The oldest agricultural bibliography is said to be the "Elekta georgika sive opuscula quaedam de re rustica", published in 1596 by J. CAMERARIUS in Nürnberg citing 100 prints, 170 manuscripts and 190 Greek and Roman publications. The "Handbuch der ökonomischen Literatur..." by F.B. WEBER, published in 1803 - 1832, quoted more than 6.000 titles. A first serial repertory of agricultural literature was published as long ago as 1860 by A. KROKER (4). Compared with modern application fields, however, as for instance nuclear science, agricultural documentation has already a long tradition.

So far it has been more nationally than internationally oriented, in spite of the fact that there are already some 45 services sponsered by international bodies. The trend towards more international cooperation has recently increased. The main reasons seem to be the decreasing chance, even for the countries which thus far led the world in agricultural research and documentation to complete their collection by their own efforts because of the growing scattering of the literature produced all over the world as well as the role of computers which invite the centres, at least to some extent, to cooperate by enabling them to handle larger amounts of data. Apart from the already existing international centres like the Commonwealth Agricultural Bureaux (CAB), the Instituto Interamericano de Ciencias Agrícolas (IICA) in TURIALBA and others, there are two major projects of international cooperation in agricultural documentation: a worldwide project under the auspices of the FAO, and a regional one of the European Communities.

In the framework of the preliminary studies the working groups of both the above projects have carried out a common inquiry into agricultural documentation and information services (5) in order to get a picture of the actual state as background for the design of the systems to be proposed. The inquiry was based mainly on the "Survey of Abstracting Services ..." by S.v. FRAUENDORFER, published in 1969 (3), and will probably be continued in the form of a permanent inquiry. Due to the limited scope of the first stage and pressure of time, the following data and the conclusions drawn are a first approximation of what really exists in agricultural documentation.

Publishing countries

The output of the documentation services which were examined are published in 50 countries. The biggest producers are the USA with 16,5 % and the U.K. with 16 % of the total number of references, followed by the Soviet Union with 12%, and France and the German Federal Republic with 9,5 % each. The percentage given for the USA is certainly underestimated because we were not able to get figures from certain important services. The output from certain eastern European countries is probably also underestimated because our collection of services of the eastern world is less complete than that of the western world.

Coverage by subject

It would have been extremely instructive, for example, to have been able to secure definitive data on the actual extent to which the primary agricultural literature is covered by the secondary services operating within particular subject areas, and also the extent to which different services cover the same bibliographical items in the primary literature. But to obtain such information would have required technical resources beyond the scope of those which were available for the present survey. Therefore the data presented in table No. 1 should not be taken literally, but as an indication of magnitudes.

Apparently there are some gaps as well as some well covered fields. As to the first there appear to be no or few secondary services specifically covering the following areas within the framework of the FRAUENDORFER scheme which has been used for the classification of the subjects (2):

- Agriculture in cold climates and high-mountain farming
- Geography of agriculture
- Literature about extension and advisory services
- Agricultural administration and legislation
- Farm labour, agricultural cooperation
- Agricultural prices and statistics
- Agricultural meteorology and climatology
- Judging and testing, rearing and housing (animal husbandry)
- Equidae
- Surveying and levelling
- Environmental factors

On the other hand there are some subjects fields which appear to be strong points and therefore probably well covered, for example:

- Soil science, plant nutrition, and fertilization
- Phytopathology and plant protection

TABLE 1 SUBJECT : SERVICES AND REFERENCES / YEAR

Remark: It must be emphasized that the values given in the table are only estimated averages and have no more than indicative value. The figures given for subject category "Z" (agriculture, wide-ranging, not specified) represent the output from services which in many cases, cover a wide subject range which it has not been possible to subdivide further. The figures given below should therefore be increased by approximately 30 % to take account of the additional output from these sources.

SUBJECT	TITLE SERVICES	TITLES PER ANNUM (ESTIMATION)	ABSTRACT SERVICES	ABSTRACTS PER ANNUM (ESTIMATION)	TOTAL SERVICES	TOTAL REFERENCES PER ANNUM	PERCENTAGE OF TOTAL MINUS "Z"
Z Agriculture (wide ranging)	64	422.920	54	245.500	118	668.420	-
A1 Documentation			1	300	1	300	0,04
A6 Agric. research			3	1.295	3	1.295	0,16
A7 Basic sciences	10	30.475	19	27.150	29	57.625	7,00
A8 Tropical agric.	5	11.775	6	7.895	11	19.670	2,38
B History + Geography	1	200	2	185	3	385	0,05
B1 History of agric.	2	5.300	2	875	4	6.175	0,75
B2 Geography of agric.	1		3	1.300	4	1.300	0,16
C Agric. education			4	2.400	4	2.400	0,29
C2 Extension. Advisory work			1	150	1	150	0,02
D Agric. administration + legislation			2	1.370	2	1.370	0,17
D1 Internat. Organisation	2	1.125			2	1.125	0,14
D4 Agric. legislation			3	1.800	3	1.800	0,22
E Sociol.+econ. aspects	5	13.300 •	5	7.500	10	20.800	2,52
E1 Rural sociology	2	800	3	1.625	5	2.425	0,29
E2 Agric. economics	7	2.600	8	6.110	15	8.710	1,05
E3 Farm organisation			5	2.650	5	2.650	0,32
E4 Farm labour			1	350	1	350	0,04
E5 Agric. cooperation	2	650			2	650	0,08
E6 Marketing	1	1.000	1	1.000	2	2.000	0,24
F Plant production			10	24.950	10	24.950	3,02
F1 Agric. meteorology + climatology			3	1.950	3	1.950	0,24
F2 Soil science	1	800	18	13.575	19	14.375	1,72
F3 Soil cultivation			2	700	2	700	0,08
F4 Plant nutrition	2	800	12	8.875	14	9.675	1,17
F5 Plant cultivation	1		3	15.250	4	15.250	1,85
F6 Plant breeding			6	15.350	6	15.350	1,86
F7 Phytopathology	5	31.930	11	15.580	16	47.510	5,75
F8 Plant protection	6	30.725	15	20.250	21	50.975	6,17
G Special field crops			8	6.550	8	6.550	0,79
G1 Cereals	1		5	8.550	6	8.550	1,03
G2 Alimentary legumes	1		3	530	4	530	0,06
G3 Forage crops			11	7.125	11	7.125	0,86
G4 Root crops			1	250	1	250	0,03
G5 Sugar plants	1	200	5	3.425	6	3.625	0,44
G6 Textile plants			6	920	6	920	0,11

SUBJECT	TITLE SERVICES	TITLES PER ANNUM (ESTIMATION)	ABSTRACT SERVICES	ABSTRACTS PER ANNUM (ESTIMATION)	TOTAL SERVICES	TOTAL REFERENCES PER ANNUM	PERCENTAGE OF TOTAL MINUS "U"
G7 Oil crops			2	270	2	270	0,03
G8 Other industr.crops	1	825	5	4.800	6	5.625	0,68
H Horticulture	1		8	13.550	9	13.550	1,64
H1 Fruit	2	300	4	480	6	780	0,09
H2 Vegetable culture			6	2.000	6	2.000	0,24
H3 Field vegetables			3	750	3	750	0,09
H4 Viticulture	1	300	4	1.250	5	1.550	0,19
H5 Ornamental gardening			1	850	1	850	0,10
H6 Landscape gardening			1	300	1	300	0,04
K Forestry	5	10.700	12	25.955	17	36.655	4,43
K1 Natural environment	2	40	1		3	40	
K2 Silviculture			3	3.200	3	3.200	0,39
K3 Forest protection	1				1		
K4 Forest engineering	1				1		
K5 Forest products	2	1.000	6	13.750	8	14.750	1,78
K6 By-products			1	2.250	1	2.250	0,27
K7 Forest management	1		2	300	3	300	0,04
K8 Forest economics	1		1		2		
M Animal husbandry	2	1.700	13	28.675	15	30.375	3,68
M1 Animal breeding	1	8.000	[illegible]	2.500	2	10.500	1,27
M4 Feeding			3	850	3	850	0,03
M5 Utilization			1	600	1	600	0,07
M6 Hygiene			4	8.750	4	8.750	1,06
M7 Veterinary medicine	5	23.850	24	39.850	29	63.700	7,73
N Particular domestic animals			4	8.575	4	8.575	1,04
N2 Larger ruminants			4	5.000	4	5.000	0,61
N3 Smaller ruminants			1	250	1	250	0,03
N4 Swine	1	2.500	3	2.350	4	4.850	0,59
N5 Other mammals			2	200	2	200	0,02
N6 Poultry			4	7.100	4	7.100	0,86
N7 Fish culture	1	400	3	700	4	1.100	0,13
N8 Insects			6	2.020	6	2.020	0,24
N9 Other animals			1	2.500	1	2.500	0,33
P Rural build. Land development	2	500	2	2.800	4	3.300	0,40
P1 Farm buildings	1	1.200	2	7.100	3	8.300	0,99
P2 Land development	3	40	19	12.720	22	12.760	1,54
Q Agricultural machinery	5	1.000	15	27.400	20	28.400	3,44
Q2 Power machinery			1	600	1	600	0,07
R Technology of agric. products	2	100	14	30.250	16	30.350	3,69
R1 Milling. Baking	1	1.700	3	4.800	4	6.500	0,79
R2 Starch and sugar industry			4	3.275	4	3.275	0,40
R3 Fermentation industr.	2	4.300	7	3.135	9	7.435	0,90
R4 Technology of fruit + vegetables	1	300	2	1.280	3	1.580	0,19
R5 Oils and fats			1	600	1	600	0,07
R6 Meat and fish industry	2	3.500			2	3.500	0,43
R7 Dairy industry	2	3.500	5	3.700	7	7.200	0,87
R8 Textile industry			1	100	1	100	0,01

114

SUBJECT	TITLE SERVICES	TITLES PER ANNUM (ESTIMATION)	ABSTRACT SERVICES	ABSTRACTS PER ANNUM (ESTIMATION)	TOTAL SERVICES	TOTAL REFERENCES PER ANNUM	PERCENTAGE OF TOTAL MINUS "Z"
R9 Other industries	2	550	3	28.100	5	23.650	3,48
S0 Preservation of nature	2	·825	8	3.905	10	4.730	0,57
S1 Hunting	2	1.900	6	2.700	8	4.600	0,56
S2 Fresh water fishing			7	13.370	7	.13.370	1,60
S3 Sea fisheries	1	5.000	4	9.500	5	14.500	1,76
T Human nutrition	2	2.300	10	19.950	12	22.250	2,70
T2 Vitamin research			1	400	1	400	0,05
T4 Dietary diseases			2	190	2	190	0,02
T5 Food policy and aconomy			1	1.200	1	1.20[illegible]	0,15
U Food stuffs in general	2	100	10	22.940	12	23.04[illegible]	2,80
U2 Processing			3	500	3	500	0,06
U3 Preserving			5	3.630	5	3.630	0,44
U5 Storage and transport	1		1	1.000	2	1.000	0,12
U8 Cookery. Domestic economy			3	970	3	970	0,12
W Environmental factors	4	3.865	11	9.955	15	13.820	1,68

•) 6 x 1.800 of the same titles in different longuages

- Forestry
- Veterinary medicine
- Technology of agricultural products.

Number of references a year

Table No.2 shows the cumulative totals of the annuel increase in references and as percentages of the total increase. Starting with the biggest producer, the services are in decreasing order of the number of references. Thus the cumulative total and percentage can be defined for every number of the respective largest services. The 10 largest documentation services, for example, publish 28 %, and the 20 largest services 40 % of the total number of references.

The total sum of 1.5 million titles and abstracts must contain, of course, plenty of overlaps. The "Bibliography of Agriculture" covers (with about 110.000 titles a year) according to NAL's own studies 50-60 % of the whole agricultural literature (1), for which we may conclude a total number of 200.000 titles per annum. Compared with 1.5 million references stated in the survey, this would mean a duplication rate of 7-7,5, in other words, on the average every original document is quoted more than seven times in the secondary literature. Taking only the abstracting journals, which produce 855.000 abstracts a year, the application rate would be nearly 4, thus converging, for example, to the abstracting situation in nuclear energy documentation.

In spite of the high duplication rate it must be emphysized that none of the existing bibliographies and abstracting journals do at least cover approximatively the whole field of agriculture. For most of the subjects various reference tools must be consulted to be sure of getting all literature wanted.

Data handling methods

Mechanisation in the sense of using only machine punched cards or punched tapes is not very widespread in agricultural documentation. Mechanised centers process about 11 % of the total number of references. The percentage of electronical processed items is 20 %. The USA have the highest rate (51 %) followed by the U.K. with 20 %. The rapid development of computerisation, however, will probably render obsolate every statistical feature.

Output data

150 (29 %) of the documentation services examined are bibliographies. They comprise 640.000 titles a year (43 %). 380 services (74 %) are abstracting journals, which produce 855.000 abstracts a year (57 %). The relations mean that the average number of titles per bibliography is much higher than the average number of abstracts per abstract journal. There is a number of large wide-ranging bibliographies, but most of the abstracting services are specialised.

Subject arrangement

348 documentation services use their own subject arrangement, the Universal Decimal Classification (UDC) is used in 192 services, 96 services group the items by means of subject headings, 24 services arrange them alphabetically

TABLE 2 REFERENCES / YEAR

SERVICES	CUMULATED TOTAL	REFERENCES PER ANNUM	CUMULATED TOTAL	PERCENTAGE OF THE TOTAL
1	1	78.500	78.500	5.25
1	2	66.000	144.500	9.67
1	3	55.000	199.500	13.34
1	4	36.000	235.500	15,75
1	5	30.000	265.500	17,75
1	6	30.000	295.500	19,77
1	7	30.000	325.500	21.78
1	8	30.000	355.500	23.77
1	9	30.000	385.500	25.78
1	10	30.000	415.500	27,78
1	11	28.000	443.500	29,66
1	12	22.250	466,000	31,16
1	13	21.500	487.500	32,60
1	14	20.000	507.500	33.94
1	15	16.300	523.800	35,04
1	16	15.000	538.800	36,04
1	17	15.000	553.800	37,03
1	18	13.000	566.800	37,90
1	19	12.300	579.100	38.73
1	20	12.000	591.100	39,54
3	23	36.000	627.100	41,94
3	26	32.800	659.900	44,14
3	29	30.000	689.900	46,14
3	32	29.300	719.200	48,10
3	35	24.800	744.000	49,74
3	38	24.000	768.000	51,37
3	41	24.000	792.000	52,97
3	44	23.230	815.230	54,53
3	47	22.000	837.230	56,--
3	50	21.000	858.230	57,40
6	56	38.300	896.530	59,96
6	62	35.500	932.030	62,33
6	68	30.500	962.530	64,37
6	74	29.800	992.330	66,37
6	80	27.300	1.019.630	68,20
6	86	25.400	1.045.030	69,90
6	92	24.000	1.069.030	71,50
6	98	22.100	1.091.130	72,97
6	104	21.000	1.112.130	74,38
6	110	20.000	1.132.130	75,73
10	120	30.100	1.162.230	77,73
10	130	26.750	1.188.980	79,52
10	140	24.800	1.213.780	81,18
10	150	22.350	1.236.130	82,67
10	160	20.200	1.256.330	84,02
10	170	18.400	1.274.730	85,25
10	180	18.000	1.292.730	86,46
10	190	16.650	1.309.380	87,58
10	200	15.100	1.324.480	88,59
10	210	14.500	1.333.980	89,56
10	220	12.900	1.351.880	90,41
10	230	12.000	1.363.880	91,23
20	250	22.500	1.386.380	92,73
20	270	19.760	1.406.140	94,04
20	290	16.450	1.422.590	95,14
30	320	21.400	1.443.990	96,57
30	350	16.200	1.460.190	97,66
40	390	16.430	1.476.620	98,76
45	435	11.780	1.488.400	99,54
50	485	6.745	1.495.145	100,--

by author name or scanned journal. Sometimes more than one arrangement system is used by a service.

Language of abstracts

22 different languages have been used by the analysed bibliographies and abstract services. English has, with 36 %, the highest percentage of the total annual production. German follows with 12 %, Russian and French with 11 % each. The percentages for English and Russian are probably higher because of the above mentioned reasons.

Indexes

Only 46 % of the secondary services provide subject indexes representing 68 % of the annual production. This means that subject indexes are more a feature of large than of small services.

111 indexes are conventionally alphabetical ones, 13 are KWIC and 8 are KWOC indexes. For another 103 indexes we have no further indication as to which of the types they belong.

The indexing depth (average number of key words per item) varies between 1 and 20. The analysis of 59 indexes results in an average depth of 0,5 keywords per 100 references a year; only, the amplitude of variations is very large - lying between 0,007 and 4,0.

The subject indexes of 31 documentation services with an annual production of 148.000 references are based upon controlled vocabularies. Most of these services are special abstracting services, only a few are bibliographies.

21 agricultural documentation services provide geographical indexes; 248 provide author indexes or 48 % of the services for 68 % of the references.

Information services

23 bibliographies and 43 abstract services with an annual output of 225.000 references offer an information service. In fact, the numbers are probably higher because only a few of the secondary publications advertise explicitly additional services like retrieval, SDI, etc., and only those could be taken into account.

REFERENCES

(1) BOURNE, C.P. Characteristics of Coverage by the Bibliography of Agriculture of the Literature relating to Agricultural Research and Development. Report No. IGC - PA - 69 - 38. Information General Corporation (IGC), Palo Alto, California 94 303, USA; and other reports ot the same series.

(2) FRAUENDORFER, S.v. Système de classification des sciences agricoles. . Classification Scheme of Agricultural Science. Stoffeinteilung der Landwirtschaftswissenschaft. Third revised and enlarged edition. BLV Verlagsgesellschaft München Bonn Wien (1960)

(3) FRAUENDORFER, S.v. Survey of Abstracting Services and Current Bibliographical Tools in Agriculture, Forestry, Fisheries, Nutrition, Veterinary Medicine and Related Subjects.
BLV Verlagsgesellschaft München Basel Wien (1969)

(4) FRANZ, G. Die Entwicklung des landwirtschaftlichen Bibliotheks- und Dokumentationswesens. Mitteilungen der Gesellschaft für Bibliothekswesen und Dokumentation des Landbaues, Stuttgart, Heft 13 (1970) S. 6 - 28

(5) BUNTROCK, H. (Ed.) Survey of Agricultural Documentation Services (Working Draft), 1. Dec. 1970. Food and Agricultural Organisation of the United Nations (FAO), Rome, and European Communities, Center for Information and Documentation (CID), Luxembourg.(including statistical analysis and comments, main list with individual data for each analysed service, various indexes for country, language, subject etc., and KWIC index) (1971)

A SPECIAL-LIBRARY INFORMATION-CENTER MODEL FOR A SOCIETAL-PROBLEM FIELD

Mark Keller
Center of Alcohol Studies, Rutgers University, New Brunswick, N. J., U.S.A.

SYNOPSIS

The new information centers, and specialized libraries conformingly, are adopting technologies for coping with the growing flood of reportage which provide quanta of informational bits. This policy overlooks the fact that information is not the same thing as education. Described is a model of an information center for a complex societal-problems field (alcohol problems) which aims to preserve the educational function of the classical library.

Within two hundred years--a short time, as history is measured--mankind has experienced three grand revolutions: the industrial revolution, epitomized by the enshrining of belief in machine rather than muscle power; the scientific revolution, epitomized by the idealization of belief in biological evolution rather than divine intervention; and the sociopolitical revolution, epitomized by the substitution of belief in economic force rather than human idealism. Although this aphoristic epitomization of complex events over-simplifies their history, I cannot here expand on them; yet it is necessary at least to mention them lest it should seem that the ideas about the human and technological problems of information, which make my topic, are insufficiently grounded in awareness of their bases. The three revolutions need remembering because they are the background for the idealization of technology which I shall be discussing here.

Equally briefly I may only mention two of the three grand explosions which have shaken mankind in the past generation--the atomic explosion and the population explosion--before dealing with the third, the information explosion. In fact, I mention the first two only because the means which have been proposed to deal with them are the same as those proposed for the third--namely, suppression. They say, let us not have so many atomic explosions; and they say, let us not have so many births. And they have also said--as by a Presidential Scientific Advisory Committee in the U.S.A.--let us not have so much information; authors--referring particularly to scientists, of course--should not write so many articles, especially unnecessary ones (1). This recommendation is offered, paradoxically, where

---

Aided by a grant (MH-05655) from the U.S. National Institute of Mental Health.

no one has yet dared to specify what is unnecessary for whom. But it ignores the future realities: We shall have more scientists and scholars; and even if not--a possibility which my thesis deplores--then all the more certainly we shall have more professionals and technologists and academics. These will also inevitably produce not fewer but more writings, owing to the fact that they are, and in profound respects can be expected to remain, human. Since the academicians, technologists and professionals and, hopefully, scientists and scholars, will continue to proliferate articles as well as books and proceedings reports, I feel safe in predicting that they will get them published--even if they have to found new societies and journals every week, as they are already doing, and eventually every day.

Recognition of the problems of coping with the information explosion came to the fore after World War II. Whether or not it may be possible--and here for the moment I slight the not unimportant issue whether it is desirable--to reduce the average literary productivity of the information-giving world, the volume is already enormous, is sure to increase, and surely requires a special technology for coping with it. We have been developing that technology of information processing during these last 25 years or so, with all the features and promise of up-to-date sophistication, including the predictable plunge into the over-utilization of computers. Whether there is such a phenomenon in the world as an information science is sometimes still argued. But only a few years ago the American Documentation Institute changed its name to the American Society for Information Science; Bowker (7) has just published a $25 "Introduction to Information Science"; and here we are holding an International Conference on Information Science; so how can there not be an information science? Perhaps we have an informatology (6). Lest I appear in a false light personally, I confess, for what it's worth, that my membership in the American Society for Information Science is the only basis of any claim I would make to be thought an information scientist. But I don't mean to deny that I am somehow somewhat of an informationist.

Libraries--please forgive the unavoidable utterance of this banality--have been one of the most useful sources of knowledge, of information and education, ever since they were invented. It is hard to imagine an educated person who has not had access to one or more libraries, or a cultured person who does not reflect a sort of library in his own personality. It was somewhat surprising, when I attended my first Research Conference on Problems of Information in Science many years ago, to discover that the frontiersmen in developing a science of information processing were nearly unanimous, and mostly rather emphatic, in believing that the place where the specialists in the new field might be trained, could not be the library schools. They asserted that library schools were capable of training only traditional librarians; the new information specialists must be educated elsewhere. The people who expressed these opinions were originally physicists, chemists, engineers, mathematicians, computerists. My modestly put forth suggestion that the library schools might reform their curricula was greeted with raised-eyebrow kindness. Apparently I was naive.

American library schools have nevertheless reformed their curricula. They now give courses in indexing, abstracting, S.I.D., computerism, and other functions conformable to the new demands on specialist types of librarians who must be fit to man the new information centers. Some have even changed their names conformably --some former Graduate Schools of Library Science have become Graduate Schools of

Library and Information Science. Not all the graduates of these reformed schools and curricula, however, are working in the industrial special libraries or the few specialized academic information centers. Some of them are filling jobs in the classical or traditional specialized libraries. For, conformable to these changes in the preliminary training of their personnel, the large and important class of specialized libraries have been reforming themselves. Apparently they have heard the rich voices of the new information scientists pronouncing the old libraries obsolete, and so they are converting to the new movement, reforming themselves functionally, to become information centers. A thoughtful and prophetic analyst of modern library problems (5) has indeed urged them to this but with the wise caution to maintain their most important function, the integration of knowledge.

Potentially the reformist trend of the specialized libraries is a beneficent one. Undoubtedly the irrepressible flood of new information must be managed in new, more efficient, and automated ways, so that needed information on what has been learned or done, what is known on specific problems, should be accessible efficiently and economically. And it is logical that the specialized libraries should adopt the newest and most effective technologies and be equipped to serve clients needing specialized information. My concern is with the circumstance that what I call--I hope not unjustly--the engineering focus of the new informationism threatens the fundamental values of these institutions as libraries. The trouble lies in the fear that has been aroused by the tidal-wavelike aspect of the volume of new information and the resulting rush to try to cope with it by mechanistic gadgetry. Marvelous technologies are developed to in-put the information into a system, and by equally marvelous technology it is retrieved again with subsystemic and supersystemic packaging of the informational bits. And what is the expectation from it all? The dream-ideal is that a person who needs to know will ask a machine a question, and out will come the answer. Or the system will deliver what he needs to know, based on his profile, before he even asks. The precise processes of button-pushing and instant out-printing are unimportant. What is relevant is that the institution to which the person goes with his question, with which he communicates, is organized as an information center. He formulates his question and it formulates "the answer," it responds with a package of information.

The trouble with this ideal system is that information is not education, it is not even the same thing as knowledge.

Now I was myself early in pointing out that the classical, the traditional, the old-fashioned library could no longer serve the informational needs of specialists. I was myself one of the first to try to develop an information system for a special problem field--and it was a tough one, as complex and multidisciplinary as you can find--the field of alcohol problems. This system is among the pioneers of the genre. So it is not likely that I would be saying that the traditional library, general or specialized, is superior to the new information center. What I am concerned about is that we in the information field, puffed up by the power of our newfangled technology, should not mislead the world into thinking the traditional library is obsolete, useless, only of museum interest for the man of the future. The truth is that our clever new systems and technologies, if and when we get them really working effectively, will be able to deliver only information.

Information is not a commodity I denigrate. It has been my preoccupation a long time. But I would be dismayed by the prospect of information taking the place of education. I am not so old-fashioned as to be alarmed by the prospect of teaching machines. They can efficiently relieve teachers of much routine drudgery and give students enhanced opportunities for needed individual instruction. But I would be dismayed if the notion should prevail that these machines can replace the teachers, or displace books, as the media of education. Machines can facilitate the mastering of essential quanta of precise information on discrete topics. They cannot provide that something which every scientist, every scholar, needs continuously in his unending role as continuing student--education.

I will not attempt what I might not be able to do at all, and certainly could not within the limitations of the present effort, to define what I mean by education. I think we all know to a sufficiently near degree what I mean. It is that process of learning which continues to enrich the mind and expand the perspectives of the potentially creative personality. It is gained by interaction with the minds. with the thinking, of other learned, knowledgeable, wise, people. The interaction may occasionally be direct communication by conversing with them; mostly it has to be indirect, by reading what they have written, or seeing the works they have produced. If this conception of education, of the absorption of knowledge and the integration of understanding, as distinguished from acquisition of information, is correct, then obviously we cannot replace the library, or the museum, by however efficient an information center.

Old-fashionedly, I dare to say that to be educated one must live in a library. For the younger student this is practically a physical truth. Far-advanced students don't so much need the opportunity for random browsing and serendipitous discovery--they have learned how to select the best chances. But they too must somehow receive more than processed quanta of information, they too must continue to be educated by refreshing themselves repeatedly in the primary lively pools of knowledge, if they are to remain creative and avoid becoming mere technicians in narrow specialties.

How is the opportunity for education--of students and scientists and scholars --to be preserved in face of the information flood, the imperative need to deal with it in automated quanta. and the inevitable conversion of libraries, certainly the specialized libraries, into information centers?

The answer I propose is a model of the information center of the future which conserves the specialized library. Its staff must understand that the primary function is indeed librarianship--an essential element in education, though often misunderstood and neglected even by university faculties. The secondary function-- even though it may absorb a majority of staff, space and budget, is information processing and retrieval. That this form is possible, and that it can function successfully, is the burden of the remainder of this presentation, in which I shall describe an actual model of a specialized library so conceptualized, the Library-Documentation-Information-Publication System of the Center of Alcohol Studies at Rutgers University.

Essential features of this model have been described elsewhere (2, 4) so I can deal with them briefly here. They are displayed as a list of instruments in the accompanying Chart 1.

Chart 1

Abbreviations--

CAAAL: Classified Abstract Archive of the Alcohol Literature
IBSA: International Bibliography of Studies on Alcohol
QJSA: Quarterly Journal of Studies on Alcohol
RCAS: Rutgers Center of Alcohol Studies

Main Instruments of the System

1. RCAS Library: A special library with collections of periodicals, books, pamphlets and other materials related to alcohol problems and allied subjects; catalogued by standard and special headings.

2. Master Catalog of the Alcohol Literature: in preparation; housed in RCAS library.

3. CAAAL: a cumulative Abstract Archive with an index on edge-notched cards for automatic retrieval by topics. Allows also chronological and limited author retrieval. Housed in RCAS Library (special search sets in other offices) and in 55 Depositories in 15 countries. Accession aid: CAAAL Manual.

4. CAAAL Matching-Originals Collection: Full-text copies of all documents incorporated into the Archive, housed by CAAAL serial number in RCAS Library.

5. International Bibliography of Studies on Alcohol: a serial publication.
   Vol. I, References, 1901-1950 [published].
   Vol. II, Subject-Author Indexes for Vol. I [published].
   Vol. III, First Decennial Supplement, 1951-1960 [in preparation, 1971].
   Vol. IV, Second Decennial Supplement, 1961-1970 [1972-73].
   Vol. V, [tentative,] The Literature Before the 20th Century [1974-75].

6. CAAAL BiblioFile: Collection of topical bibliographies produced for requesters, keyed to CAAAL and QJSA abstracts; housed in CAAAL headquarters, RCAS.

7. Current Bibliography: New titles and titles not meriting abstract, published quarterly in QJSA, Part B (Documentation).

8. Current-Literature Abstracts: published quarterly in QJSA, Part B.

9. Current Periodical (original articles, critical reviews, notes): The Quarterly Journal of Studies on Alcohol, Part A.

10. Current-Literature Subject Index: Open-ended; exhaustive. Published in QJSA annually. (In future, quarterly with annual, quinquennial and decennial cumulations.)

Item 1 of the "instruments" or "fixtures" is the Library. The system was developed in and functions in a problem-oriented center in a University--the kind of interdisciplinary problem-oriented center that a recent editorialist in Science (9) has just told us ought to be created in universities. We created this one over 30 years ago. The problem field happens to be alcohol, and I may emphasize that alcoholism is only one of the problems. It happens that alcohol interacts with multifarious human interests and activities, sometimes as a "problem" only in the sense that everything is a problem to the student and the intellectually curious. So we get our "information" from a fascinating mix of disciplines, professions and other sources: chemistry, pharmacology, biochemistry, physiology, genetics, experimental, social and clinical psychology, neurophysiology, anthropology, sociology, economics, political science, psychiatry, linguistics, clinical medicine, biometrics, public health, law, enzymology, penology, theology, oenology--yet this list is by far incomplete. Our library therefore is naturally classified as a specialized library. But it really is that. It houses special collections of periodicals, books, pamphlets, films, and other materials relevant to the study of alcohol problems. It acquires the materials needed by a staff of information processors--bibliographers, abstracters, indexers, archivists, programmers-computerists, searchers-retrievers, and consultants. It houses the output of this staff and makes it available, with expert human help, to visiting searchers. It serves the system's staff and the institution's staff as a library, a source of education as well as information. It provides to visiting students and scholars, from as near as our own University--in which this is one of the integrated special libraries--and from as far as halfway across the world, not only the informational bits which it can send out in relatively inexpensive packages in response to questions, but also and especially the opportunity to sit in a library environment--surrounded by catalogs and reference works and books and periodicals and pamphlet boxes--and to educate themselves.

From this library heart of the system branch out interconnected offices housing the specialist documentalists or information scientists who process "the literature," convert it into the classified informational bits, and, on request from anywhere, provide very specific answers to most specific questions, mostly in the form of bibliographies and photocopies of abstracts. Except that, sometimes, we tell an inquirer frankly: We think what you need is to come and spend some time in our library.

Other instruments of the system designed to document the complex alcohol-problems field may be housed in the library proper or in some of its connecting offices, and elsewhere as Chart 1 shows. Item 2, the Master Catalog of the Alcohol literature, is of course maintained in the Library. It now consists of some 200,000 cards but in an incomplete state, and divided up into segments required for special projects under way. When complete, it is intended to be a prime single source of reference to all the literature, giving not only bibliographic detail but main subjects and whether processed into other parts of the system (CAAAL, IBSA, QJSA, with accession information)--and where a copy exists if not in our own Library.

CAAAL (Chart item 3) is a cumulative system of informative abstracts, printed in readable type size on 19.1 X 16.8 cm edge-notched cards. The subjects are indexed on the notched margins according to a Code Dictionary in the CAAAL Manual (3) and are retrievable by hand sorting with a needle (or machine sorting with multiple needles). The CAAAL Manual also provides an alphabetic Index Key to Sorting, a guide to topical search codes. Besides Library and office sets for use at the Center, there are complete sets of CAAAL in 30 Depositories in the U.S.A. and 25 Depositories in 14 other countries. There are also 20 Subscribers to annual CAAAL card issues in the U.S.A. and 14 abroad which have incomplete but useful sets. The CAAAL subject index is formed of 24 Main Categories (e.g., Metabolism of alcohol, Clinical physiology, Nutrition, Mental disorder, Experimental psychology, Alcohol addiction, Socioeconomic statistics, Education), each of which may have up to 92 subtopics (e.g., in Sociology, 1, women; 6, drivers, motor traffic; 11, ethnic groups; 19, social class; 30, sex behavior; 38, abstinence; 40, kinds of beverages; 57, prevention; 66, polls, questionnaires). By sorting the CAAAL cards at combinations of numerals within one of the Main Categories (sometimes in more than one) the very specific bibliography and abstracts on nearly infinite varieties of topics are quickly assembled. For example: effect of hormones (or insulin) on oxidation of orally (or intravenously) administered ethanol in fasted (or fed) habituated men (or other animals, or alcoholics, or abstainers); or, effect of parental alcoholism on intelligence (or personality, or morbidity) of offspring as measured in twin studies. The CAAAL system has exceptional "browsability"; an individual searcher can sort, resort, combine and recombine groups of cards to order a complex topic in varieties of aspects. In January 1971 CAAAL incorporated 12,045 abstracts covering 13,309 documents on 14,355 cards; it is growing presently at a rate of over 600 cards per year. Eventually to purge CAAAL of excess, the system contemplates generating serial CAAAL-oriented critical topical reviews.

The collection of full-text originals is sufficiently noted as item 4 in Chart 1, and is available to back up needs going beyond the informative CAAAL abstracts. In January 1971 this collection was 98% complete for serial numbers 5001-13000+, and 38% complete for earlier documents. The missing documents are being systematically acquired (10).

The contents of the International Bibliography of Studies on Alcohol (IBSA, see Chart 1, item 5) are selected from the Master Catalog; all items incorporated into CAAAL or listed in the Quarterly Journal of Studies on Alcohol (QJSA) are included. IBSA Volumes I and II include over 25,000 titles, from separata and over 2500 periodicals in 31 languages, keyed for abstracts in CAAAL and QJSA. Volume III, for the period 1951-1960, will have over 15,000 titles. The cumulation of new bibliography is approaching 2000 items per year.

The CAAAL BiblioFile (Chart 1, item 6) consists of the master set of several hundred topical bibliographies compiled over the years in response to inquiries. They are re-used (with updating) only if they fit a new inquirer's needs exactly--otherwise a new bibliography is compiled. In preparing such bibliographies, the QJSA and IBSA indexes may be used to supplement CAAAL.

The Quarterly Journal of Studies on Alcohol, published at the Rutgers Center of Alcohol Studies (RCAS), "houses" the items numbered 7-10 on Chart 1. A firm conviction based on experience is the decisive advantage of a documentation or information system being closely allied to the publication of an important segment of the primary literature. The RCAS Publications Division publishes not only the QJSA, IBSA, and a Monograph Series, but also various other types of materials.

No strict division exists between the documentation-information staff and the publications staff; planning, problems and productivity are shared in convenient collaborations. One result is that original articles undergo professional abstracting; bibliographies are professionally checked; and subjects are deeply and expertly double-indexed (for QJSA and for CAAAL).

All new bibliography is gathered by an acquisitions librarian and prepared for listing in the next issue of QJSA under a series of standard main subject divisions. Documents are next acquired and evaluated before further treatment. Duplications (obvious or masked) are identified and relegated to cross-reference only. New periodical literature is abstracted informatively for inclusion in QJSA and CAAAL, or by carefully elaborated criteria, abstracted only indicatively for publication in QJSA but exclusion from CAAAL. Books are assigned to specialists for review. Abstracts and reviews, as well as articles published in QJSA, are deeply subject indexed. A master vocabulary serves as a prime guide. The index has been published annually in the past, with occasional cumulations. Shortly, with the help of automated techniques, indexes will be published in each issue, with regular annual, quinquennial and decennial cumulations. Abstracts incorporated into CAAAL are indexed also according to its Code Dictionary, as described above.

The features of the system are also described in terms of 10 functions: acquisition, evaluation, abstracting, indexing, storage, retrieval, review, research, communication, and service. These do not need elaboration but a few points are worth mentioning. Under review we include the projected serial CAAAL-oriented topical reviews. Under research we include beginning efforts to test the effectiveness of the system more objectively than by the gratifying feedback from well-served users. Under communication we include the output of special informational materials, including the IBSA series, and certain special publications such as an Alcoholism Treatment Digest. And under service we include much of the activities and output of the Library, the librarians and the whole staff of the documentation-information system which we think of as, hopefully, a model of the specialized library of the future. The system can serve those who seek information by providing bibliographies, on discrete topics, which lead to both old and new knowledge from a mix of disciplines and professions, and informative abstracts of the most useful literature as determined by evaluation. Such services to individuals are given often after interchanges of communication that lead to refined definition of the requester's problem. The system can provide copies of full-text originals of a substantial part of the literature (8). It also gives general service to the field by serial and special publication of primary articles, monographic works, bibliography, abstracts and reviews. And not least important, it provides the classical multidisciplinary library resources needed beyond information, essential for education. This policy is based on the belief that the long-range well-being of mankind depends not on technical experts but on the creative efforts of educated people, and that the specialized library of the future has an important role in the forming of educated people.

The system is moving cautiously into automation. We do not believe machines can produce abstracts or indexes of adequate quality. This is the work of educated specialists. Computers can take over the mechanical processes of storage, man-guided retrieval, and print-out. Even so, until computers are universally available to everybody, the advantages of retaining, e.g., the old-fashioned CAAAL system, which can be made available in many libraries and institutions at small cost, seem obvious.

My final emphases are these: The Documentation-Information-Publication system of the Rutgers Center of Alcohol Studies is neither a substitute for the traditional library nor a mere information center. It is a reconceptualized specialized library combining the capacities to serve both the immediate-information needs of specialists and the broader educational needs of scholars. Its successful operation depends only secondarily on machinery, and primarily on the human resources of a trained staff of educated librarians-informationists, with a supporting cadre of constantly available consultants in the multidisciplinary faculty of a problem-oriented center located in a university, and the additional support of the personnel and institutional resources of a university with a broad range of graduate schools and institutes.

In spite of a history of over 30 years, this system shares the permanent fate of science and scholarship and education: it is still only in the stage of development and presumably will never be perfect or finished.

References

1. The President's Science Advisory Committee. Science, government, and information. Washington, D. C.; U.S. Govt Ptg Office; 1963.

2. Keller, M. Documentation of the alcohol literature; a scheme for an interdisciplinary field of study. Quart. J. Stud. Alc. 25: 725-741, 1964.

3. Keller, M., Efron, V. and Jellinek, E. M. CAAAL manual; a guide to the use of the Classified Abstract Archive of the Alcohol Literature. New Brunswick, N. J.; Publications Division, Rutgers Center of Alcohol Studies; 1965.

4. Keller, M. Documentation of an interdisciplinary field of study: alcohol problems. Proc. 28th Int. Cong. Alc. 2: 145-163, 1969.

5. Adams, S. The library function in science. (Address given at the dedication of the Library of Science and Medicine, Rutgers University.) New Brunswick, N. J.; 20 Nov. 1970.

6. Otten, K. and Debons, A. Towards a metascience of information: informatology. J. Amer. Soc. Inform. Sci. 21: 89-94, 1970.

7. Saracevic, T. Introduction to information science. New York; Bowker; 1970.

8. Documentation and information services of the Rutgers University Center of Alcohol Studies. New Brunswick, N. J.; Rutgers Center of Alcohol Studies; 1971.

9. Long, F. A. Interdisciplinary problem-oriented research in the university. Science 171: 961, 1971.

10. Bacon, S. D. The McCarthy memorial collection in 1971. Quart. J. Stud. Alc. 32, No. 2, June 1971.

AN INFORMATION SCIENCE APPROACH TO NATIONAL LIBRARY SERVICE FOR THE BLIND, WITH SPECIAL REFERENCE TO SOUTH AFRICAN EXPERIENCE.

D.E. Schauder
Director, South African Library for the Blind

SYNOPSIS

A system for eventual integration with normal facilities is being developed. The program envisages the introduction of reading machines with non-visual output, non-visual access to information stores, and computer-selected catalogue printout in ink or braille. This work should facilitate the introduction of mechanical cataloguing and participation in cooperative lending and on-line networks.

Information Science is concerned largely with the problem of the "information explosion" - of sifting an over-abundance of literature. Yet many of the techniques which it has developed can help solve the problems encountered by those to whom the great mass of printed records is inaccessible. Blind information users can benefit perhaps more immediately than any other group by creative departures from current modes of written communication and by the selective retrieval and presentation of information. Similarly the recreational blind reader can best be served, and the limited literature available in reading media for the blind best be exploited, by the use of advanced communication techniques to keep librarians, readers and potential readers constantly aware of the services being offered.

Problems in providing effective library services to the blind in the past have been due mainly to three factors:

1. the low population density of blind readers,
2. their wide range of needs from light leisure reading to specialised documentation, and
3. the nature of their reading media which seemed to exclude blind library and information users from normal sources.

Postal distribution has been the traditional answer to the first problem. In Britain, a small country with a relatively high population density, it has been the dominant mode for many years. In the United States and the U.S.S.R., large countries with high concentrations of blind people in many areas, there is more decentralization.

II Information Systems for Specialized Applications

The problem of a wide diversity of reading needs to be served by extremely limited resources is one which South Africa shares with other countries. The situation is made more complex because, with South Africa's multiplicity of languages, the amount of literature in a given field available in the reader's home language is often drastically restricted. The obvious answer here is that gradually blind readers should be enabled to make increasing use of normal library facilities, with some means of converting the printed page into audio or tactile form. In South Africa, as in Britain, the special nature of the reading media, both embossed and recorded, developed for the blind has tended to strengthen the idea that library service to visually handicapped readers is very specialised and outside the scope of normal library work. The relatively small number of trained librarians, in contrast to welfare workers, who have been active in this field up to now, has tended to limit the total amount of first-hand professional experience of library services for the blind and to confirm this traditional misconception.

A basically new approach to national library service to the blind is, however, possible: the approach of the Information Science oriented librarian. This approach is an important element in a new policy which has been formulated and implemented at the South African Library for the Blind (S.A.L.B.).

The policy rests on three main foundations:

1. the National Union Catalogue of Literature for the Visually Handicapped,
2. co-operation with libraries and information services for the sighted, and
3. the comprehensive acquisition of foreign catalogues and bibliographies of material for the blind, and the building up of loan co-operation with foreign libraries for the blind.

In other words, a service pattern which can allow eventual integration of a large area of library service to the visually handicapped with normal library and information facilities is being developed parallel to the traditional method of postal service. This programme anticipates the eventual introduction of reading machines capable of converting optical input to audio or tactile stimuli, and of possible non-visual means of access to information banks.

Clearly essential to all these departments, and the key to full exploitation of the small quantity of material available is high-grade cataloguing. The S.A.L.B. has thus given priority to the production of a book-catalogue with a high degree of subject access. Since the S.A.L.B. does not hold all the literature available to the blind in South Africa, this must be a union catalogue.

The first two parts of the National Union Catalogue covering the Mark IV and Mark I cassette holdings at the S.A.L.B. have already been published. Like future parts of the Union Catalogue, they are self-contained classified catalogues arranged along the lines of the British National Bibliography. Added features are full bilingualism and a certain degree of subject access to works of fiction. Every entry has an annotation indicating scope and treatment. Each part of the Union Catalogue will be kept up to date by accessions lists and cumulations as required. In due course the parts will be consolidated into single printed catalogue.

The Mark IV catalogue was produced manually. The Mark I catalogue was produced on a semi-mechanised basis using the Friden Flexowriter, but with a great deal of hand sorting and filing. It is planned that all future parts and cumulations will be compiled by computer at the South African Council for Scientific and Industrial Research (C.S.I.R.), with the Flexowriter at the S.A.L.B. being used for the initial punching of the coded paper tape, and the final print-out for lithography.

This co-operation with the C.S.I.R. marks an important new phase in the history of library services to the blind in South Africa, and probably represents a pioneering step in bibliographic work in this field in the world.

The system being used by the C.S.I.R. is based on that developed for the South African National Bibliography (S.A.N.B.) by the Zentralstelle für maschinelle Dokumentation (Z.m.D.) in Frankfurt, with alterations made to accomodate the special features of the Union Catalogue and the computer equipment available in South Africa. The system provides for the following:

a. the creation of all index entries and cross-references,
b. the sorting of all entries and the cumulation of these entries as required,
c. the insertion of guide headings in the classified file, and
d. the capacity for selective print-out according to criteria of medium, language, and location.

During input preparation publications are catalogued and conventional catalogue cards are drawn up. (The form of entry based as far as possible on the descriptive principles laid down in the Anglo-American Cataloguing Code.) Codes are then added to these cards in preparation for the paper tape punching operation. Information contained on the cards is punched on to paper tape.

The paper tape records which are the computer input are checked for errors which can be programmatically determined and are used to create magnetic tape records. The hard copy produced as the paper tape records are punched is checked visually for any further errors. A new paper tape is then punched containing all corrections to be made to the original tape records. These checking and correction procedures are repeated until an error-free file of magnetic tape records has been created.

The required computer processing is then done by some ten programmes which:

1. create magnetic tape records from the paper tape input, and check and correct these records,
2. create from the magnetic tape records all index entries, create sort keys and sort the index entries,
3. create the entries required in the classified file, including cross-reference entries, create sort keys and sort the entries,
4. build up a list of guide headings linked to the Dewey Decimal Classification for introduction into the classified file as needed, create running guide headings for successive pages,
5. cumulate or extract entries with all relevant index entries, cross-references and guide headings as needed, and
6. provide page numbering and correct spacing for use of continuous stationery during Flexowriter print-out.

The Frinden Flexowriter used is the 2301 model. The computers are the I.B.M. 360 model 65 and the C.D.C. 1890. The I.B.M. 360 computer has no paper tape output unit.

The required paper tape output records are written by the I.B.M. 360 computer on to magnetic tape. The magnetic tape records are then read into the C.D.C. computer which punches them out on to paper tape.

It must be stressed that the mechanization of the National Union Catalogue of Literature for the Visually Handicapped could not have passed beyond the more limited possibilities of print-out from Flexowriter tapes after hand sorting and filing, were it not for the fact that the S.A.N.B. programmes, planned and drawn up with great effort by Dr. H.J. Aschenborn and his colleagues at the State Library, Pretoria, Z.m.D. and the staff of the Information and Research Services Division of the C.S.I.R. under the leadership of Mr. D.G. Kingwill, were generously made available to the S.A.L.B.

The production of entries in machine readable form not only allows swift book-catalogue print-out in ink-print, but also holds the possibility of producing a computer braille edition with no intermediate manual stage. This idea, however, requires further investigation and will probably entail the use of overseas computer brailling facilities.

Great emphasis has been placed on adhering as far as possible to internationally standard cataloguing procedures. The way is thus open, on the one hand for library service to the blind eventually to benefit from projects such as MARC, and on the other for participation in any national or international on-line network which may develop.

After the complete Union Catalogue has been compiled, South Africa's national bibliography of literature for the visually handicapped will be extracted. This will facilitate reciprocal interchange with other countries. The impact of the policy on the service potential of the S.A.L.B. can already be felt. A large number of public, university, and special libraries in South Africa which include visually handicapped people among their present or potential readership have placed standing orders for the parts of the catalogue as they appear. Use of the Flexowriter facilitates correspondence with libraries on a national scale.

In regard to public libraries, co-operation is aimed primarily at the group of readers for whom large-print books, already distributed by public libraries, are no longer suitable. Such readers, who often at first do not wish to be directly identified with agencies serving the blind, may continue to be served by their public library which receives tape players and cassette books on inter-library loan from the S.A.L.B. If the reader has no psychological barriers, he may be offered the choice of postal service, or service by his public library. Inter-library loan service through the local library's "shut-in" facilities, is also invaluable for old people incapable of carrying on their own correspondence and with no-one to help them select from a catalogue, and for temporarily blind hospital patients.

In regard to university and special libraries, inter-library co-operation with the S.A.L.B. on the basis of the Union Catalogue gives blind users the means for "one-stop" access to their own library's reference facilities, to available literature for the blind in all media, and to the combined bibliographic services of the co-operating libraries thus providing maximum access to the material that they require both in ink-print and in media for the visually handicapped.

The adoption by national library services for the blind of policies and techniques similar to those outlined above can only result, in the short term, in more effective use of available resources. In the long term, it will ensure for the visually handicapped library or information user a service pattern capable of incorporating without delay the products of any successful new reading-machine technology and of deriving maximum benefits from advances in the fields of library and information science. Documentary material which is stored in non-visual form can often yield hard copy in tactile or audio media with little extra expenditure or technological effort. It is earnestly hoped that workers in the fields of librarianship and information science will bear in mind, when designing new services in the future, the exciting possibilities that these techniques hold for visually handicapped users.

References

1. Schauder, D.E., The South African Library for the Blind: its future as a national library. South African Libraries. 37 (2), 1969. p. 81-9.

2. National seminar on the library needs and study problems of visually handicapped students. South African Libraries. 37(2), 1969. p. 95-7.

3. South African National Bibliography Computerized System. Unpublished document made available by the Information and Research Services Division, South African Council for Scientific and Industrial Research to the South African Library for the Blind. [1971].

4. Schauder, D.E. and N.M. Lodder, Mechanisation and library co-operation - a new era for the S.A.L.B. To be published in South African Libraries, 1971.

AN INFORMATION SYSTEM FOR AUTOMATED DATA PROCESSING FILES

Peretz Wollman
Israel A. Cohen
National Center of Scientific and Technological Information, Tel Aviv

SYNOPSIS

Large organizations can utilize an Information System to control decentralized data processing operations and to facilitate data exchange among their divisions. The heart of the system is a completely documented catalog of automated files. Implementation of the system requires a standardized vocabulary to describe the technical attributes of files and their contents. The system itself serves as a data-base for the continuous operation of an information center serving the needs of large organizations.

Large organizations with semi-independent divisions (for example, government administrations), often use numerous data-files in their data processing operations. Each division typically deals with the data used in its own operations and such data is usually collected, processed and disseminated independently. Consequently, data items are often duplicated several times, both between divisions and within the same division. Furthermore, as Mr. Aitchison said this morning, "computers will not talk to one another". That is, differences in hardware and software impede the flow of data between organizations.

Two methods have generally been used to minimize the effort required for handling data in large organizations. One method is to create a comprehensive, centralized data-base which can simultaneously serve all areas of an organization. This approach has several disadvantages. The continuous acquisition of new data and the extensive preparation required to create the initial data-base make the preparation stage very long and costly. This disadvantage is compounded by the need to define all final output requirements at the initial design stage. Technological, personnel, and economic resources required to operate a centralized comprehensive data-base are difficult to obtain and maintain. Also, security considerations become more crucial. Collecting all of an organization's data in one place increases the loss if security safeguards are breached. The security hazard associated with large collections of data often exceeds the sum of the individual hazards associated with parts of such data.

A second method generally used to minimize data processing effort is to establish and rigidly enforce uniform standards relating to files and their contents. Disadvantages of this approach include the need to continuously update such standards and the difficulty of enforcement when some areas of an organization refuse to properly apply the uniform standards.

The information system approach which we offer provides an alternative to these customary methods. An information system avoids most of the problems just mentioned while enabling large organizations to exercise control over decentralized data processing operations with a minimum of additional effort. In addition, such a system provides a standardized mechanism for transferring and exchanging data between and among divisions of an organization. The heart of such an information system is a completely documented catalog of automated data processing files.

We, at COSTI, have recently completed a pilot project designed to produce an inventory of Israeli government population data files. The results of this project enabled us to make decisions about what data items to collect about files and file contents in order to produce a workable catalog.

Data items about file contents included the following: name of file, names of fields within the file, sources of input, destination of output, descriptions of the population involved in the file (such as drivers, students, homeowners), system used to select the population (total group or various types of samples), who controls the file, who has access to the file, etc. Technical data items about the files included the following: input and output media, storage media, character sets used, format data, type of file organization (sequential, indexed-sequential, addressable), record type (fixed or variable length, blocked or unblocked), a status code to indicate whether the file was in the planning, testing or operational stage, a code to indicate whether the file is updated, not updated, historical, statistical or designed for one-time usage, and related technical data.

The information collected for this catalog will be listed by office responsible for file maintenance, by major subject matter of files (such as health, education, employment, or taxation), and by field name. Of course, the field name listing will indicate in which files the fields are located.

We found quite a few special problems associated with making our catalog and implementing the system. Utilization of the information system concept requires a standardized vocabulary to describe both the technical aspects of files and the contents of the files. The variety of terms used - often within the same division of an organization - to describe the same phenomena, attribute, or entity is truly amazing. Care must be taken so that the inputs to the information system are made with a uniform, consistent terminology. Whenever possible the system should incorporate existing uniform standards and new standards must be devised to fully describe automated files when necessary.

It is important to properly characterize the various user-levels of the system. The system will be of value to programmers, systems analysts, managers and administrators, but the types of information and services these people want from the system will be quite different.

At some installations and in some divisions of an organization, the current level of file documentation will not be sufficient to provide the required data about their files for input to the information system. Consequently, such installations will need to produce sufficient documentation to implement the system.

We have discovered that different divisions in an organization will treat the same data items with different levels of confidentiality or security. Sometimes this difference in treatment is justified. At other times it is not. The formulation of a "Security Index" for data items will enable the organization to determine whether such differences should be maintained or eliminated. In general, using a Security Index will enhance the likelihood that data items requiring a given level of security will receive that level in all divisions of an organization.

Finally, a word about what we are doing at COSTI right now. We are building such an Information System, and we are producing SWAP/File programs as a tool for information retrieval from documented files. SWAP/File is our acronym for "System for Working with Automated Processing Files". These programs will not manipulate the data contained in such files but will provide a mechanism for transferring and exchanging data among organizations and their subdivisions. The SWAP/File programs perform a retrieval function and a creation function. The retrieval function extracts content items from existing files by exchanging the definition of the content item for the item itself. The creation function produces new data files utilizing content items previously extracted from existing files.

The Information System itself will serve as a data-base for the continuous operation of an Information Center serving the needs of large organizations. The first application of this Information System at COSTI is to the population data files utilized by the ministries and departments of the Israeli government. We expect to apply the system to the following additional areas in the future: property and taxation files, corporation and business enterprise files, and inventory and equipment files. Eventually the System may be utilized to coordinate the usage of meteorological, agricultural, and water resource data files.

THE TISA PLAN

Margrett B. Zenich
Chief, Scientific and Technical Information
Office Chief of Engineers
1000 Independence Avenue, S.W.
Washington, D.C.

SYNOPSIS

The objective of the TISA Plan is the provision of more effective technical information support activities to Army elements through experimentation with information techniques and concepts, the development of information handling procedures, the discovery of basic principles of information acquisition, organization, retrieval and transfer, and the preparation of guidance documents for information support activities.

The TISA plan is a long range, multi-faceted project to increase the effectiveness of Technical Information Support Activities with the United States Department of the Army and its parent agency, the Department of Defense. It is research based, but includes in its design provision for evaluating and implementing solutions to problems and operational recommendations stemming from the research. The TISA project is administered by the Corps of Engineers for the Army. Though it is administered by the Corps of Engineers, TISA is responsible for research to improve the flow and utilization of information within the entire Department of Defense.

As it is now constituted, TISA can be characterized as having three major thrusts, as follows:

First, the provision, through research, of a comprehensive theoretical foundation for the flow and utilization of information in the Department of Defense, especially the Department of the Army;

Second, the resolution, through research, of operational problems in Technical libraries, technical information analysis centers and information centers, especially the resolution of operational problems encountered in implementing theoretical research findings;

Third, the dissemination of findings emanating from the theoretical and operational TISA research and from other sources, as well as research about dissemination formsts appropriate to a variety of utilization needs.

The ultimate objective of TISA is to provide dynamic and effective technical information support activities to all units of the Department of Defense through experimentation with information concepts and techniques, and through accelerated dissemination in various media.

Examples of Types of TISA Projects

To illustrate the kinds of work related to each major TISA thrust, I shall present three examples of current projects. The first is a project in the Identification of Technical Information Problems. The first phase of this project was designed to identify technical information problems relating to Army and Department of Defense technical libraries, information centers and information analysis centers. All recorded research completed by other agencies but related to information flow and utilization was screened. That which is potentially valuable for military libraries was coded for retrieval and stored in a computerized data bank. Analysis of this data bank of research indicated the characteristic types of problems, the intensity of research effort in relation to these problems, and the gaps in the existing research. This phase of the project has been completed.

The second phase, now beginning, will derive from the analysis of relevant research a list of the highest priority needs. Each problem on this list will become the topic of an in-depth synthesis of research and state-of-the-art review. Decisions about the priority and nature of future projects will be based on these reviews.

The second example is the Model Technical Library Project, which stems from the TISA emphasis on implementation and evaluation of research findings. A model technical library has been established to serve as a test and evaluation base for developing new techniques for use in Army technical libraries and for testing the solutions to problems as they emerge from other projects. It is designed to provide a means of evaluating the feasibility of implementation of results from theoretical research in a real environment, rather than in a theoretical or a simulated environment.

Thus far, two phases of the project and two special studies in the model library context have been completed. In Phase I, a model technical information facility to serve a working sub-unit of the Corps of Engineers was designed and is now operational, and a staff was recruited. In Phase II, basic policies and procedures were established, orientation and on-site training of the staff was completed, the technical collections were updated, and preparations were made for the anticipated series of special studies and experiments.

The first special study undertaken was a study of user needs in the agency which sponsors the model library. This study supplied data for policy and procedural decisions and recommended appropriate follow-up studies. The methodology of this study may well serve as a prototype of user studies within the Army and the Department of Defense technical libraries.

The second special study contained within the model library project was a comparative study of commercial microfilm systems for vendor products information and specifications.

This study profiled available commercial systems in terms of the distinct user requirements of the model library's parent agency, and it provided recommendations for decisions relating to appropriate systems for the model library.

In Phase III of this model library project selected techniques and equipment incorporated in the model library will be tested and a systems analysis of the library's operations will be carried out complete with manpower and cost/effectiveness studies. Special studies yet to be completed include the development of an automated periodical data file. Pending experiments include an SDI experiment, a comparison of microfilm vs hard copy provision of documents in response to selected needs, and a feasibility study of the utilization of external information services as adjuncts to the model library's in-house resources.

The third example of a project, which exemplifies the TISA emphasis on rapid dissemination of research findings, is called the Dissemination of Library Technology Information Project. Begun within the last three months, this project is for the purpose of establishing an information service for Army and Department of Defense information personnel about advanced technological developments which can affect programs and operations of technical libraries, information centers and information analysis centers.

The planned information service will provide resumes and critiques of new developments in mechanized information handling, with emphasis on the potential application of such developments to military technical libraries.

The service must present update summeries of research for use at several levels of expertise, in a variety of presentational modes, and at frequent intervals of time. For example, one appropriate topic about which there is a considerable amount of recent research is color microfiche. Ideally, the new project will prepare a resume of the research in textual form, in audio form and in visual form, or in a combination of these media appropriate for use by different levels of interested personnel, such as other researchers, library administrators, procurement officers or library clerical personnel.

The formats finally selected for this information service should be equally valuable to several secondary interested groups as well as to the primary target group of military libraries and information facilities. Manufacturers of equipment should find it useful because in critiquing new equipment modifications needed for special purposes will be discussed. Technical libraries generally, both in and out of the Federal library community, should benefit from its critical approach, its succinct presentations, and its currency in relation to other, similar sources of comparative data.

Structure of TISA

In order to accomplish its many and varied specific objectives, TISA is structured in five task areas, as follows:

1. Problem Identification and Technical Direction
2. Organization of Technical Information
3. Information Operations
4. Networking
5. Testing and Evaluation

Each of these task areas is subdivided into appropriate work units, twenty-four in all at the present time. General staff management is provided by the Chief, Information Systems Office, Department of the Army, The Chief, Scientific and Technical Information Program, Corps of Engineers, provides the technical competence for monitoring. Expert advice on the design and implementation of each task area and work unit project is provided by carefully selected advisory boards and consultants. These individuals are chosen for their specialized knowledge and their relationships to libraries, to the TISA sponsors, and to the Federal government. Criticism from these advisory boards is fed back to each work unit operational staff during the course of the research. Thus, on-going research and future planning benefit from the immediate critical appraisal of a cross-section of the technical community, both the Federal and the private sectors.

Summary of Results

It is difficult to summarize an effort that includes as many diverse elements as the TISA plan, especially when many of the individual projects are still in their early stages. Nevertheless, there are tangible results, and some of them can be stated briefly. Thirty-seven technical reports have been published. These include reports of original research, literature surveys, state-of-the-art reviews, bibliographies, handbooks, equipment evaluations, directories and other types of documents. These technical reports have been widely distributed to Federal and non-Federal libraries, many are still available through the National Technical Information Service, Bureau of Standards, U.S. Department of Commerce. The model library is one of TISA's most important tangible results.

It will continue to serve as the laboratory for a wide variety of experiments and tests. The data bank of completed research projects from the total library information community is, perhaps, the TISA result with the most significant implications for future projects. This data base will serve directory, bibliographic and primary informational purposes for the anticipated wide range of continuing TISA projects.

The TISA plan is the coordinating umbrella under which the U.S. Department of the Army is seeking solutions to problems of technical information utilization. Providing a theoretical base, testing proposed solutions and disseminating results will continue to absorb TISA's attention in the foreseeable future.

AN INTEGRATED INFORMATION SYSTEM FOR PHYSICS AND ASTRONOMY

Arthur Herschman
American Institute of Physics
New York, New York, USA

SYNOPSIS

The National Information System for Physics and Astronomy, being implemented by the American Institute of Physics with support from the National Science Foundation, is initiating procedures and services which will integrate primary, secondary, and tertiary information channels in the disciplines of physics and astronomy. In addition to applying current technology to the large and established primary publication program of the AIP and establishing a correlated computerized secondary information file, NISPA is developing new facilities and procedures for the dissemination of primary and secondary information, as well as encouraging the writing of reviews.

Introduction

While you may be somewhat familiar with the American Institute of Physics, let me give you a few brief facts concerning it which may also aid in our discussion of the National Information System for Physics and Astronomy (NISPA).

The AIP was established forty years ago as a federation of the leading societies in physics; it presently has seven member societies (shown in Figure 1) representing some 50,000 physicists and astronomers all over the world; in addition it has 19 affiliated societies with an interest in physics, 110 Corporate Associates, and a Society of Physics Students with 6000 student members in over 400 universities. The Institute provides a number of services on behalf of physics and astronomy, including programs in education, physics history, public information, etc. However, its principal activity is its primary publication program, one of the most concentrated in any discipline. Some 100,000 scientific text pages were published by the AIP in its own name or on the behalf of its societies during 1970 (the growth is shown in Figure 2), in 18 primary journals, 5 society bulletins, and 13 Russian translation journals. The AIP processes over 30% of the world's total of primary physics information, as publisher, translator, or marketer.

AIP MEMBER SOCIETIES - 1970 MEMBERSHIP

The American Physical Society	28,532
American Association of Physics Teachers	13,275
Optical Society of America	6,437
Accustical Society of America	4,663
Society of Rheology	830
American Astronomical Society	2,701
American Crystallographic Association	1,853
	58,291

FIGURE 1

Thus, we see the Institute as the central agency for America's leading societies in physics and astronomy, performing numerous services which can best be done by the societies jointly, publishing and processing enormous quantities of scientific literature, and in constant dialogue with the scientific community. It should be viewed as a logical extension of its mandate for "the advancement and diffusion of the knowledge of physics" and as a response to the needs of the physics community, that the Institute has assumed the responsibility for the design and implementation of a National Information System for Physics and Astronomy, with the support and encouragement of the National Science Foundation. The specific responsibility for this Program, within the AIP, lies in the Information Division.

The Public Archive Of Science

The AIP's Information Program assimilated and extended earlier studies into all aspects of communications in the physics community: the channels used and preferred, publication patterns and problems, new services needed, etc. Work proceeded on the planning of a system and the design of specific elements (Information Division, 1969 and 1970).

During the course of our studies and planning, one point assumed overriding significance, viz., the central importance of the publication process in the very fabric of science and its communication, and within the sphere of publication, the pre-eminent role of the scientific journal. As this has been discussed extensively elsewhere (Herschman, 1970), let me only remind you of the salient points.

1. Publication represents a prime motivation for scientists to do research, affording them "property rights" to their ideas through peer recognition of their priority (Merton, 1969).

2. The formal publication process permits the scientific community to control the quality of what is entered into the Public Archive (Pasternack, 1966).

3. The existence of the Archive permits the community of scientific scholars to arrive at a consensus as to what is "public knowledge" (Ziman, 1968).

4. Individual and library subscriptions to copies of the Archive (e.g., journals) are still the preferred means for scientists to obtain information (Information Division, 1970, Ch. 11.2).

AIP SOCIETY MEMBERSHIP AND PAGES PUBLISHED

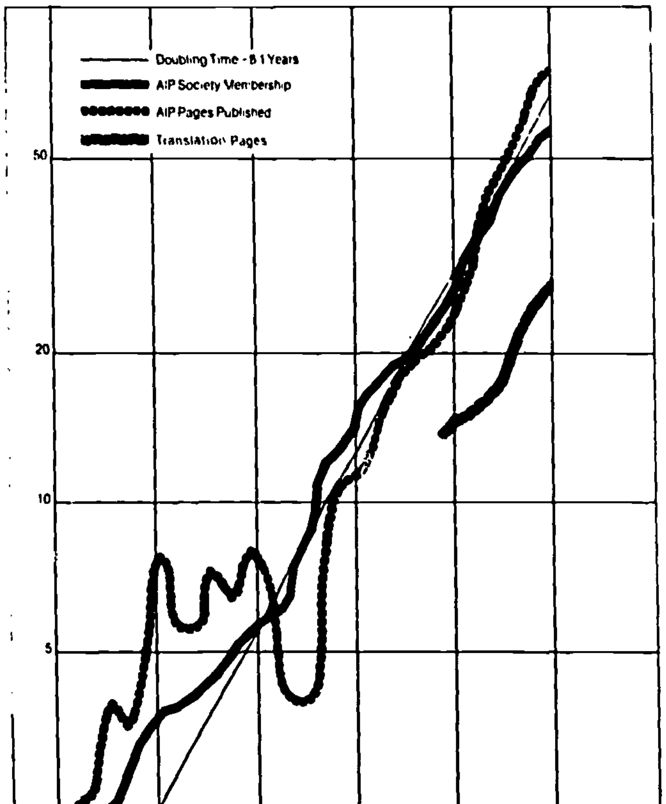


FIGURE 2

The gist of these points is that scientific information systems must be predicated on the wide dissemination of quality publications to appropriately interested scientists, in a timely manner. It is not sufficient to merely supply information about these publications, i.e., secondary information, although secondary information has an important role to play in aiding in the selection of the primary information appropriate to a given user.

Our studies have shown that there are two additional characteristics of the Archive which are worth noting. One is the limited number of journals which are involved in such an Archive for physics; this is shown by the data in Figure 3. Only about 120 journals account for 90% of the journal coverage of Physics Abstracts, compared with about 2000 for Chemical Abstracts. Of these, in physics, only 10 journals account for more than 50% of all of the references in articles, i.e., considering the citing patterns of authors as an indication of use, a very small number of journals account for the greatest amount of useful information. The second characteristic is also based on citations or references in the articles, and is indicated in Figure 4, which shows how important the recent literature is in physics, relative to the older material. The probability that a paper is cited is halved every seven years.

To put the role of the Archive in better perspective, let us look at how information flows in three channels: that of informal communication, the Archive itself, and the secondary information channel, as shown in Figure 5 (adapted from Garvey and Compton, 1967). This diagram indicates the interrelationship between the elements of the information flow until a sufficient concensus is reached about this information so that it can be taught as a part of the public knowledge of the subject. It does not indicate, in any systematic manner, the flow of information to the user before and during his initiation of research.

Horizontal vs. Vertical Integration

The traditional information system is based on secondary information, usually represented by an abstract journal, but more recently also including a computer-readable data base of such information. Secondary services have always strived for as complete a coverage as possible within their respective subject areas. As a natural consequence of very broad coverage and of the availability of computer-readable files, secondary systems are moving more and more into a mode which I shall call "horizontal integration", i.e., integration of data bases across subject-matter lines. The more traditional approach to this is evidenced by Chemical Abstracts, whose coverage transcends chemistry into all surrounding disciplines to pick up any item which might be of use to chemists in the broadest sense. Another approach is that of the INSPEC system which publishes the most comprehensive abstract journals in the English language in physics (Physics Abstracts), electrotechnology (Electrical Engineering Abstracts), and computer technology (Computer and Control Abstracts). Having gone to a computerized data base, the system is able to produce a single cross-disciplinary magnetic tape from which it is possible to produce each of the three original services. However, the most interesting examples of horizontal integration can be found in the operations of the information reprocessors who accept input from several disciplinary systems and offer information services for some mission which traverses disciplinary lines. Such a system is that of the American Petroleum Institute which serves the missions of its clients in the petroleum industry. Such horizontal integration is

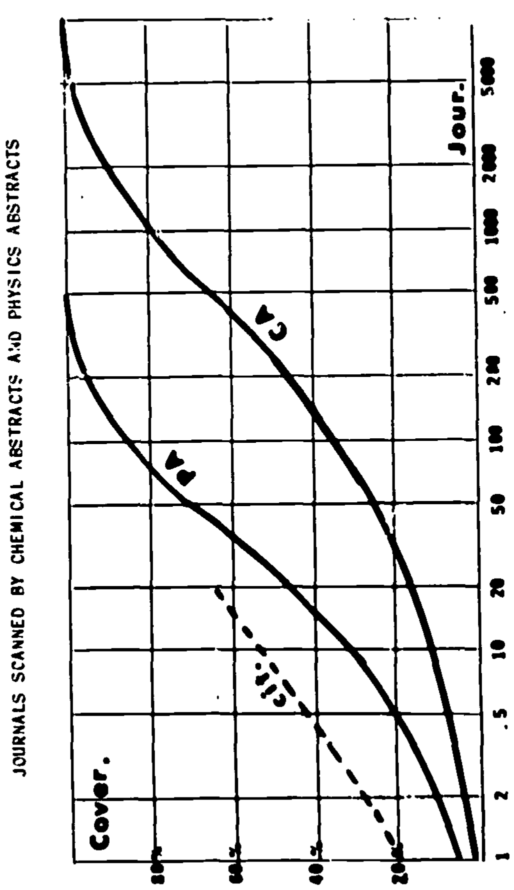


FIGURE 3

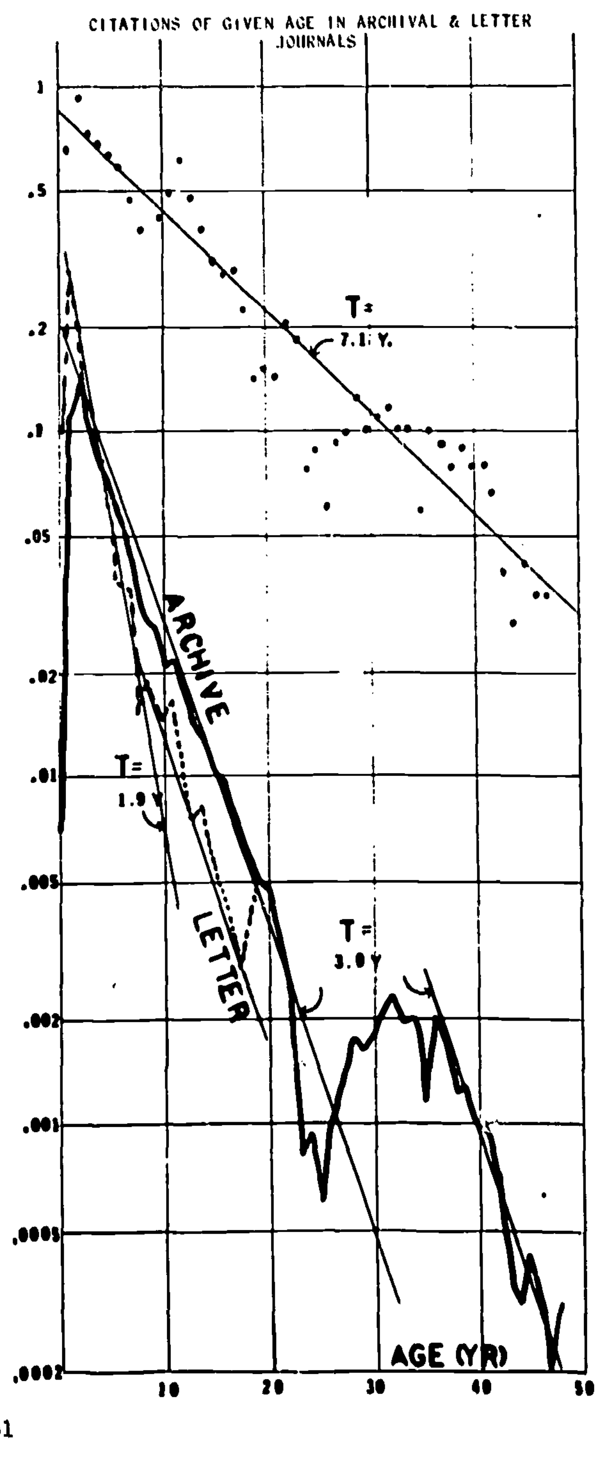


FIGURE 4

THE EVOLUTION OF SCIENTIFIC INFORMATION

TIME (yr)	INFORMAL	ARCHIVAL	SECONDARY
0		(STARTS RESEARCH)	
1			
	COLLOQUIA		
2	SEMINARS		
	CONVENTIONS		BULLETINS
3			
	PREPRINTS	(SUBMISSION TO JOURNAL)	
4	REPORTS		
	SYMPOSIA	JOURNAL PUB.	CURRENT AWARENESS JOURNAL
5	REPRINTS	PROCEEDINGS	ABSTRACT JOURNAL
			DEPOSITORY FILES
6			
		COMPILATIONS	
7		REVIEWS	
		PATENTS	
8		TEXTBOOKS	

FIGURE 5

possible because the secondary information is sufficiently compact so as to be processable by computer. The user of the service receives information about the original information, which sometimes suffices and which at other times overloads his library resources.

A complementary approach to horizontal integration is what I will call "vertical integration", which is the integration of the stages of information processing indicated in Figure 3. Such an approach has been taken by the American Institute of Physics in designing the National Information System for Physics and Astronomy. This integration is both of production and distribution. The secondary information file is going to be obtained directly as a by-product of the keyboarding for the initial journal production, and indexes to the journals are obtained as a by-product of the secondary file. Primary information is also made available in a form designed to supply ready hard copy backup for secondary services, e.g., as microfilm of the full text announced in a limited-coverage magnetic tape service. The tape service will contain the microfilm reel and frame number for each article on the film in addition to the bibliographic citation. Thus secondary search services and printed services based on the tapes would supply unambiguous leads to the full text from an easily copyable format. This possibility of supplying meaningful services from a limited coverage of the journals is due to the compact nature of the literature of physics and the fact that such a large fraction of the set is handled by one organization, AIP.

Further integration in the processing and dissemination of information is made possible at AIP through its strong connection with the physics and astronomy community. The Institute is able to establish mechanisms for identifying suitable individuals and groups which could undertake the preparation of tertiary information, i.e., the reviews and compilations essential to the total information process (Herring, 1968 , National Academy of Sciences, 1968), and offer them bibliographic support including textual backup, as well as financial support to encourage the preparation of these condensates of the literature. The complementary nature of the two approaches of horizontal and vertical integration can be seen in Figure 6 where the first approach, limiting the amount of information handled for each item, is able to extend its scope widely across disciplines. The second approach, limiting the scope of its coverage, is able to extend the amount of information processed per item through the whole process. A combination of both of these approaches helps to "span" the full informational "space" and thereby comes closer to representing a total system. At the present writing, AIP is actively negotiating with the Institution of Electrical Engineers (London), operators of the INSPEC system, to "integrate" the two systems, both in production and in the dissemination of products and services, so as to make available the most complete system in physics, on a world-wide basis.

User-Oriented Services

By concentrating on a limited coverage of only the "core" journals of physics and by integrating its operations throughout the information process, AIP is able to offer a number of services in direct response to the needs of the users, which were not heretofore available. Thus, abstracts of articles accepted for publication will be available in a new indexed journal called Preview Abstracts in Physics and Astronomy. This concept of offering the user access to selected, current literature in several formats is further carried forth through the package of services of SPIN (Searchable Physics Information Notices), a magnetic-tape service; CPT (Current Physics Titles), a sectionalized (by subject area) printed version of the basic data elements in SPIN; and microfilm of the full text of the items announced in SPIN. Thus an institution with adequate computer facilities may run searches on the SPIN tape and determine the microfilm reel and frame numbers of articles on the film, so as to be able to supply their clients with a complete SDI service. Individuals at institutions without such computer facilities may utilize this efficient back-up service in response to their scanning of the printed "Index" to the film, CPT.

In addition, the film-tape combination makes it possible for AIP to search the tape according to established "group profiles" and use these to reproduce sets of articles to form repackaged journals, of greater utility to the individual user than the traditional journals whose content is largely determined by the authors and editors.

In 1972, AIP hopes to offer its film-tape-index package based largely on its own journals, to test this concept in the market place. Plans are underway to extend this to the other core journals of physics and astronomy. A schematic of this planned system is shown in Figure 7.

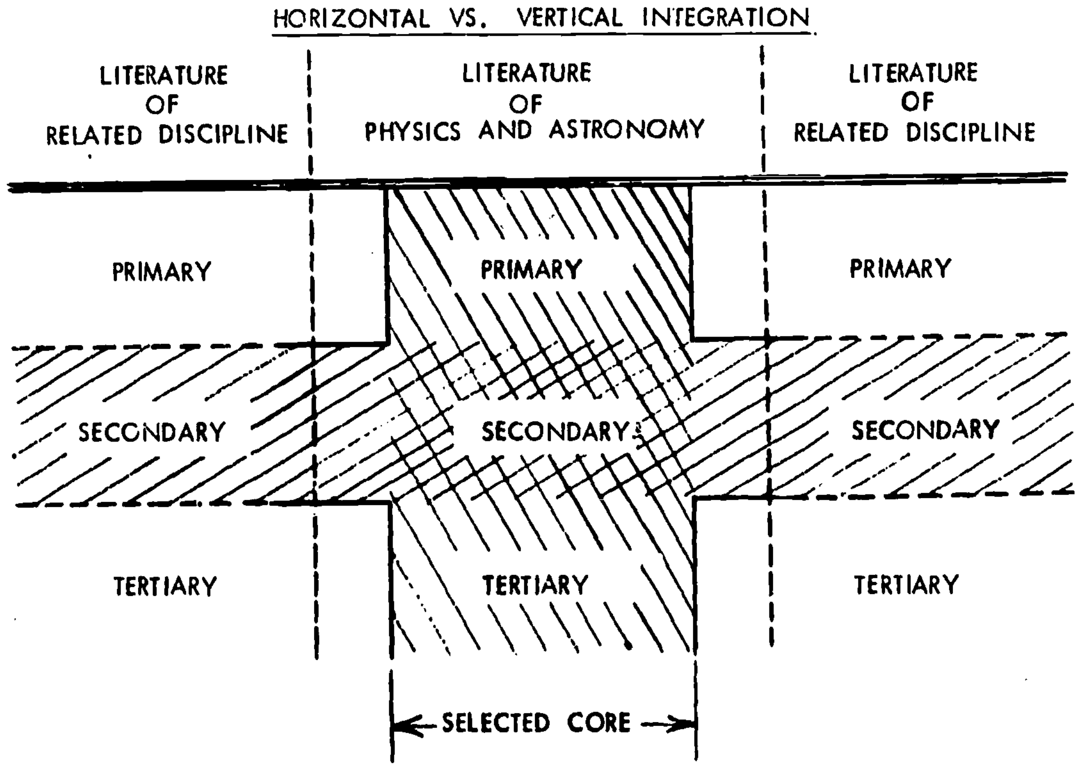


FIGURE 6

FLOW DIAGRAM OF THE
NATIONAL INFORMATION SYSTEM FOR PHYSICS AND ASTRONOMY

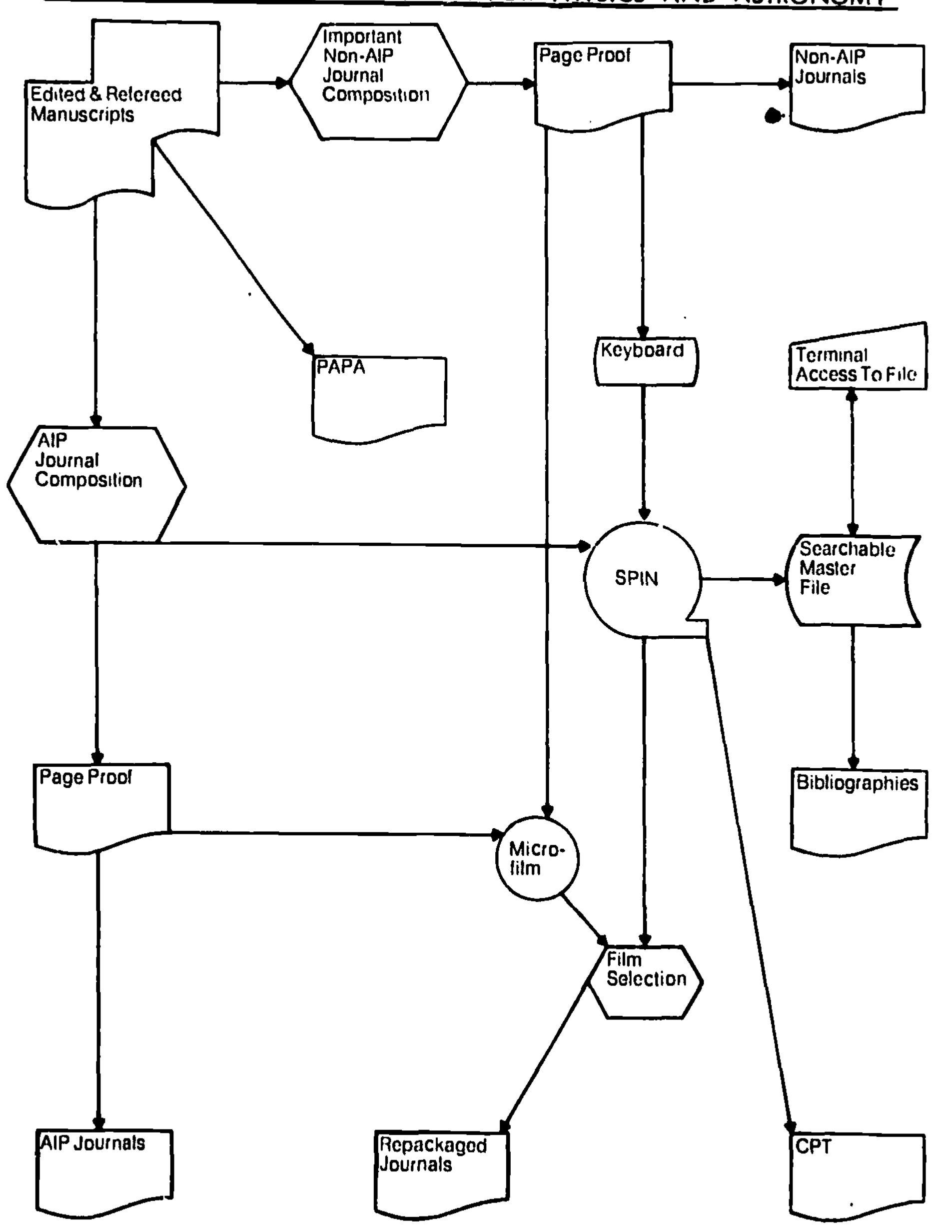


FIGURE 7

REFERENCES

1. American Institute of Physics, Information Div. A Program for a National Information System for Physics, 1970-1972. Report AIP ID 69R (PB 185 388) (August 1969).

2. American Institute of Physics, Information Div. A Program for a National Information System for Physics and Astronomy, 1971-1975. Report ID 70P (PB 192 717) (June 1970)

3. Herschman, Arthur. The Primary Journal: Past, Present, and Future. J. CHEM. DOC. 10 (1), February 1970, p. 37-42.

4. Merton, R. K. Behavior Patterns of Scientists. AM. SCHOLAR 38, Spring 1969, p. 197-255.

5. Pasternack, Simon. Is Journal Publication Obsolescent. PHYS. TODAY 19 (5), May 1966, p. 38-43.

6. Ziman, J. M. Public Knowledge: An Essay Concerning the Social Dimensions of Science. London, Cambridge Univ. Press, 1968.

7. Garvey, W. D. and Compton, B. E. The Flood and How to Survive It. JOHNS HOPKINS MAGAZINE 18 (4), 1967, p. 3-12.

8. Herring, Conyers. Distill or Drown. PHYS. TODAY 21, September 1968, p. 27.

9. National Academy of Sciences-National Academy of Engineering Committee on Scientific and Technical Communication. Scientific and Technical Communication; A Pressing National Problem and Recommendations for Its Solution. Washington, D.C., National Academy of Sciences, 1969.

OPERATION OF THE NUCLEAR DESALINATION INFORMATION CENTER
AT THE OAK RIDGE NATIONAL LABORATORY*

Karl O. Johnsson
Director, Nuclear Desalination Information Center
Oak Ridge National Laboratory
Oak Ridge, Tennessee

SYNOPSIS

The Nuclear Desalination Information Center is operated at ORNL to provide a source of technical information on distillation processes for desalting and on the use of nuclear energy for these processes. More than 1500 abstracts of technical literature have been prepared and stored in a computer-based system with provisions for retrieval by subject, author, and organization. The mechanics and philosophy of operation are described.

The Nuclear Desalination Information Center at the Oak Ridge National Laboratory originated in 1965 when the Office of Saline Water (OSW) of the U.S. Department of the Interior requested the development of a computerized system for the storage and retrieval of literature on distillation processes for desalting seawater. Initially the research and development reports published by the OSW were abstracted and indexed for the information system. Subsequently, the OSW has supported continued work of the Information Center to add to the system all published technical information related to distillation processes for saline water. In 1966 the U.S. Atomic Energy Commission (AEC) provided support for additional information services covering material relating to the use of nuclear energy for desalting.

Purpose of the Center

The basic objective of the Center is to provide ready access to published information on desalting distillation processes and the use of nuclear energy for desalting. It should be possible to retrieve information according to subject, author, organization, and other descriptions. The information system access should be provided in a form so the scientist in the field will have at his disposal a comprehensive coverage of the existing literature.

---

*Research sponsored by the U.S. Atomic Energy Commission under contract with the Union Carbide Corporation.

The Center is meeting this objective in several ways. First, it publishes indexed lists of abstracts of the literature on a continuing basis. Second, it provides information for the preparation of review articles and conducts special surveys on various aspects of desalting. Third, it answers specific requests for information. The Center endeavors to provide information to workers in the desalination field or to give specific directions as to where the information may be found. It does not act as a library or supply center and does not try to keep full-sized copies of all reports or documents included in the Information Center. Microfiche copies of almost all articles covered are kept on file; these usually serve when a member of the ORNL Nuclear Desalination Program staff needs the complete report. Enlarged copies of any page or pages can be made in a few minutes.

Operation of the Center

An outline of the general process the Center uses to prepare its publications and to have on hand the answers to questions may be of help in understanding the extent of information available. The steps include locating new information as it is published and obtaining a copy of it, writing an informative abstract (if one is not included with the article) and selecting indexing terms, putting this data into a form for computer storage, and then storing the information on magnetic tape. For retrieval it is necessary to structure the questions for machine searching so as to obtain all the desired information without too much unimportant data.

Taking these steps one at a time: To become aware of the existence of published information, we scan current periodicals which provide considerable coverage of information of interest to desalting. These include such journals as Desalination, Environmental Science & Technology, Chemical Engineering Progress, Power, Water and Water Engineering, and others. We also consult several abstracting journals such as Nuclear Science Abstracts, Chemical Abstracts, Desalination Abstracts, and Water Resources Abstracts. Most of the articles listed in the abstract journals will already be in the system, but these journals do provide a valuable check list and also provide a supplement of sources not otherwise covered. We have permission to copy abstracts appearing in Desalination Abstracts and they, in return, use any of ours which they may not have already covered. Reports by various government agencies and their contractors constitute another source. ORNL automatically receives copies of reports issued by many of these groups and has information exchange agreements with a number of information centers and libraries. Another group of reports received on a regular basis is the Office of Saline Water Research and Development Progress Reports. These latter reports cover the field of desalting by many processes; however, we include only the ones relating to the distillation process.

The next step is to obtain a copy of the desired article or report. In many cases we already have a copy of the report or journal, but in some cases a copy must be obtained from outside the group. ORNL has some 10 information centers and a large Central Research Library. One member of the library staff is assigned the responsibility of obtaining copies of papers, books, or reports for all the information centers.

New information is first scanned briefly to see that it actually is pertinent to the desalting program and within the scope of the Center, since titles can be very misleading. (Including keywords in the titles certainly helps in locating material in permuted lists such as Chemical Titles.) Articles containing useful information are sent to an appropriate member of the Nuclear Desalination Program staff for abstracting and indexing. There are about 15 members of the staff who abstract and index articles on a part-time basis. We try to match the subject matter of the articles with the special knowledge of the staff member so that he is abstracting material closely related to his field of specialization. Many times the article is one which the person would normally read as part of his regular work. We prepared an indexing and abstracting guide to give instructions to the staff and for others who might be interested. The guide is written in an informal manner with a minimum of jargon or information science terminology. The emphasis in the guide is on helping the abstractor-indexer determine how to abstract and index the material in the most informative and useful manner possible.

The abstracts include key data and results either in words or numbers as far as is possible within a limitation of 200 words. Rather than just telling what the report is about, the abstract should provide the information a researcher needs. In this way it is often possible for a person using the abstract to obtain the needed information there without ever seeing the original paper. We do not try to judge or evaluate the work in the abstract but rather to paraphrase the reported results in a few words. As more and more journals are requiring 150- to 200-word abstracts to be supplied by the author, the task of abstracting will become easier. The Distillation Division of OSW now requires all of its contractors to include an abstract and a list of indexing terms with their reports to OSW.

The indexing terms used by NDIC are patterned after the ones listed in the Engineers Joint Council Thesaurus. Many of the terms in our thesaurus are the same as theirs, and most of the ones we have added are more narrow terms to provide for a greater depth of indexing than is possible in a general engineering listing. Indexing is done on the basis of use for the information as well as the specific content of an article. That is, we will index an article on packed towers under feedwater treatment as well as deaerators and carbon dioxide since packed towers are sometimes used in feedwater treatment. The thesaurus contains at present about 550 terms. These terms were selected by asking Program staff members to supply a list of terms they felt would be needed to describe information pertaining to their work. The lists were then combined and each term compared with the Engineers Joint Council listing to determine which was more logical or descriptive. In the course of abstracting some 1500 papers, about 50 new terms have been added to the thesaurus.

For the transfer of the abstract and keywords to a computer storage and retrieval system we use the form shown in Figure 1. As is shown, there is an accession number for each item in the system and then a set of coded information upon which a machine search can be made if desired. These include the type of original material such as journal article, report, or patent; a category indicating the area of the desalination field covered; the corporate author; and the date of issue. Next appear the names of the authors, the title, the corporate author, and the reference. This is followed by the abstract. At the bottom is the list of keywords assigned to the article and a place for new keywords to be added to the master list.

All the information shown here, along with a few extra numbers and symbols for computer control, are keypunched and then transferred to magnetic tape at the Computing Technology Center, which is part of Union Carbide Nuclear Division's operation in Oak Ridge. They have an Informations Systems Department specifically to handle the ORNL and AEC information work in Oak Ridge. Of the 10 centers at ORNL, four use

NUCLEAR DESALINATION INFORMATION CENTER

BIBLIOGRAPHIC REPORT

REVIEWER: W. H. Kelley

REPORT NUMBER OR JOURNAL REFERENCE: ESTH-02-06-420

DATE OF REVIEW: 4/16/69

ACCESSION NUMBER	CARD NO	DOCUMENT TYPE	EVALUATION	SUBJECT CATEGORY	JOURNAL ABBREVIATION	SIGNIFICANT DATE MONTH	DAY	YEAR	CLASSIFICATION	FILE	SUBJECT NUMBER
00767	01	JE		50	ESTH	06		68			

CORPORATE AUTHORS: PRC OSW KAISR

ABSTRACT

02 Arad N + Mulford SF + Wilson JR

03 DESALTING PLANT DOWN TIME PREDICTED BY FORMULA

04 Planning Res. Corp., Washington + Office of Saline Water, Chula Vista, Calif. + Kaiser Engineers, Oakland, Calif.

05 Environ. Sci. Technol. 2(6), 420-27 (1968)

06 An analysis of the factors which affect the plant operating factor of proposed large dual-purpose electrical power-desalting plants. The plant factor for large ... can be conservatively estimated to ...ctor and electrical power plant) factor is ... ...wer, there is no above. The product of all these items is the overall plant ope... which is found to be 0.846 (the water plant factor is 0.942 for a 30-year plant life).

KEY WORDS

07 *PLANT AVAILABILITY + *PLANT FACTOR + *BREAKDOWNS (FAILURES) + *LEAKAGE + CORROSION PROBLEMS + DUAL-PURPOSE PLANT + EQUIPMENT FAILURES + PUMPS + STEAM JET EJECTORS + TUBING (METAL) + OSW SPONSORED

PROPOSED NEW KEY WORDS

08

Figure 1. Bibliographic Form

essentially the same computer program which was written by the Computing Technology Center. The computer services which can be provided with such an arrangement are far greater than would be practicable for any single center unless it is very large. The information programs are written in COBOL for use on an IBM 7090 computer.

Output

When several hundred additions have been made to the storage tape, the computer people prepare a printout of these additions indexed by the keywords and by authors. These listings, which include the reference data, the abstract, and the keywords, are being prepared at three- to six-month intervals at present. In the listings, there are nine subject categories to organize the information under specific areas of technology and interest. Since many reports cover information in more than one area, the abstracts for these reports are listed in two, or sometimes three, categories. Although categories have been established for several desalination processes, the distillation process is the only one covered in appreciable depth at the present time. The other processes may be included at some future date. Some reports in the information system may not appear, from the title, to be appropriate; however, they are included because information or data contained in the document was found to be of use in studies of some phase of nuclear desalination. The categories and their general coverage are:

Energy Sources: Information on sources of energy for desalination plants, especially nuclear reactor power plants, but also including other energy sources such as coal-, gas-, or oil-fired plants for economic comparison as alternatives to nuclear energy.

Energy Utilization: Covers all aspects of the coupling of the energy sources to the desalting plant, with emphasis on efficiency, economy, reliability, and flexibility to adjust to various load conditions.

Seawater Distillation Processes: Information on the various distillation processes, plants, and equipment, including materials of construction, heat transfer, fabrication methods, and design information useful in the development of distillation processes.

Other Desalting Processes: Processes other than distillation, such as hyperfiltration, freeze process, ion exchange processes, etc.

Overall Plant Studies: Broad studies which cover the overall or integrated desalting plant.

Siting Considerations: Factors involved in the choice of a site for a desalting plant.

Industrial Applications: Studies on by-product chemical production, recovery of chemicals from seawater, agro-industrial complexes, and waste water treatment.

Water Utilization: The needs and uses of water for agriculture, industry, and domestic consumption. Also studies on water sources, water distribution problems, and projections of future water needs in various areas of the world.

Miscellaneous: General papers on desalting and studies covering the physical and chemical properties of seawater.

Services Available

The Center's services are available without charge to groups in the United States including governmental agencies, research and educational institutions, and industry. The services include:

1. Periodic lists of abstracts of new additions to the information system.

2. Reports or surveys on special topics relating to desalting.

3. Answers to technical questions on nuclear desalination.

4. State-of-the-art reviews prepared by the technical staff of the Nuclear Desalination Program at ORNL.

5. Computer-prepared searches of stored information selected by keywords applicable to the users' needs.

Reports from the Center are issued with ORNL-NDIC numbers and distributed according to USAEC standard distribution under the Reactor Technology category (UC-80). Copies can be purchased from the National Technical Information Service, Springfield, Virginia, and are available for use at the USAEC Depository Libraries throughout the world.

Philosophy of the Center

The Nuclear Desalination Information Center is operated by scientific personnel for scientists. That is, the coverage is intended to cover technical information rather than popular press type items. We do not include general news releases, congressional budget hearings on desalting, or the reports of the advantages of desalting for water-short areas of the world. There undoubtedly are needs for cataloging all printed references to a technical subject, but to include these in a technical information system would only serve to inflate the size of the system without contributing to its technical usefulness. In fact, including numerous nontechnical references in a system tends to discourage the use of the system by a scientist who is looking for data to use in his work.

Another question which arises is whether or not to include references to poor or even erroneous data (this, of course, is much easier to recognize in hindsight than at the time an article is accepted by a journal for publication). We have taken the position that all material must be included if we are to serve the technical community honestly. Omitting a "poor" reference will only lead the information system user to wonder if the reference was overlooked. The place to judge or evaluate the work of others is in review or state-of-the-art papers wherein an evaluation is given in the opinion of the author of the review.

INFORMATION SYNTHESIS IN TECHNOLOGY FOR SOCIAL BENEFIT

P.K. Bhattacharjya
Central Fuel Research Institute, Dhanbad
Bihar, India

SYNOPSIS

While attempting to enunciate the gamut of knowledge contained in the term "technology", the quantitative growth in its literature and its consequent impact on users are analysed. Information science has the responsibility to prevent repetition of research in technology and in industry; to make the consumers aware of the facilities offered. Monitoring intelligence through different media on research and development activities is the rightful domain of the professional information worker. It is ultimately the society that benefits most by the proper interpretation of technology via information.

Man is living now in the exciting new world of technology. A word much in fashion today, technology may basically be termed as the assorted human knowledge in science, engineering, economics, sociology and law in their broader sense, which engages itself in the materialistic uplift of man. Elaborated a little, science unravels nature's mysteries for potential employment; engineering shapes scientific findings for corporeal utilisation;economics studies the utilitarian demand and supply of resources made available; sociology records peoples' needs and their fulfilment; whereas law ratifies or rejects human rights and obligations posed by the preceding aspects.

Due to this growing knowledge, there has been a virtual information explosion in the modern world. By way of profession or passion, more and more people try hard to educate most others through all sorts of spoken and written words in journals, books, radio, television, stage, films and live lectures, discussions and even gossip. From this babel of audio-visual media of information dissemination arises the need for specialization, particularly in technology where information is generated constantly and causes the inevitable unrestricted publications inflation. The number of scientific periodicals in 1950 were about 50,000 according to the compilers of the "World List of Scientific Periodicals" and the growth decadewise exceeds (in all probability) 60,000 and 70,000[1]. Another author says that about 27,000 scientific periodicals, ranging from weeklies to annuals are published each year, where there were less than 5,000 in 1900 and 2 only about 300 years ago[2]. Yet another source quotes very modestly that in physical sciences alone, there are more than 3,500 journals[3] (he is perhaps not including engineering and other journals.) Although estimates on the number of technical journals produced differ, the number of papers published and volume of printed words used are increasing at a formidable rate.

II Information Systems for Specialized Applications

The modern author perhaps feels that if an impression is to be created, it is to be effected in the quickest time. Undue haste in many cases diminishes the quality of a publication as premature and defeats the very purpose of communication between the producer and consumer of information. Technological readers being professionally disciplined in science or in any other field mentioned above, can certainly better imbibe information than the administrator or manager, thus enabling them to be more contemplative, penetrative and discerning. As they are not impressed by mere journalistic "jargon", or "gimmicks", caution has to be exercised in presenting them the information, to sustain their interest for direct employment into their work.[4]

All industrial organizations and technological institutions now have an information service unit. Recent advances in technological research and consequent industrial development have created a new profession of information workers, to fulfil the growing need for selective information dissemination.

In technology, the professional information worker is designated the "Information Scientist" who is responsible for monitoring intelligence on all aspects of research and development activity. Now when known traditional demarcations between isolated areas of knowledge are being progressively dismissed, and findings in one field may be of help to a researcher working in a seemingly quite unrelated subject, it has become the added responsibility of the information scientist to educate the latter, so as to virtually police against costly and wasteful repetition of research by introducing a methodical system of retrieval, storage and dissemination of technical knowledge.

In industry (or utility), information service is handled by the "Public Relations Adviser" whose duty involves informing outsiders about the inside functionaries and facilities offered and recording the social reaction for use of the management. Obviously, information in this case has some commercial bias, unlike that in research.

A modern producing firm tries to know about the activities of competitors, the developments taking shape in similar or allied industries at home and abroad. Also, it should take stock of situations arising out of new techniques envisaged and new products evolved elsewhere. This type of information exploration is pursued in research institutions too. As technological research precedes industrial production, the researcher is just as keen to know how others are faring in the related subjects, which the industry can profitably exploit for the service of humanity.

Every now and then, something happens in technology and industry. It is the job of the information service to sort out what really makes news, from what merely makes noise. In this aspect, silence on reality is a guilt. If information is the watchward, sophistication of techniques should be introduced as far and as early as practicable with the help of gadgets, machines and equipment. When the demands of growing population are soaring more steeply than nature can sustain, the onus lies on the technologists in forcing nature to change her nature[5]. Certainly organised information in technology is vitally important now, to help derive the maximum benefits from minimum resources available.

Presentation of technical information varies according to the purport and readership eliciting information. The media through which research information are communicated are classed as "Technical Documents" which require expert processing to bring out their intrinsic value. They comprise, (i) Technical Write-up, (ii) Research Note, (iii) Technical Report, (iv) Research Paper, (v) Patent Specification and (vi) Technical Monograph.

Technical Write-up - Only a publicity hand-out released for general information on suitable occasions. This does not qualify as technical literature properly. It is akin to a "sales leaflet" in industry which is more carefully worded to guard against any infringement and legal right or obligation[6].

Research Note - Generating from the laboratory experimental data and/or plant operation log-book, it serves as an important ingredient in creating technical reports and research papers. It too does not generally constitute an item of technical literature.

Technical Report - Embodying some definite research findings or an appraisal of investigations undertaken, the technical report (brought out occasionally, and also periodically) forms an important part of technical literature, which may not necessarily be published. As a compromise between research note and paper, it is produced for internal use, record and reference, and help in formulating a policy for subsequent research programming.

Research Paper - Logical metamorphosis of research note and technical report which ultimately breaks into print. Meant for widest circulation through a periodical, the author either establishes an important finding of his own or reviews a particular aspect of research carried out previously by other workers. However, as has been the trend today, many a research paper does not disclose completely the nature of the work, because of the competitive prospect of patentability of the research conducted.

The main difference between a technical report and research paper is that while the former is produced at the conclusion of research, the latter implies no conclusion. The research paper has one significant value - it definitely contributes to the domain of knowledge and as such, it should not be denied its due role, by restraining its spirit into the technical report form.

Patent Specification - A very important item in technical literature serial is the patent specification.

Culmination of a specific period of research, the award of Letters Patent is the ultimate prize a research worker always tries to win. However, research and patent do not always go hand in hand. There exists always a chance factor, as a patent is never contemplated from before - rather it is grabbed (sometimes as a side issue) during the course of a research assignment. It is the responsibility of the information service to judge the value of work from technical documents submitted to it and suggest further development for patenting the process, product or equipment; if it merits, to arrange filing the patent specification immediately for publication by the competent authority.

The information service can be of much help to the research worker in retrieving relevant literature on patents. The latter should also be apprised of the latest implications of the Indian Patents Act, (Amendment 1970), especially where research relates to the patentability of inventions on food, medicine and drugs (including some special chemicals), which fulfil the basic needs of the common people. It is important to educate the inventors (who draft the patent specification) that the implications suggested safeguard the patents only for the process of manufacture and not for products.

It also is the responsibility of the information service to educate a research worker and aspiring patentee about what "invention" and "inventor" mean in the patent system. The service must make the research worker aware that it is not the end to his/her inventing any new process, equipment or product (or any new and useful improvements thereof), unless he/she secures timely protection for the invention by a patent, thus preventing others from usurping the advantage.

A patent confers the legal right to economic exploitation of an invention by the State to the inventor. According to the national patent system (as it exists), an invention broadly means any of the following: (i) a new process or manner of production or manufacture, (ii) any product obtained by the new process or manner of production of manufacture and (iii) a new apparatus, machine or plant involved in creation of a new article. It is important to note that the new process of manufacture of a "vendible product" stresses "manufacture" which means something of tangible and substantial nature turned out in a factory, in contrast to something produced through agriculture or similar natural means.

The patentability of an invention depends in large part upon drafting the important legal document known as the "Patent Specification" (provisional or complete) which, submitted along with an application, describes the nature of the invention and explains clearly how best it could be put into practice. Insufficiency of claim and ambiguity in presentation of facts should be eliminated, otherwise the patent specification would fail to survive the test of validity in a Court of Law.

The growth of patents directly promotes the growth of industries, which in turn provides increased employment for the citizens and wealth for the nation. And in this context, the importance of the role of patent specifications towards conceiving the Letters Patent needs hardly be emphasised further.

Technical Monograph - As a majestic rear-guard to all the technical documents mentioned, the technical monograph incorporates their essential features to present one single subject-matter as a printed text. Information contained in a technical monograph has a uniquely permanent value as a storehouse of knowledge which definitel aids the industry via technology.

COPYRIGHT is an important point in the technical writer's career. It is helpful to note that the authorship of the technical paper or monograph is protected by the Copyright Law. India, since 1958 a member of the Universal Copyright Convention (UNESCO, Geneva, 1952), and also a member of the Copyright Union of the earlier Berne Convention since 1953, extends the same facility to the author as any member.

Thus, it is now readily accepted that synthesis or building-up of an efficient information hook-up in technology induces industry into action. "At a time when specialization makes it increasingly difficult for scientist to communicate with fellow scientist, and difficult for scientist to communicate with humanist, it is vital to utilize every means available to explain what is happening in the technological world around us. Through the media of the written and the spoken word, and through film and television we must avail ourselves of these men who can interpret scientific achievement in terms that the trained humanist mind can understand"[7].

After all, it is the people who ultimately derive benefits from technology through industry. "Technology can be directed creatively so as to bring human societ into close harmony with its natural environment. It can be made to create more wealth with less waste both waste products and waste of human and natural resources. It can be made to create beauty where we have let it spawn ugliness. It can be made to bring man both greater security and more individual freedom."[8]

REFERENCES

1. BHATTACHARJYA, K.: In: "New Trends in Scientific Communication," Oxford Book and Stationery Co., Calcutta.

2. EDWARDS, D.G.: The Work of an Information Scientist, Coke Oven Managers' Association Year Book, 1968.

3. HOLT, C.C., and SCHRENK, W.E.: Quoting a figure in American Documentation, January 1968.

4. MARDER, D: In: "The Craft of Technical Writing," The Macmillan Company, New York.

5. RANGANATHAN, S.R.: Director, Documentation Research and Training Centre, Bangalore, India.

6. GODFREY, W. and PARR, G.: In: "The Technical Writer," Chapman and Hall Ltd., London.

7. MACMILLAN & CO. LTD.: Advertisement in the Times Literary Supplement; October 25, 1963.

8. SEABORG, G.T.: In Proc. Symp. on "The Place of Values in a World of Facts, "Nobel Foundation, Stockholm, Sweden (1970) - an address.

Session Two - Discussions

INFORMATION SYSTEMS FOR SPECIALIZED APPLICATIONS
Chairman: Dr. M. Cremer (Germany)

CHAIRMAN (Dr. Cremer): We know from the SATCOM report of the National Academy of Science and National Academy of Engineering, U.S., that there are more and more special user groups with special needs. In our planning, we have to take into consideration the relation between these general systems and the special systems and special applications. The papers this afternoon deal to some extent with this problem.

DR. U. SCHUTZSACK (Germany): (Comments to Buntrock paper.) In 1968 a panel was set up by the Director-General of FAO to study the question of international cooperation in agriculture. This panel set up a study group which did two things. One culminated in the report of Dr. Buntrock which you have just heard. The second was a report recommending further study of the problem and saying that it would be feasible to divide the agricultural information into two levels: level one and level two. Level one should be a system which would contain bibliographical data and indexing, but no abstracts. Level two should be specialized information system with in-depth indexing and abstracting.

The next step was to set up a study team consisting of three groups: the first concerned with management and organization, the second with mechanical problems, and the third with scope and coverage of agriculture. The report of the latter group recommends that the scope should be the scope of FAO, i.e., food, agriculture, forestry and fisheries. The coverage should be conventional and non-conventional literature, should cover the whole field of agriculture and should be as quick as possible; it should have bibliographical data and some kind of indexing, but no abstracts.

The main responsibility for organization will rest with FAO so as to make it a truly international system.

There should be a maximum of five lineal input points. These should be responsible for the input of conventional literature, e.g., an input point for the developing countries; an input point for literature from English-speaking countries; an input point for literature from Eastern European countries, etc. The input should be on magnetic tape in the mapping center. The second input would be by national centers and the national centers would be responsible for the input of national literature. FAO would be responsible for mapping and coordinating the system. Output would be, first, a printed bibliography and second, a magnetic tape.

There now exist some large general centers like MEDLARS, Chemical Abstracts, etc., which are doing what we call level one and level two. Here is envisaged a general service which would be responsible for bibliographic data and some kind of indexing only. This service could be used by abstracting and documentation services, by specialized documentation centers, and by others as a basis for the preparation of a level two service.

There is a very pronounced need for special services. We have set up in our Institute an information service for food technology and also one for aquatic sciences, and I am of the opinion that these are serving the scientists better than a general center can. Scientists have more direct contact with these specialized centers and, moreover, these centers are preparing the information in more adequate form, with in-depth indexing and abstracting, etc.

QUESTION: To Dr. Buntrock. To apply a system which is only geared to the needs of agriculture is a mistake because agriculture ties in with a lot of other disciplines. It would be reasonable to apply here a classification and documentation system which will tie in internationally and nationally with other disciplines.

MR. [illegible]. KEREN (Israel): To Dr. Buntrock. Why in your recommendations to AGRIS, have you not taken an INIS-like approach, wherein each country puts the input of what is being produced in the country into an international system, and in return the international system supplies the local member with the total accumulated knowledge?

DR. H. BUNTROCK (Luxembourg): To Mr. Keren. Why did the AGRIS study teams not recommend the INIS system for agricultural documentation? It is not yet published so I am not in a position to comment.

DR. SCHUTZSACK: To Mr. Keren. Concerning INIS, we have studied this question and found that there have been some very bad experiences in using national centers as input centers. There are time lags. We prefer regional input to national input centers, with the regional centers responsible to the management center at FAO.

PROF. DR. P. LOOSJES (Netherlands): To Dr. Schutzsack concerning AGRIS. The figures of the Auerbach research on the Bibliography of Agriculture a few years ago were 50 to 60% coverage. The Bibliography of Agriculture people are trying to improve. If this improvement will achieve say 60 or 70%, is it still necessary to start another service in Europe on another level, with the improved Bibliography of Agriculture already on the market?

DR. SCHUTZSACK: To Prof. Loosjes. If the Bibliography of Agriculture would be willing to meet the following requirements: that it be quicker than it is now, have complete coverage, and be classified, then why should it not do the job?

MR. P. ARIES (France): To Dr. Buntrock. (In French). Summarized by Dr. Buntrock. Mr. Aries asked for a more detailed explanation of the distinction between level one and level two.

DR. BUNTROCK: To Mr. Aries. Level one comprises all universal services with the following characteristics: the system comprises more than the total of a special service; that is, it includes information from fringe and fundamental sciences. Level one need not be indexed in depth. A rough categorization would be enough, as it has to be quick and to serve primarily as a current awareness service. Level two, on the contrary, is the total of specialized documentation areas, each one specializing in a very limited field. The items in level two are characterized by in-depth indexing and abstracts, and should provide retrieval means. There is a question whether both services can be combined in an integrated system like MEDLARS. For agriculture it will be recommended to establish two levels. Level one, e.g., Bibliography of Agriculture; level two, a specialized center.

MR. KEREN: To Dr. Buntrock. Your survey found that in secondary services in agriculture, a cummulative number of about 1 1/2 million abstracts are produced each year, which means that each item is repeated about 7 times. If we consider that to prepare an item for a mechanized abstracting journal costs about $20, we have an expenditure for each item of about $120, whereas only $20 of that are really necessary.

DR. CREMER: I think that this is a point which is not only true in agriculture, but in nearly every field. To some extent you cannot avoid duplication, but I think it is one of the tasks of information policy of international organizations to handle that.

DR. BUNTROCK: To Mr. Keren. You quoted me as saying we found 1 1/2 million references. This includes bibliographic citations plus bibliographic citations and abstracts, but there are only 850,000 abstracts. So the duplication rate is "only" four.

PROF. H. ARNTZ (Germany): Duplication is not a numerical problem only. A field like agriculture includes other sciences like biology, chemistry, etc. Three or four abstracts of the same paper can be completely different if they serve different groups of users and emphasize different points of view.

DR. CREMER: To Prof. Arntz. I think exactly that is one of the differences between the generalized and specialized systems. I think it is a problem, because duplication is a problem of finance and manpower, so we must have a kind of compromise.

DR. BUNTROCK: To Prof. Arntz. I would agree that duplication is not a numerical problem only. I agree as far as abstracts are concerned, but as far as bibliographic citations only are concerned, it is a waste.

MR. U. BLOCH (Israel): I am shocked by the 1 1/2 million items which cover 60% only. I find that in other fields many papers I read say nothing new at all, and therefore quite a number of them are truly superfluous. I wonder whether it would not be worthwhile to use the $120 mentioned before to repackage what has not been said before. Then you can do four or five different abstracts for the same publication and still produce something that does not create tons and tons of words which I cannot imagine how you can control with even up-to-date computers.

DR. BUNTROCK: To Mr. Bloch. We analyzed about 1 1/2 million secondary references, bibliographic citations referring to primary publications. According to estimates there are 200,000 to 250,000 new publications only. So it is not the amount of 1 1/2 million items to be processed and computerized, but only 200,000, to 250,000.

MR. F. LIEBESNY (UK): To Mr. Bloch. On the subject of duplication, I would like to introduce a third conference language - music. An information scientist did a survey to find out how often Beethoven's Fifth Symphony is played every year. 10,000 performances are given a year, which obviously is a duplication factor of God knows how many times, but there is always one person who hears the Fifth Symphony for the first time. And there is always one person who reads an article for the first time on a certain subject.

MRS. L. VILENTCHUK (Israel): Duplication is certainly to be avoided or anyway minimized. But slanted abstracts in my experience do not answer the real needs. It seems to me that the right way to eliminate duplication is very deep indexing with many entry words giving all aspects of the article. This will enable the person who wants only a certain aspect, to find it. Slanted abstracts, however, are often completely useless because they hide part of the information and emphasize information which is hardly there in the article.

PROF. V. SESSIONS (US): *To Mr. Keller.* Many people in the States are very concerned about what is in the census tapes, whether they will need access to the census tapes and why. Perhaps Mr. Keller might mention data bases in alcoholic studies. Urban information systems have not worked. The social sciences, educational resources information people and the urban project ID for HUD are almost the only projects done. What has been done at Rutgers University? When we got started, people thought the techniques in the hard sciences were not applicable to the social sciences. There are some problems, but they are not insoluble ones.

MR. M. KELLER (US): *To Mrs. Sessions.* Our system constitutes a complete data base for everything about alcohol, excluding technology of manufacture and industrial use. You can find out from us anything that happens when people use it - whether for pouring into the mouths of live people or for washing the bodies of the dead. Alcohol problems do not belong to the social sciences only, but are multidisciplinary, so that information in physiology, chemistry, biochemistry, neurophysiology, experimental psychology, etc., are incorporated. The "biological" part today is actually larger than the "social" part.

DR. CREMER: *To Mr. Keller.* The difference between educational projects and information projects exists also in the humanities. A university library for the students in philosophy is an educational project, but in our country we have at the Duesseldorf University the first attempt at fully computerized information systems in the field of philosophy.

DR. N. HENRICHS (Germany). At the Duesseldorf University we are indexing in-depth philosophical journals, 20 to 30 to 40 entry words. We don't use any classification. This is a special problem in the humanities. We have two retrieval systems. Those interested in more details may write to the Philosophical Information Institute, Duesseldorf.

DR. M. KESSLER (US): I'd like to mention one information system in musicology which can serve as a lesson in other fields. When I came to see the system, it was beaitufully successful. I started to search for the reasons for this success, and I think one of the reasons was that the musicologists didn't have a computer expert, so they went and hired one. In other fields, such as chemistry and physics, everyone is a computer expert and knows exactly what the system ought to be. But these people just went out and hired someone to do the job. I think the other reason for their success is that they had nothing to begin with and so had no vested interests. They have succeeded where many of the sciences have not. They have an international network; every country provides abstracts and indexing and sends them to New York, and a musicologist there collects it and manages it as a part-time job. It is a sizeable organizational problem managed as a pilot project by the American Council of Learned Societies (ACLS). This experience has been so cheap and so successful and so universally satisfying that there are plans of going into the history of art on a similar basis.

PROF. ARNTZ: I would like to mention the educational function of scientific journals. When scientific journals were created in 1665, their function was to educate the scientist, to give him the background from which he could study a special field. The scientific journal has largely lost this function because of overburdening. Chemical

Abstracts, for example, has to go through 11,000 journals to find the chemical literature worthwhile abstracting. If we can develop computerized information centers giving narrow, specific information with nothing of the educational background, and can do it in such a way that the scientific journals can go back to what they originally did - to give background to scientists - the big information centers can give the detailed information, while the journals can take over the educational function so badly needed for scientific work.

MR. KELLER: To Prof. Arntz. I'd like to agree very strongly with your remark about the need for the primary scientific journals to regain their function as educators of scientists. If a man sends in an article he is told it will be published only if he cuts it by 40 to 60%. What he cuts out are his ideas. So we get a lot of short articles, and lots of tables. I don't say thay are not useful and shouldn't be published in some form or other. But we have to have some journals that will go back to the longer article in which a man says what he thinks and describes what he has been doing.

MR. LIEBESNY: To Mr. Kingwill concerning Mr. Schauder's paper. Do the studies carried out relate to the use of computers in the preparation of Braille material?

MR. D.G. KINGWILL (South Africa): To Mr. Liebesny. They do produce "talking" books using existing technology, but not in embossed forms. This they have obtained from overseas sources.

MR. LIEBESNY: To Mr. Kingwill. Is any such work being conducted not necessarily by Mr. Schauder?

MR. KINGWILL: To Mr. Liebesny. I am not quite sure. The catalog is for use by librarians, but not the other materials.

MR. KEREN: (Comments to Wollman-Cohen paper). I think it should be explained that this project is a typical example of the plight of the information scientist. The project was started several years ago with the idea of preparing a data-base on the information data files which are available. But when we couldn't get any financial support from our professional community, we turned to the administrative departments, and of course, had to adjust the project to suit their needs. When about five years ago, we started to investigate how we could use existing information about information storage and retrieval as applied to automated systems, we had the choice either to learn about the programs which are already operative or to develop them ourselves. We soon found out that with the traditional non-confidence one programer has in the program of another, the ready program approach would not work. Also, due to the rapid changes in hardware and the concurrent rapid changes in software, the updating of any detailed program data bank would probably be quite illusory. So we decided on the second approach, as the master file would probably change more slowly than the program itself. A system which describes the master file would be easier to update. The system has been in active operation at the Center as a pilot project, and it works. We are also working on a system for using data files in science and technology for other information types.

DR. CREMER: To Mr. Keren. You mentioned the technical data base in the field of science and technology and that this system may be used for other information types. What types?

MR. P. WOLLMAN (Israel): To Dr. Cremer. What we think we can do with this system is to combine data from different files. So if we have files in agriculture or biochemistry or other disciplines, we can know what we have in those files and gather parts of the files and do not need to collect previously collected data. Our system can help us know what there is in other files and help us decide how to utilize it.

PROF. SESSIONS: Bibliographic data files is the subject we are usually discussing. Very little attention has been paid to the documentation of substantive data files. I discovered at the University of Cologne that similar work is being done there. Incidentally, they seemed to have the best ideas on how to use U.S. census tapes, more than most of the people in the U.S. I think librarians will have to learn about the documentation of substantive data files, how to get into them, etc.

DR. CREMER: To Mrs. Sessions. Substantive data files are very important for the new types of information systems we are concerned with now: for data banks, data files of statistical offices, etc. In view of the real needs of governments and society, we should become more conscious of the information demands of society in fields other than science and technology. It is really important to prepare systems which will answer this demand.

SESSION THREE

information analysis

Authors of Papers

FOREIGN LANGUAGES IN INDUSTRY.

FELIX LIEBESNY
Transociates Ltd.

SYNOPSIS

The problem of foreign literature in the fields of science and technology is analysed with respect to its magnitude, the degrees of competence required in bridging the language gap, the availability of various tools used in translating and interpreting, the economic factors of translating, and the possibility of using computers.

The importance of the barrier which is caused by the existence of the many foreign languages used in the communication of scientific and technical information is often not fully appreciated. Even the very fact that there are languages other than one's native tongue and that there are altogether more than 3000 languages - leaving out many more dialects - is frequently overlooked or conveniently forgotten. It is of course also true that most of these languages do not make a significant contribution to the vast amount of documentation which is produced and that perhaps only some ten or so languages play an important part. Yet the size of the language gap is so great that it is really dangerous to ignore it or to minimize its significance. Although it is difficult to be very precise in assessing its quantitative nature, some recent analyses of the language distribution in the literature of science and technology - compared with earlier investigations - show some startling figures.

In 1957 the first serious attempt was made to obtain an accurate picture of the language problem when Dr. J.E. Holmstrom and G.A. Lloyd (1) studied 1000 scientific periodicals selected at random and found that 43.6% were published in English, 14.4% in German, 12.6% in French and 8.1% in Russian. These figures were accepted without much question for almost a decade. In 1965 A. Tybulewicz and the author (2) conducted a further study on a slightly different basis by analysing the languages of the original articles abstracted in a variety of abstract journals. This investigation showed three important points, viz.

a.) that English is still the most widely used language in the scientific and technical literature, accounting for more than half the world's output,
b.) that the second most important language is now Russian, representing some 20% of the literature, and
c.) that half of the world's scientists cannot read half of the world's scientific literature purely because it is in languages which they do not understand.

To put these percentages into absolute figures it means that about 500 000 scientific and technical articles are published annually in English and some 200 000 in Russian. The reasons for the preponderance of Russian over other languages in the field of science and technology are numerous and are obviously interrelated, but must include:

1.) the quantity of Russian literature per se has increased considerably, owing to the rapid rate of technological advance prevailing in the Soviet Union, which rate is certainly greater than that in many Western countries;
2.) the means of reporting Russian literature have improved greatly since 1957 when the launching of Sputnik I took place. This event prompted the whole,

perhaps rather wasteful, set-up of cover-to-cover translations which, however, has most definitely multiplied our coverage of the Russian output of scientific publications;

3.) the growing number of Russian abstracting journals (the Referativnye Zhurnaly) which scan comprehensively the less accessible Soviet periodicals, and

4.) a more noticeable willingness on the part of Russian authors to publish the results of their work.

One aspect which may have considerable significance in years to come is that the Chinese contribution, although still small, has risen by a factor of about 100 compared with 1961. Although the present picture shows that in the field of physics (3) the English, Russian, French and German contributions together make up some 97% of the total it may be useful to remember the old riddle in the translation industry: "What is the difference between an optimist and a pessimist?" Answer: "The optimist learns Russian, while the pessimist learns Chinese".

To make this large amount of foreign language material accessible and comprehensible to the potential user - the research scientist, engineer, manager, patent officer etc. - it is essential to bridge the language gap in an efficient and effective manner since the user is rarely capable of properly understanding an original article in a foreign language; he will almost invariably demand a rendering of it into his own language. But in industrial information work it is very rare to find people who are equipped to deal with much more than one or two foreign languages. The term 'equipped' may perhaps require somde definition because there are obviously several degrees of competence any of which would be quite adequate to meet certain needs of the organization which the information centre is serving. The principal levels of competence are:

i.) to identify a foreign language: this may not be such a mean task at times since quite a few so-called linguists can be quite easily baffled when trying to distinguish between Spanish and Portuguese, or between Chinese and Japanese. Having identified the language the information scientist would then pass the document to the translator specially versed in that language to process the original either by abstracting or translating it. Thus at this level the total requirements would be merely a knowledge of the respective characteristics of languages such as the use of capital letters for nouns which indicate a German language text or the tilde (~) sign to indicate a Spanish origin or the occurrence of the ø sign to show that the author wrote in Danish or Norwegian, but not in Swedish. In this respect the following publications may be useful:

PIETTE, J.R.F. A guide to foreign languages for science librarians and bibliographers. 2nd edn, London, Aslib, 1965, 53 pp.
VON OSTERMANN, G.F. Manual of foreign languages. 4th edn, New York, Central Book Co.Inc.,1952, 414 pp.
'THE INCORPORATED LINGUIST', the journal of the Institute of Linguists, London, has published a series of 'Language Information Notes' on more than 20 languages.

ii.) to transliterate the title and name of the author of the document: this skill is required for languages using an alphabet other than the Roman one, e.g. Greek, Russian, Bulgarian, Hebrew etc. It is easily acquired, but care should be exercized that a standard method of transliteration is used. Thus, for example, Recommendation R 7 by ISO (International Organization for Standardization) for the transliteration of Cyrillic characters should be used.

iii.) to translate the title of a document and headings, subheadings, captions to diagrams or illustrations etc.: when carried out with the aid of dictionaries this ability may often provide sufficient information to the user about the contents of the document so that he can then decide whether or not he requires a full-length translation of it.

iv.) to prepare a proper translation or an abstract of a document: at this level of competence it is absolutely essential for the translator to combine two different abilities, viz. a knowledge of the languages from which and into which he is translating as well as a real understanding of the subject matter of the document. It is normally recognized in professional linguistic circles that a translator should only translate into his mother tongue; similarly he should confine his activities to those subjects with which he is fully acquainted. It is unfair to the user and the translator if the latter is asked to deal with an article on, say, biochemistry if his expertise lies in metallurgy. It is obvious that the special terminology of the subject will be unfamiliar to him, probably in both languages unless he has made a special study of it.

In order to be efficient at any of these levels of proficiency it is however equally important to have available really reliable tools of the trade, viz. dictionaries. Although such works are usually classed as reference works and thereby given a status of infallibility, they only rarely deserve it. It must always be remembered that they are compiled by human beings who - alas - are frail and fallible. Hence the usefulness of these works which are compiled by such beings is sometimes adversely affected by inaccuracies, omissions, distortions, typographical errors and the like. Some compilers of dictionaries appear to rely on people with mainly linguistic qualifications and thereby overlook the importance of specialized subject knowledge. For example, the publishers of a well-known French-English dictionary use schoolmasters to check the entries; it is reasonable to assume that few of these teachers have ever been inside a foundry and would therefore be utterly unaware of the fact that the French word 'fonte' has at least three different meanings in foundry technology alone. As a result the translation of that word into English - when based on the definitions in that dictionary - may be quite quaint and even inaccurate. Perhaps Alice's friend Humpty Dumpty in that immortal work 'Through the Looking-Glass' is the patron saint of all dictionary compilers by virtue of his motto "When I use a word it means just what I choose it to mean - neither more nor less". It is often an amusing little exercise to compare the translations of a fairly simple technical term in several dictionaries and to find the many different renderings. It should be noted that only rarely does any dictionary give examples of context to illustrate various shades of meaning. Many others consider it sufficient to provide a few strange symbols such as intertwined anchors, for example, to denote the whole wide world of marine terminology.

There are of course some good dictionaries such as Harrap's French-English and English-French dictionaries, many more mediocre works and hundreds of poor dictionaries. Frequently their quality can only be assessed after long use by the critical translator. Some lists of dictionaries - which do not evaluate their usefulness - are:

Bibliography of interlingual scientific and technical dictionaries. 3rd edn, Paris, Unesco, 1953, 178 pp.
WALFORD, A.J. A guide to foreign language grammars and dictionaries. 2nd edn, London, Library Association, 1967, 240 pp.
The world's languages: grammars/dictionaries. 15th edn, New York, Stechert-Hafner Inc., 1969, 173 pp.

In linguistic work the role of the monolingual dictionary or glossary (e.g. the Concise Oxford Dictionary, Le petit Larousse or a British standard specification such as the one on terms used in the aluminium industry -BS 3660:1963) should not be underestimated since a correct detailed definition of words and concepts in one language can often be much more helpful than a guessed attempt at translating one word in a foreign language into one word of another language. For that reason many standaridizing organizationa publish such glossaries in order to normalize terminological usage. Other useful tools in the armoury of a professional translator are the illustrated dictionaries of which the German Duden series provides an outstandingly good example. By showing a picture of the article whose definition is translated it proves the truth of the old

Chinese proverb "One picture is worth ten thousand words".

It would be idle to pretend that translations are inexpensive. Because the production of a good translation requires the two above-mentioned skills of knowing the input and target languages as well as subject expertise, the exercise of these skills must be suitably remunerated. The cost of a translation therefore depends on several factors such as

a.) the languages concerned: this aspect follows the age-old pattern of supply and demand; thus while it is comparatively easy to find someone willing to translate from French into English, it is certainly much more difficult to find a translator able to convert a Polish text into Spanish. Therefore the latter task will be correspondingly more expensive.

b.) the type of text: obviously a simple business letter of the type "Dear Sir, unless you settle the invoice outstanding since....etc." requires a lower degree of skill than a learned text on the spectro-chemical analysis of heterogenous alloys of face-centred crystals.

c.) the length of the text: even if one uses a standard system of charging (such as 1000 words of original text) the question of length is still a salient point since a text of 100 words requires a considerable amount of clerical processing (typing, despatching, invoicing etc.) which is out of all proportion to the cost of the translation itself. Equally an advertising slogan of five words - say, 'Guiness is good for you' - requires a great deal of time, ingenuity and skill which is quite incommensurate with the charge for five words. In such cases a more equitable basis for calculating the cost is to allow for the time spent. This method is also applicable to the typing of mathematical equations, tables etc.

d.) the time required for completion: while a normal output of a competent translator is generally between 2000 and 5000 words per day - depending on the complexity of the task - extra efforts to complete a big translation in a short space of time must be rewarded by appropriate surcharges.

The multitude of languages mentioned at the beginning of this paper creates a considerable barrier to mutual understanding and numerous attempts have been made to overcome it by creating a universal language. Most of these endeavours have not been crowned with success, but reference must be made to the following:

LATIN: this was used by scientists (such as Bacon, Newton, Descartes) for a very long time as a universally understood medium, but its complex structure and the fact that it is spoken by only very few people have caused it to be used very rarely in modern times.

VOLAPÜK: the first artificial language was invented in 1880 by Johann M. Schleyer, a priest of Constance, Baden. Its vocabu-

lary is mainly based on English, though with a greatly simplified grammar. The word means 'world speech'.
ESPERANTO: the best known of the synthetic languages was devised by a Russian physician, Dr. L.L. Zamenhof, in 1887. Although it has more than 1250 local societies and national affiliates in 21 countries it is always a second language to its speakers. Its use in industry is very limited.
IDO: this language was developed from Esperanto in 1907 (its name is the Esperanto word for 'offspring'), but its growth has been rather slow and its influence is almost insignificant.
BASIC ENGLISH: this simplified form of English with a vocabulary of 800 words was introduced during the Second World War, mainly to help foreigners to learn English. Some scientific texts were produced in this language, but they were not considered to be very successful.

With the present preponderant position of English in the field of scientific and technological documentation it may be expected that English will become the major international means of communication in the future. Many of the developing countries have accepted this probability and their scientists now tend to publish their findings in English, perhaps to thereby gain a wider readership for their contributions.

A final word must be said about speeding up the process of translation which is time consuming and expensive. It is obvious that thoughts had been devoted to the idea of using computers to assist in this task. However, mechanical or machine translation sounded very promising in its early days and it seemed so easy to translate CAT (En) into its digital form 00011 00001 10100 which equalled CHAT (Fr) in its digital form 00011 01000 00001 10100. But the complexities of semantics, syntax, grammar etc. proved to be too great and today MT is considered to be moribund. There must be many texts whose very dullness and repetitiveness cry out for the use of a moron translator of which the computer is the prime example. And he can work 24 hours a day without a tea break. It is indeed regrettable that our hope that the machine will relieve the human translator of some of his drudgery seems to be in vain. But it must also be admitted that no computer will ever be able to produce a really good translation of Shakespeare's Sonnets or the Song of Songs.

Being a translator myself I must therefore declare my interest when saying 'DOWN WITH THE COMPUTER'. But in so doing I firmly believe that the human translator can be of real assistance in overcoming the formidable language barrier.

REFERENCES:

(1) Scientific and technical translating and other aspects of the language problem. Paris, Unesco,1957, p.15.
(2) Tybulewicz, A. and Liebesny, F. The relative importance of various languages in scientific and technical literature. Inc. Linguist, 4(1),Jan.,1965, pp.12-13.
(3) Tybulewicz, A. Languages used in physics papers. Physics Bull.,(120), Jan.,1969, pp.19-20.

TOWARDS THE DEVELOPMENT OF A CUMULATIVE DICTIONARY WITHIN A DYNAMIC INFORMATION SYSTEM

D.KRALLMANN

SYNOPSIS

It is reported about a flexible dictionary system for the german language, called cumulative dictionary (CD). The CD is based not on words as lexical units but on basic concepts, which are as minimal units elements of a semantic network. The CD is not constructed for special applications within the range of information systems, it is conceived to manipulate any language data in verbal communication processes.

Our work on the development of a cumulative dictionary (CD) as part of a dynamic information system is to be seen within the range of communicative aspects of verbal behavior. The re-interpretation of behavioristic stimulus-response schemes led to the design of special models of thinking and infering. With respect to this topic information systems can be interpreted as mechanistic attempts to explain the functioning of cognitive processes. With regard to realization, methods are used which, having the result in view, seem to be justified.

Consequently, information systems are not isomorphic copies of mental activities, but interpretable as models of verbal communication processes which are goal-oriented. In order to reach the goal linguistic methods are used mutually in analysis and synthesis in a manner which is controled by interactions of human users.

Within the area of information sciences following problem situation may be sketched. The user has an information lack,simultaneously he wants to fill his information gap. This effects that he carries out certain actions, e.g. he consults an information system (IS). The goal of these actions is to get data, which are interpretable as information.

Hereto several means of accessing and using an IS are in use ranging from simple searching procedures with fixed query languages to more flexible dialogue languages. These methods are oriented with human language knowledge. The user employs verbal means in communicative function for the description of facts of a case and/or situations and for the determination of certain operations on the basis of which analysis and search procedures are carried out so that the system itself is able to react in an informative manner.

The system which is based on a general communication model can be described in the following manner; the computer as information store and processor is not to be seen as an isolated entity, it is to conceive as component of communication system which is constituted of a computer C,a human user M and a goal-oriented cooperation K between M and C. M interacts with C in order to fill an information gap which prevents him from solving a given problem. To solve the problem is the goal of the cooperation between M and C. M communicates with C about the problem, about possible problem solving procedures and about lacking information.

Such a system is called a PIC-system, that means a problem solving, information retrieval, communication system.

Within such a system the computer has the function of an active communicator which has to fulfil a certain role in a verbal communication process.

In the development of IS essential parts of problem solving and answering of questions in document and fact retrieval systems are done by dictionaries, thesauri and lists of key-words, descriptors, and the like. These dictionaries very often must be considered as lists with a rigid structure and a limited size. A user of the system must know at first the list of key-words or descriptors which are pre-assigned by a himan cataloguer; in addition to that he must accept the rigid representation of the data and correspondingly a rigid formulation of the retrieval requests, so that the questions are able to be answered. This in turn makes necessary the existence of a special dictionary constructed for the range of application of that system.

In contrary to this we assume that within a dynamic system the vocabulary of the language used and the set of information terms should not be stated as constant and rigid list, but should be defined as "potential", as capability to perform retrieval requests as part of actual communication processes.

The cumulative dictionary system is to be seen as an attempt to build for the German language a flexible and variable form of dictionaries for interactive processes. The hypothesis of the CD is that the German language inventory - such as the vocabulary - is reducable not to words as basic elements but to semantically oriented terms which are actually smaller in size than the number of words. These semantically oriented terms, we call basic concepts; words and phrases are build up or generated according to the theme of the communication process or the intention of the user.

'Cumulative' means that the system is build up and enlarged user- and data-oriented, but only in an actual interactive process the system becomes a concrete dictionary. Up to now the CD is not provided for a special range of application, it is rather conceived as mechanism to manipulate any language data within the area of information systems.

As to word formation, the following hypotheses are formulated. Each word consists of at least one basic concept. This basic concept is defined as a word formative which determines the semantic conception of the word. Thus the basic concept of the German word 'BEIN' (leg) is identical with the word itself, whereas the words 'HALTBARKEIT' (stability) and 'ENTHALTSAMKEIT' (abstinence) have a formative 'HALT' as basic concept.

For the first approach basic concepts are comparable to what German philologists call 'root'. The difference can be demonstrated when comparing the words 'FLIESSEN', 'FLUSS', 'FLOSS', which have three roots or basic formatives, but one basic concept.

As a function of the basic concept, further classes of formatives are defined. Formatives in front of a basic concept are called LFs (left formatives), formatives which follow a basic concept are called RFs (right formatives). For a word material of about 68.ooo German dictionary words, an analysis of the formative frequencies results in the following distribution:

words of one formative	2.781
words of two formatives	2o.412
words of three formatives	27.777
words of four formatives	12.858
words of five formatives	3.563
words of six formatives	98o
words of seven formatives	137

Each formative gets one or more classification features according to its position and function in the word. An analysis of 68.ooo dictionary entries resulted in a set of 4.5oo basic concepts, 2oo RFs or/and LFs and about 1.3oo formatives of non-inflectable words, such as prepositions and interjections.

The CD-system consists of three parts: an analysis part, a verb inflection part and a synthesis part.

As a result of the analysis formatives as minimal units of the CD are identified and classified. This yields to a general word structure scheme which can be written as

$$\left[(\mathrm{NE})/\ \{\mathrm{LF}\}_i^{*}\ \ \mathrm{BC}\ \ \{\mathrm{RF}\}_j^{*}\ /(\mathrm{NE})\right]$$

Here NE stands for 'no entry in this position of the word', i and j are position indicators, * indicates an iteration, and LF, BC and RF are left formatives, basic concepts and right formatives, respectively.

Thus the word 'EINHEITLICHKEIT' (uniformity) has no left formatives, one BC and three formatives as RFs, whereas the word 'UNVORSICHTIG-KEIT' (lack of foresight) has again one BC and two LFs and RFs.

All formatives are classified in relation to the corresponding basic concepts. The word formation operates on an actual basic concept or on a basic concept position.

In the first case a BC is activated by an user who regulates the word formation by a selection of BCs and of combination rules for word formatives. In the second case the user is able to select any formative in order to define it as BC within a special word. That means, the word generation operates on a word structure, chosen by the user; a BC is defined as the formative which occurs in BC-position. Thus the user is able to generate even words like 'ing-ness', or 'ibility-ness'.

The function of the verb inflection part is a double one. During the analysis it is used to perform a reduction of basic formatives to BCs. Thus the basic formatives 'FLIESS', 'FLUSS', 'FLOSS' are normalized to one BC 'FLIESS'.

On the other hand all verb forms are analysed and synthesized within a communication process; also, different basic formatives are synthesized from one basic concept.

The formatives as minimal units of the CD are classified according to different points of view. BCs are interrelated as elements of a semantic network. These relations are determined partly from word combinations, partly from special contexts. Relations between words are not stated as fixed entries in the dictionary, they are generated from BC-relations.

The dynamic aspect of the system is given by the reduction of words to semantical-oriented minimal units. New combinations and concatenations are won from language material, like texts or dictionaries, and from concrete interaction processes within a general communication model.

THE MANY USES AND FORMS OF
SUBJECT REPRESENTATION

B. C. Vickery
Head of Research, ASLIB

SYNOPSIS

There are many kinds of subject representations within bibliographic records - titles, abstracts, descriptors, codes and so on. Their variety is discussed, and how their structures are related to their differing functions.

Information retrieval today is usually a process that takes four steps: (1) we search for words, often the keywords by which documents are indexed; (2) using these words, we search an index for references; (3) armed with these, we locate documents in a store; (4) from these documents we extract information. The first step uses word lists of various kinds - dictionaries, thesauri, classifications, code books. The second uses indexes, catalogues, and many other types of reference retrieval system. In such systems, the full texts of documents are not usually present. Instead, there are document representations such as catalogue entries or machine-readable records. It is such representations that I wish to discuss.

A document representation or bibliographic record consists of a set of elements such as author, title, publisher, keywords and so on. The elements can serve two main purposes: collectively, they describe and identify the document represented; individually, each element can in principle serve as a retrieval key by which records that share a common characteristic can be selected. Descriptive cataloguing is a well-established discipline that will not be considered here. I want to discuss the retrieval functions of document representations.

The traditionally accepted retrieval functions of a catalogue were set out by Cutter in his classic Rules (ref.1) :

"1. To enable a person to find a book of which either the author, the title or the subject is known.
2. To show what a library has by a given author, on a given subject, or in a given kind of literature.
3. To assist in the choice of a book as to its edition or as to its character."

The traditional means of serving functions 1. and 2. was the provision of catalogues containing single-purpose records (author, title, subject, form and language entries), each entry carrying only a single retrieval key by which it was filed and selected in search. Function 3. was served by elements in the descriptive part of the record (imprint and notes).

III Information Analysis

In the century since Cutter first wrote his cataloguing rules, the need for more selective retrieval has steadily increased. First, indexes to articles and reports now loom much larger than catalogues to books, and this has vastly increased the number of document records from which selection has to be made. Second, the output of these articles and reports has greatly grown - there are perhaps a hundred times as many in our libraries as there were in Cutter's day. As a result, a greater number of bibliographic elements are called upon to serve as retrieval keys. More importantly, each document is now indexed from a number of different subject aspects. Instead of the 4 or 5 retrieval keys with which Cutter was concerned (author, title, subject, form or language), we may now wish to retrieve a document via 10, 20 or more keys.

To provide 10 or 20 records in a conventional catalogue has usually been considered impractical. With the coming of machine-readable files we now have the possibility of providing a single multi-purpose record that can carry many retrieval keys by which it may be selected. One may cite as an example the bibliographic records of the U.S. National Aeronautics and Space Administration (ref.2), each of which can be searched via a dozen subject terms, author, report series, corporate source, broad subject category, document type, security classification or accession number.

Moreover, each bibliographic record may perform several functions. The NASA record, for example, is used in a printed abstracts journal, in a service for selective dissemination, in retrospective search, and in on-line consultation of the file. Each function may impose its own requirements on the structure of the record. In the remainder of this paper I am going to concentrate on subject representations within bibliographic records, and consider how their functions and structures are related.

The figure below is a typical bibliographic record. It is in fact taken from U. S. Government Research and Development Reports, 1967.

6T. Toxicology

PROBLEMS IN AERIAL APPLICATION: DETECTION OF MILD POISONING BY ORGANOPHOSPHORUS PESTICIDES USING AN AUTOMATED METHOD FOR CHOLINESTERASE ACTIVITY.
Civil Aeromedical Inst Oklahoma City Okla
Patsy R. Fowler, and Jess M. McKenzie. Apr 67.
15p
FAA-AM-67-5

Descriptors: (*Cholinesterase, Chemical analysis), Blood chemistry, Tissues (Biology), Toxicity, Pesticides, Insecticides, Automatic, Organic phosphorus compounds, Carbamic acids, Poisoning.

An automated method, capable of measuring cholinesterase activity in blood and tissue samples, was modified to provide increased reliability. The technique was evaluated as a means of detecting and measuring the inhibition of enzyme by organophosphorus and carbamate insecticides. As many as 200 specimens a day may be analyzed by the improved method, which provides precise estimates of cholinesterase activity in normal and poisoned samples. (Author)
AD-656 211 HC$3.00 MF$0.65

Figure 1.

Four areas of the record involve subject representation. At the top is a class code, 6T, which is in fact a COSATI subject category (ref.3). The word that follows it, Toxicology, is the natural-language equivalent of the code. Next there comes the subject of the document as expressed by its title. There follows a set of subject index terms, descriptors, drawn from a controlled thesaurus. Lastly, there is an abstract, provided in this example by the author.

As previously mentioned, all the elements of the record serve a descriptive function, allowing the reader to form a fairly clear view as to the content and likely interest of the document. The subject keys can in addition serve a variety of retrieval functions, such as :

(a) Topical search - e.g., to recall the item in response to a specific request for documents on "chemical analysis of cholinesterase in blood" or "enzyme inhibition by carbamic compounds".

(b) Generic search - e.g., for items on "toxicity of pesticides".

(c) Selective dissemination - e.g., to produce the item in response to a profile that includes "Physiological effects of phosphorus compounds".

(d) Group dissemination - e.g., inclusion of the item in a Toxicology bulletin.

(e) Subject indexing - e.g., to provide an index entry for visual search, such as "organo-phosphorus poisoning: detection: by cholinesterase activity".

(f) System transfer - e.g., to enable the item to be switched from class 6T in one system to class 615.9 in another.

All these are legitimate and necessary functions, found separately or together in information systems. Their variety is such as to make it doubtful whether any single subject representation (say, the title of the document) could serve them all. Let us consider each of them in turn and how they are handled.

Topical or specific search.

To include in a document representation an indication of the specific topics dealt with by the document, permits the item to be recalled in a topical search. The specific topics may be represented in natural language (the title and/or abstract of the item), or in a controlled indexing language (a string of assigned index terms or descriptors that may be in some ordered sequence), or in some combination of both.

Depth of indexing (i.e. the degree of content detail that is represented) usually increases along the series title/assigned index entry/abstract. In the example given above, only the abstract specifically mentions "enzyme inhibition", and the title does not include the concepts "blood" or "carbamate". Consequently, title search for either "chemical analysis of cholinesterase in blood" or "enzyme inhibition by carbamic compounds" would fail to retrieve the item, and the latter topic would produce a recall only with abstracts search.

Apart from depth of indexing, what differences are there between natural and assigned index languages, as shown in the above example? Some words are identical (cholinesterase, blood, etc.); for some the assigned language gives a standardization of word form (analyzed - analysis, automated - automatic, organophosphorus - organic phosphorus); some assigned terms are specified more closely by scope notes, e.g. tissues (biology); in one case, a near synonym is assigned (toxicity, to supplement poisoning). The effect of these differences is that higher recall and precision can be achieved with a simpler search formulation in assigned language.

<u>Generic search and selective dissemination.</u>

These two types of search are usually aiming at a wider subject than topical search. To retrieve relevant items, it is necessary <u>either</u> to sum specific terms, <u>or</u> to include in the assigning more general terms.

With natural language systems, summation is usually essential. Where the summed terms simply differ in spelling, they can be specified, in mechanically searched systems, by truncation. Thus, *CELL* will recall Cellulose, Cellophane, Celluloid, Hemicellulose and other similar materials. Truncation is a tricky technique, for *CELL* also recalls the unwanted Cells, Cello, Cancellation and Miscellaneous. Summation may also be carried out by searching for Cellulose OR Cellophane OR Celluloid, etc. An example using both OR logic and truncation is

FUNG
SPOR*
MYCE
YEAST*
MOULD*
BLIGHT*
MILDEW*
MICROORG*
MYCOS
*CILLIUM

Assigned index languages can provide for generic search by various forms of generic tagging. One form is generic posting. Specific descriptors used in indexing automatically have added to them a series of more generic terms that can be used for search. The only instance of this in the example given above was Poisoning <u>add</u> Toxicity. In the Euratom system (ref.4), generic posting is much more extensive: e.g., we have Sprouting <u>add</u> Plant growth, Plant growth <u>add</u> Plants, Plants <u>add</u> All Agriculture. As a result, a generic search for All Agriculture will retrieve Cattle, Cereals, Eggs, Fruit, Plants, etc., and all their more specific index terms. A second form of provision for generic search is to use hierarchical coding, as in the American Institute of Physics system below (ref.5):

.8 **<u>BIOPHYSICS</u>**

.81 PHYSICAL PROPERTIES OF BIOLOGICAL ENTITIES (for electrical and magnetic properties of biological entities combine with .23 section)
(for optical properties combine with .24 section)
(for chemical reactions in biological entities combine with .49 section)
(for thermal and thermodynamic properties of biological entities combine with .51 section)
(for mechanical and rheological properties of biological entities, including acoustical properties, combine with .52 section)

.811	Mass, density
.812	Size, shape, structure
.813	Chemical Composition
.82	SENSORY PHENOMENA
.821	Vision, visual perception
.8211	Visual threshold
.8212	Color perception
.8213	Depth perception
.8214	Defective vision
.822	Hearing, auditory perception
.8221	Auditory threshold
.8222	Pitch perception
.8223	Binaural hearing
.8224	Defective hearing

The Indexer assigns one or more specific codes to each item, e.g. 8214 Defective vision. The item will be selected by a search at any level of the hierarchy (such as 821 Vision, or 8 Biophysics).

Figure 2.

The essence of this provision for generic search in assigned index languages is therefore to draw descriptors from a thesaurus (alphabetical or hierarchical) that links specific and more general terms.

Group dissemination.

For the generic search purposes just discussed, each and every subject term in a representation may give rise to generic terms, so that the item can be retrieved in many generic searches. For group dissemination, an item may need to be assigned to only one or two groups. In the example given above, the abstract is assigned to group 6T, Toxicology, and the representation must include this key. This form of tagging might be used in the production of a series of group bulletins, such as the NASA/SCAN notifications (ref.6).

It can also be achieved by tagging the one or two most important specific descriptors assigned to an item, which give rise via a thesaurus to added generic group keys. For example, in a Nuclear thesaurus, terms might be linked as below to lead up to groups 14 (Nuclear energy) and 19 (Fuel and power) via sub-groups such as 14/2 (Nuclear power).

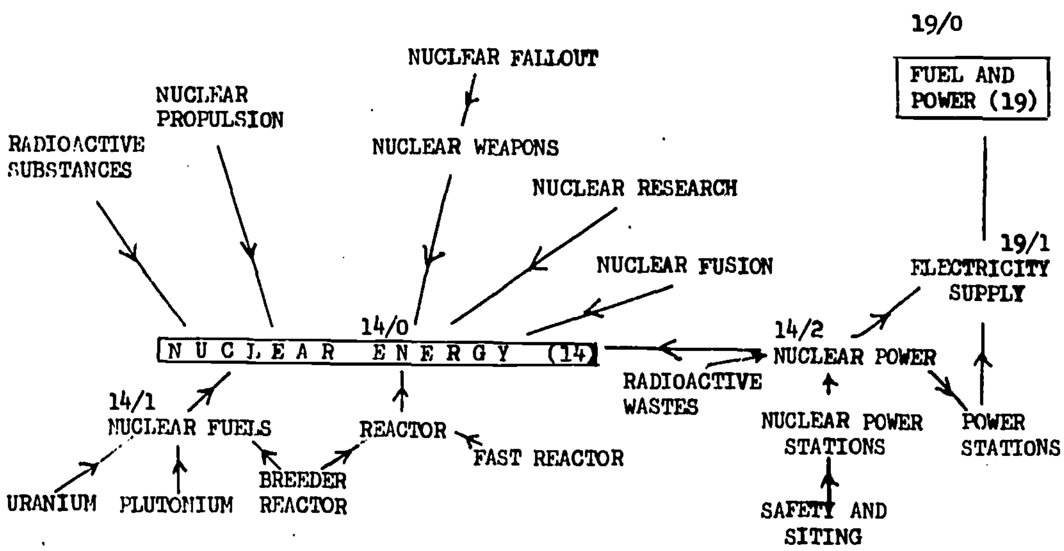


Figure 3.

Visually searched subject indexes.

Subject indexes for visual search remain very important retrieval tools. Techniques such as summation and truncation are not suitable for visual search. Faced with a set of subject terms - such as either the descriptors or the abstract in the example we have been using - the visual searcher needs two things: (a) that the set of terms be broken up into a series of intelligible index phrases, and (b) that each phrase be accessible from each significant term within it.

Where the subject representation of a document is its title, this can serve as an intelligible index phrase. The well-known KWIC routine then serves to rotate the title to make it accessible from each significant term. The latter are selected either by "stopping" common words, or by tagging words considered to be significant. The same procedure can, though less effectively, be applied to abstracts. Still more "intelligibility" can be built into the resulting index by such techniques as "double-KWIC" (ref.7).

Where the subject representation consists of terms selected from or assigned to the document, intelligible index phrases can be generated by (i) linking subsets of terms into "themes" or "splits", several to a document, (ii) ordering the terms within a theme into some relational pattern, and (iii) providing a rotation program that will produce an index entry for each significant term within the theme. Recent examples of such procedures are Skolnik's multiterm index (ref.8), the articulated subject index of Armitage and Lynch (ref.9), and the preserved context index system of Austin (ref.10). Such techniques may be based upon assigning index terms to categories, that are related in a standardised way. For example, Skolnik designates each term as either a chemical product (P), reactant (R), process (S), condition (C), equipment (E), use (U), or property (Y), and formulates sequences such as PRSC, that may be rotated to give additional index entries CPRS, SCPR and RSCP.

System transfer.

The use in one retrieval system of bibliographic records prepared in another is fairly common: often, specialist information services utilise published abstracts. The title of an item can clearly be transferred without modification from one system to another. An abstract can be similarly transferred, provided that its slant is acceptable to the receiving system.

Records can be transferred en bloc if a coded group, such as 6T Toxicology, is used in both the systems concerned. The COSATI subject categories (ref.3) were in fact developed in the hope that they would be widely used to group items in bibliographies. The recent UNISIST report on the feasibility of a world science information system (ref.11) has stressed the need for a set of standard subject groups for international use.

A much more difficult problem is that of relating the different controlled indexing languages used in retrieval systems. To do this it is necessary to establish concordances, either directly between individual languages or via an "intermediate lexicon". As an example of the direct concordance we have the very careful analysis and linkage of the Euratom theesaurus and the UDC, undertaken by Marosi at the Soreg Nuclear Research Centre (ref.12).

The idea of the "intermediate lexicon" is that one particular controlled language (and the UDC is often advocated) should serve as an intermediate standard. Each other specialist indexing language should develop a concordance with this standard. Each specialist system would include in its bibliographic records, not only their own set of descriptors, but also concordant terms from the intermediate lexicon. A second specialist system would be able to go from these intermediates, via its own concordance, to its own descriptors (ref.13).

Review of subject devices.

It is very evident that it is the variety of functions that are served by subject representations that has given rise to the variety of structures that have been mentioned, We have seen in turn the use of

- natural language words in title or abstract
- tagging of these words as "significant", either positively, or negatively by means of a stop list
- truncation of these words to permit generic search
- standardised, controlled indexing terms to improve search performance
- generic relations between terms, expressed either in posting or coding
- group terms, either separately assigned or related to standard terms
- an inter-system standard set of such group terms

- linkage of natural language words or controlled terms into themes
- assignation of terms to categories to permit the formation of structured, intelligible index phrases
- rotation techniques to provide multiple access to entries
- an intermediate lexicon for supplementary "concordant" indexing

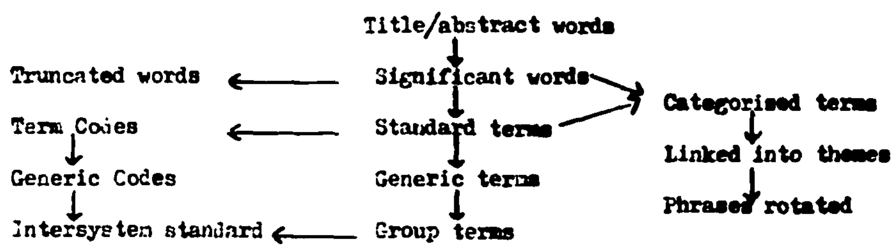


Figure 4.

Once it is clear that structural variety is only a reflection of functional variety, then we no longer have to argue as to which structure is "the best". Arguments about the alphabet versus classification, or about natural versus controlled languages, are as though carpenters were arguing as to which tool could be used for all jobs - the saw, the chisel, the hammer, the plane, or whatever. Even tests to evaluate which device is "best" are of restricted applicability, for they assume that all structures aim to serve the same function.

The moral that I draw from this review is that the many forms of subject representation now found may be justified as serving many uses. It is likely, of course, that not all variety has arisen truly in response to needs, and that some exists only because of hallowed tradition or fashionable innovation. What we need is to relate structure carefully to the function it serves, and to consider its effect not only on performance, but also on system costs. The purposes served by information systems are multiple and complex. We cannot expect these purposes to be achieved by simple bibliographic devices.

References.

(1) Cutter, C.A., "Rules for a dictionary catalog", 4th ed., US Bureau of Education, 1904.
(2) Brandenharst, W.T., and P.F. Eckert "Guide to the .... NASA Scientific and Technical Information Facility," Documentation Inc., 1966.
(3) "COSATI subject category list", U.S. Federal Council for Science and Technology, 1964.
(4) European Atomic Energy Community "Euratom thesaurus", 2nd ed., 1967.
(5) "Draft classification scheme for physics", American Institute of Physics, 1968.
(6) Wente, V.A., and G.A. Young "Operating experience with NASA/SCAN", Proceedings of American Society for Information Science, Annual Meeting 1968, vol. 5, pp. 217-22.
(7) Petrarca, A.E., and W.M. Lay "The double KWIC coordinate index", Proceedings of American Society for Information Science, Annual Meeting 1969, vol. 6, pp. 277-82.
(8) Skolnik, H., "The multiterm index", J. Chem. Docum., vol. 10, no.2, May 1970, pp. 81-4.
(9) Armitage, J., M. Lynch and others "Experimental use of a program for computer-aided subject index production", Inf. Stor. Retr., Vol.1, 1970, pp. 79-88.
(10) Austin, D., "An information retrieval language for MARC", Aslib Proc., Vol. 22, 1970, pp. 481-91.
(11) UNISIST: Study report on the feasibility of a World Science Information System, Unesco, 1971.
(12) Marosi, A., "Euratom thesaurus and UDC", J. Docum., Vol. 25, 1969, pp. 197-213.
(13) Coates, E.J., "Switching languages for indexing", J. Docum., Vol. 26, 1970, pp. 102-10.
(14) Vickery, B.C., "Structure and function in retrieval languages", J. Docum., Vol. 27, 1971, pp. 69-82.
(15) Vickery, B.C., "Document description and representation", Annual Review of Information Science and Technology, 1971, in the press.

AN INTEGRATED CLASSIFICATION AND INDEXING SCHEME FOR PHYSICS, ELECTROTECHNOLOGY, COMPUTERS AND CONTROL

B J Field
INSPEC, Institution of Electrical Engineers, London

SYNOPSIS

The INSPEC unified classification and indexing schemes are described together with the reasons for their adoption. The unified indexing scheme which has a thesaurus structure and is linked to the classification scheme has been designed to be used both as a controlled - language and as a free-language thesaurus.

The Institution of Electrical Engineers has been concerned with the classification and indexing of scientific and technical information since 1898 when it began to publish Science Abstracts. With the establishment of INSPEC (Information Service in Physics, Electrotechnology, Computers and Control) and the provision of computer-based services, such as magnetic tapes and SDI, in addition to the production of the abstracts journals and current-awareness publications by computer, the development of improved classification and indexing schemes has been greatly stimulated.

It is these schemes which form the subject of this paper, but, first, to place the developments in context, the present classification and indexing system will be briefly described.

Background

Up to the present time INSPEC has used three separate and independent classification and indexing schemes for the three INSPEC abstracts journals (the three sections of Science Abstracts: A - Physics Abstracts, B - Electrical and Electronics Abstracts, C - Computer and Control Abstracts). These schemes are used to arrange the abstracts and titles in the abstracts and current-awareness journals and to produce the six-monthly printed subject indexes to the abstracts journals.

III Information Analysis

The classification schemes for each of the three sections are hierarchical, with three levels of hierarchy and each contain about two hundred headings, which are allocated a four-digit classification number. The subject indexes may be described as alphabetic systematic schemes with a structure of 'see' references and 'see also' cross-references and a subheading structure in some cases. The index headings are controlled and supplemented by free-language modifier lines which add more detail and context to the headings: these are arranged alphabetically under each heading and give a reference to the abstract number.

An abstract is selected to appear in one or more of the sections A,B,C and is classified and indexed for the appropriate sections independently. The complete indexing operation consists of assigning classification codes, controlled subject headings and modifier lines for the printed index, and free indexing terms which are the main tool for SDI and retrospective searching in the INSPEC system.

To simplify the indexing operation the unified classification and indexing scheme is being developed to replace the sectional schemes, with four main points in mind:

1. <u>The cost effectiveness of the intellectual effort in indexing</u>

As described above, the use of three separate schemes requires extra intellectual effort. In using a unified classification and indexing scheme the information scientist will only concern himself with the subject delineation of a particular document and not with its selection for particular sections. Therefore, he will carry out one indexing and classification operation and the document subsets and the sectional classification and indexing schemes will be generated by program.

2. <u>Suitability for use in all INSPEC products</u>

These products include the abstracts journals, current-awareness journals, printed indexes, SDI, Topics and magnetic tape services. The unified classification and the sectional classifications generated from it will be used to display abstracts in the monthly and fortnightly abstracts journals and to display titles in the current awareness journals. It will also be present in the magnetic tape, SDI and Topics services where it will provide a means for making broad subject searches.

The unified subject indexing scheme, which will be linked to the classification and have a thesaurus structure of BT, NT, RT etc., fulfils a number of functions. As a controlled-language thesaurus it will be used to provide an authorised list of subject headings (and corresponding lead-in terms) for the printed indexes to the INSPEC abstracts journals, and may be used by subscribers to the INSPEC tape services who wish to carry out their own SDI or retrospective searching using a controlled language. As a free-language thesaurus it will be used by INSPEC to organise and relate the free-indexing terms which are the main tool for SDI and retrospective searching in the INSPEC system.

3. Results of INSPEC index language studies

The results of the DEVIL project (Direct Evaluation of Index Languages) which will be discussed in detail in another paper to be given by Mr. T.M. Aitchison has led to the decision that INSPEC would use free indexing together with a free-language thesaurus for SDI and retrospective searching of the machine file. Briefly, although in the index language studies the controlled language showed a superior performance to the free language it was thought that the free indexing could approach the performance of the controlled language if a free-language thesaurus were developed so that most of the advantages of the controlled language were obtained in retrieval.

4. International co-operation, and the acceptability of the scheme to the user community

As producer of one of the largest scientific data bases in the English language, INSPEC accepts its responsibility to the international network of information services, and has sought and is seeking to cooperate with organisations throughout the world in the development of its classification and indexing system.

The new schemes were formulated in connection with proposed cooperative developments with the Institute of Electrical and Electronics Engineers Inc. of New York and the American Institute of Physics. They take into account the best current practices of other operating systems, in particular Bulletin Signalétique, Physikalische Berichte, and Referativnyi Zhurnal, and the results of the comparative study of classification schemes by INSPEC on behalf of the Working Group on Physics of ICSU AB. The Working Group is presently examining the INSPEC scheme with a view to its adoption as a common basis for development.

In addition to this continuous dialogue with other information organisations, INSPEC has established an International Advisory Panel network to ensure that it continues to keep in touch with the changing needs of users.

Having described the background and philosophy behind the INSPEC classification and indexing policy, the scheme will now be discussed in more detail.

Classification scheme

The scheme is hierarchical in structure with a maximum of four levels of depth at present and each category is understood in the context of its hierarchy. Each category is assigned a six-letter code, the fifth being for future expansion and the sixth will be used as a machine check character. All levels of hierarchy can be used in classification, with multiple choices if required, and 'see' references are used when it is thought that a particular category could fall logically into more than one hierarchy. The sectional classification schemes required by each journal are generated by program using a mapping procedure from the unified scheme to the separate schemes (as shown in figure 1) which in turn predetermines their subject content. Thus in some subject fields a 1:1 overlap for different sections will be allowed for, whereas in other fields which do not merit a 1:1 overlap the hierarchy is designed to separate papers aimed at different sections of the user community. In this way we have a rigid automatic selection procedure in place of the present selection of documents for particular sections which relies upon the decisions of a number of information scientists, with all the variations and inconsistencies that this gives rise to.

		Sectional Classifications	
SAAAA	SOLID-STATE PHYSICS AND MATERIALS SCIENCE	A16.00	
SMAAA	ELECTRICAL PROPERTIES AND DEVICES	A17.20	B11.00
SMFAA	Superconducting properties, materials and devices	A17.24	B13.90
SMFBA	General theory of superconducting state	A17.24	
SMFDA	Superconducting properties and materials	A17.24	
SMFFA	Superconducting devices	B13.90	

(for superconducting stores use ZHKAA)

Figure 1
Unified classification and its mapping to the 1971 sectional classifications

The thesaurus is linked to the classification scheme in that each thesaurus term has been allocated one or more classification codes. The terms are not locked into the classification scheme but each term has a finite probability of existing in the subject space defined by a particular classification category. In other words each classification category has a cluster of terms associated with it which are likely to be of interest in that subject field and perhaps likely to be assigned by indexers. It would, of course, be possible to carry this link a stage further and implant the thesaurus terms into the classification structure, resulting in a more detailed and less flexible system, but at the moment it is considered that there are real advantages in having the two schemes essentially independent.

Unified Subject Indexing scheme

The complete indexing procedure consists of assigning classification codes, controlled subject headings for the printed indexes and free-index terms. It is these last two indexing requirements which have mainly influenced the design of the unified subject indexing scheme together with the specification for the machine file and its validation and special print-out programs. The unified indexing scheme has a thesaurus structure of BT, NT, RT and the terms are of direct entry type (i.e. the words are in a natural language order) and comprise uniterms, uniconcepts and precoordinated subject headings, the average number of words/term being about two, which gives some idea of the level of precoordination.

The basic relationships are shown in figure 2 and it can be seen that the synthesis of multiword terms or phrases (which are likely to be assigned in free indexing) is allowed for by using the relationship G <u>use</u> (H <u>and</u> I <u>and</u> J)

	reciprocal relationships
A NT B	B BT A
C RT D	D RT C
E use F	F UF E
G use (H and I and J)	H (with I and J) UF G
	I (with H and J) UF G
	J (with H and I) UF G

M - modifier to distinguish homonyms

Figure 2
Basic relationships in the thesaurus

III Information Analysis

As far as the printed index is concerned the requirements of the thesaurus are threefold as follows:

1. The specificity and number of subject headings which are available should result in a reasonable number of entries under each subject heading (i.e. about ten after a six months cumulation).

2. The thesaurus structure and its relationships should provide a sufficient number of lead-ins ('see' references) and cross-references ('see also' references): this will depend largely on the specificity of the subject headings themselves.

3. The question of adding further detail and context to the subject headings will to some extent depend on the specificity of the subject headings themselves which in turn determines the need for an alphabetical or other arrangement of 'context lines' beneath each heading.

To satisfy these requirements the main types of term in the vocabulary development file which is being used for developing the classification and indexing schemes are:

(a) a main entry in the thesaurus

(b) a 'lead-in' in the thesaurus

(c) a direct entry in the printed index

(d) an indirect entry in the printed index

(e) a classification heading

It can be seen that, in general, (c) is a subset of (a) and (d) is a subset of (b) although some of (a) will become (d).

The printed index structure results from a choice of the most important relationships in the thesaurus, which will depend largely on their frequency of occurrence in documents. In general NT/BT and RT relationships (particularly object-property and part-whole relationships) are chosen for 'see also' references. In addition, multiword terms are rotated to create signposts for the printed index (e.g. conductivity, electrical <u>see</u> electrical conductivity).

The source of terms for the thesaurus has been the present INSPEC printed index terms, the term analysis of free indexing assigned by the INSPEC information scientists, term analysis in document titles (i.e. singlets and doublets), various other thesauri, particularly the English Electric Thesaurofacet, TEST, INIS thesaurus and scientific and technical dictionaries.

The thesaurus which was manually prepared on 8 x 5" cards containing the minimum basic elements was compiled both alphabetically and systematically, using clusters of terms which contained a common word or fell into the same classification category. These elements were keyboarded and processed in batch mode, using the vocabulary development file mentioned above, and the various file processing operations (e.g. the generation of reciprocal relationships and hierarchical chains) removed much of the mechanical effort and showed up many of the illogicalities and errors which had developed. It is intended that further updating and development will be done on a time-sharing system.

Future development

One possible line of development is the use of the free indexing input to generate subject headings for the printed index and hence classification codes, which brings us some of the way towards automatic indexing. Although this would be a saving in terms of intellectual effort, it would also mean that we would lose the flexibility of the three independent indexing operations.

The other major area of development concerns, of course, the most economic and efficient use that can be made of the INSPEC classification and indexing scheme in searching the machine file, including techniques such as on-line searching and automatic profile expansion.

1. Aitchison, T.M., Hall, A.M., Lavelle, K.H., and Tracy, J.M. Comparative evaluation of index languages, Part II: Results. Report No.R.70/2. London, INSPEC, The Institution of Electrical Engineers, July 1970.

THE STOCHASTIC STRUCTURE OF RESEARCH IN THE SOCIAL SCIENCES AS A DATA BASE

Richard Hellman

Professor of Economics and Director
Research Center in Business and Economics
University of Rhode Island, USA
Consultant, IBM Co. USA.

SYNOPSIS

The author represents the primary researcher who manufactures brand new data and concepts for the information Banker and draws upon him for those deposited by others. The researcher's problem of quickly identifying the tiny fraction of newness in the banker's depositary, and the cognate problem of a nonsystematic research anatomy are defined. Solutions are proposed for the banker.

This conference, as one may see from the subjects of the speakers, is primarily from the viewpoint of cataloguers, keepers and purveyors of man's world of knowledge -- the memory Bankers. But I speak here as a manufacturer of the data which goes into your bank, and as a grateful user of new informatinon manufactured and deposited with you by others. My viewpoint may be of interest and even of some value to you in your deliberations here on how to improve your services as Banker.

One more thing by way of background for what I shall say -- I speak from 35 years of work as a general research economist with the U.S. government in Washington, where we are continuously challenged to find solutions to urgent problems at the interface of the present and the future.

Definition of research : implications for Banker

Research is the endeavor for invention of the first of its kind under conditions of uncertainty. The inventory of already existing information is not pure research as I have defined it, although much of what is ordinarily called research is really simply compiling of inventory.

Let us now see what implications for the Banker flow from this simple definition. He serves two kinds of users: the passive receptor who simply wants to draw upon the existing storehouse of knowledge without adding to it, e.g., the undergraduate writing a term paper; and the researcher, the voyager into terra nova , who needs to know primarily what in a book or article is new, what is being added to the inventory of knowledge published before his particular writing. I shall call this incremental knowledge, Delta-new or, for short, D-new.

D-new is the essence of the problem of the research man in surveying literature, and it is in parallel the crux of the problem of the Banker in serving him. In my experience, perhaps 1/10 of a percent of the volume of new writings have something truly new in them in the research sense. When I pick up a book or article, what I look for is the truly new. Too often, there is little or nothing worthwhile of D-new. Once in a while, one comes across a gem, a repository of much D-new, and it is truly gratifying. The problem, as you know by your lack of library shelves or by the multiplication of your memory discs, is accelerating. Population is exploding at a rate "r", writers at 1.2r, publications at 1.3r and the selectivity or D-new problem at 2r.

How does all this affect you as guardians of the knowledge bank? As a customer, I would urge you to find a way to indicate on your catalogue "printout" that which is D-new. To the researcher, the man with one foot here and another in the future, this is vital. We now spend too much time running down spoors which take us nowhere. If a man is that rare breed, a good researcher, we as a society cannot afford to reduce his output and dissipate his scarce hours by leading him down these unnecessary blind alleys.

How can you do this? May I suggest what seems to me to be a simple solution ?

Compel, force each author to list on the equivalent of a library index card his D-new. The card catalogue would then contain two items: the usual topical index, and his D-new. You as Bankers would not be responsible for what the author lists as his D-new. That is for him to live with. But you can be responsible to force him to list his D-new.

To show you that I try to practice what I preach, permit me to quote from the opening sentences of a book which I just submitted to my publisher in June:

> "A preface is an easy chair from which, his labor done, the author may furnish cartographic notes to guide the voyager about to embark upon a reading of his work. Perhaps the best service to be performed is to answer the question that is uppermost in the professional reader's mind as he approaches a book : what in it is new? It is rarely that one is privileged tointroduce something entirely new. More often it is a matter of fresh treatment or emphasis. The more significant items of newness in these two senses, as seen by the author, are listed below."

The solution which I have proposed to you has real world practicality. If you could achieve it you, as Bankers, would add magnificently to the benefit/cost ratio of the world's researchers. You would bestow the following benefits:

1. You will force authors to list their D-new, and thereby may call to their attention its near or total absence.

2. This may not only result in reduced literary output, but what is left may be of some value.

3. Happily it may reduce your work loads.

4. For the researcher, it will reduce the task of extracting the D-new from an exponentially exploding library.

The stochastictic anatomy of research

The nonsystematic origination of research in the social sciences, and to some parallel extent in the natural sciences, is a complicating factor for both the researcher and the Banker. What an individual researcher undertakes to study is often the accident of personal selection, rather than a piece of a total vision, of a Weltanschauung . Or it is an instant reaction to an instant problem requiring an instant solution -- what may be called "firefighting".

Let us look at the categories in which research is generated.

1. In universities, a research subject is chosen by a particular interest of the individual professor, either for himself or for his degree candidate. There is no systematic framework of research values or priorities into which he fits his particular choice. The research may be good and relevant, but it is stochastic.

2. At nonprofit institutions of research, there may be greater sensitivity to priority of social problems, but again the subject chosen is dependent largely on the particular interest of those who have the say, and again it is essentially stochastic.

3. In the U.S. government -- a good example because the resources devoted to research are great -- "research" is largely a matter of firefighting. On two occasions I have had large staffs of professional researchers keep records of their time spent on firefighting and on long range economic research. The results, whether by coincidence or inherent pattern, in each case showed about 85% of time devoted to firefighting, about 5% on long range studies, and the balance for administration and miscellaneous purposes. Even the 5% was interrupted and diluted by firefighting to a very low scale of effectiveness. Looking at research in the social sciences throughout the government, I find very little true long range study, very few true researchers, and a stochastic pattern largely determined by nonsystematic forces.

A solution ? Is there a solution to the stochastic structure of research ? I think yes, to a worthwhile extent. Research goals are divisible into two categories: the unforeseeable, for which firefighting is inevitable; and the foreseeable, for which long range studies are suitable. Much more of the latter is practicable, and there is no reason why in the U.S. government, e.g., a Long Range Economic Studies Staff could not be housed in, say, the Executive Office of the President. Its mission would be to foresee the major problems upcoming in the next 5, 10, 20 and 50 years, assign priorities of study and see that they are studied somewhere. Success would anticipate and forestall the need for much firefighting. For years there has been complaint that government agencies "react" to rather than actively anticipate problems. Systematic planning of research would anticipate problems, permit anticipatory adaptation, and the allocation of resources to research according to a value structure expressed in priorities. At relatively little expense, a great payoff would be obtained. I should like to give two actual cases in my own experience.

About 15 years ago, at the Department of Commerce, it was foreseen that with the Common Market, changes in scale of production in favor of its countries would affect the competitive position of both the USA and countries within the market vis-a-vis each other. A case was chosen of refrigerators, where it was obvious that the Italians, having modernized their production lines, would submerge the French producers, displace American exports, and even export to the USA. It was possible 3 to 5 years before the dropping of the sword to foresee the situation and for the French and the Americans to decide either to renovate their own production lines or let the Italians take over. No studies were made, certainly in the American government, and probably not by the French because suddenly the sword did fall and they were pleading with the Common Market to give them time to adapt, but it was already too late.

My second case is pollution. The scandalous accumulation of poisons in our waters, our air, our earth were certainly present and known 25 and 50 years ago in the USA, and earlier in other countries. It is a neglect for which all are to blame, the experts and government officials more than others. Long range studies, publicized with skill and understanding, certainly would have placed us in a much better position than we are now mired in.

<u>Implications for researchers and Bankers</u>. For researchers, the non-systematic structure of research complicates the problem of searching for D-new in the vast output of government, the universities, institutional studies and unattached scholars. For Bankers, it also complicates the problem of identifying and cataloging in conveniently retrievable form the D-new.

I need not elaborate any more than I have already, except to say that here again the solution in my opinion is to persuade, cajole, compel the researcher to identify explicitly his original contributions to the human storehouse of data and concepts, and to move in the direction of a more systematized structure of research. There need not be a straightjacket, there need not be a stifling of creativity and imagination. The individual would still be free to work as he wished, certainly outside the government. But it would provide a frame of reference, a value structure expressed in a scale of priorities, for adherence or departure. Why not ? Surely it is worth thinking about.

APPROACHES TO ESTABLISHING DIMENSIONS AND CRITERIA FOR EVALUATION OF REVIEW PUBLICATIONS

Lawrence Papier
Program Officer, Research and Program Development, Division of Library Programs, Bureau of Libraries and Educational Technology, U.S. Office of Education, Washington, D.C. 20202

SYNOPSIS

Four independent approaches to the formulation of a taxonomy of review publications and the subsequent development of criteria useful to their planning and evaluation are presented. These approaches are based on the type and degree of intellectual processing applied, a transfer model, the purposes of users and producers, and their innate dimensions.

In recent years scholars from a number of disciplines have begun to urge the production of more and better review publications as a way of controlling an increased amount of data and information. These scholars recognize the utility of current efforts, particularly the use of those that involve computer-driven information retrieval systems, in supplying access to this information. They point out however, that the simple provision of efficient access and retrieval of pertinent documents does not solve the problem. Users of these systems are merely presented with a number of documents or citations, some containing the necessary information and some not. They must still carefully peruse these documents, recognize and extract the required information, and convert it to a form that can be absorbed and used. In some cases this effort taxes not only their patience but also their capability. An example of the latter is the case of the practitioner who is confronted with advanced mathematical formulae or theories or research data. True control and ultimate utilization can only be brought about through the processes of purposeful condensation and tailoring the information to the needs of particular audiences. To be useful, information must be of a form and quantity that can be absorbed by human beings.

By themselves, retrieval systems cannot accomplish these things. They are unable to condense, analyze, interpret, synthesize or evaluate information, and they are unable to convert information into usable knowledge. Such transformations are commonly accomplished, however, through the process of reviewing. Reviews therefore have a very basic, almost unique, role to play in the generation and in the ultimate utilization of knowledge.

III Information Analysis

One would think that in view of the importance of reviews, information scientists would be diligently studying their characteristics, developing a taxonomy, suggesting procedures for their preparation and evaluating and designing courses for training reviewers. Such is not the case. Reviews, reviewers, and reviewing have all been neglected by information scientists.

In a recent literature survey covering the past ten years this author found that less than three dozen pertinent papers on the subjects of either reviews or reviewing have been published in the English language. The list of these papers was circulated to a number of authorities who were able to add only three citations. Compare this to the several thousand papers on indexing that have been produced! The dearth of scholarly studies on this important subject has prompted the author to share some observations and thoughts on alternative ways to distinguish types of reviews, to define their dimensions and to make a start toward developing planning and evaluation criteria. These measures are necessary to assure quality, to provide guidance to those preparing them, to establish bases for resource allocation and, finally to structure the field in such a way that developed techniques can be taught to others. From a scholarly standpoint these measures can provide a start in an area rich in opportunity.

Three possible approaches to the formulation of a taxonomy of review publications and the possibility of subsequent development of criteria for design and evaluation will be considered: (1) the degree of kind of intellectual effort necessary for their preparation, (2) the purposes of their producers and consumers, (3) their inate dimensions, that is, their substance, characteristics and intellectual level. While all three approaches have their uses the last seems most promising for those interested in further scholarly development.

The intent of this paper is to break ground not by providing a single unifying theory but by offering alternative conceptualizations. The approaches presented are general and are not intended to apply to any one specific subject field. Even though reviews prepared for administrators are a relative rarity in science and technology, reviews for administrators are included since they are fairly common in education. Finally in keeping with the purpose of this paper, reviews are broadly defined. A review is considered to be a narrative presentation or reformulation of existing information on a topic derived from a variety of printed sources.

<u>Intellectual Processing</u>

Discussions of reviews often center about a somewhat simplistic polarization. Typically, in such discussions, mention will be made of the need for expert or critical reviews in a subject field of common interest and concern will be expressed about the unwillingness of experts to prepare them.

The point may then be made that while there are plenty of reviews of the non-critical type published, particularly the bibliographic, this is not what is needed. Predictably, interest will then turn to the "bread and butter" issue of how experts can be remunerated or otherwise encouraged to do the necessary work.

The need for critical reviews is real and the remuneration issue is important. Yet the discussions are simplistic in that they assume there is only one "real" type of review, the critical, and presumably that the other forms are hardly worthy of the name. Completely overlooked are the many other types of reviews which may be prepared by persons with skills other than expert subject knowledge.

These types differ in the kind and degree of intellectual effort necessary for their preparation. Clearly the lowest form is the bibliographic. A distinction, borrowed from the field of abstracting, might be made between the indicative bibliographic review, which is simply a bibliography in narrative form, and the informative. The latter may be considered to require greater "criticality" since significant or comparable data must be extracted from a variety of publications and presented. From these two we could proceed through reviews requiring greater effort such as the interpretive, the state-of-the-art, the evaluative and finally reach the true critical review perhaps written by the outstanding expert. Critical reviews survey entire fields or sub-fields, analyze and evaluate developments, synthesize the distilled results and suggest the direction the field should take.

The various types of reviews can thus be considered to be points that can be located on a continuum based on the intellectual effort that has been applied to them as in Figure 1.

FIGURE 1

REVIEWS ON A CONTINUUM OF CRITICALITY

TYPE OF REVIEW	MORE CRITICAL
EXPERT EVALUATIVE ANALYTICAL	CRITICAL
STATE-OF-THE-ART	TOPICAL
INTERPRETIVE POPULAR	EXPLANATORY
INFORMATIVE INDICATIVE	BIBLIOGRAPHIC
LESS CRITICAL	

As one proceeds up the scale the reviews become less literature oriented and more topic oriented and progressively more sophisticated in both the intellectual skills applied to them and the subject expertise involved.

This structuring has several merits. It reveals to the scholar an "intellectual dimension" of reviews which may be capable of further development. It helps the designer to orient himself and, since he knows the type with which he is dealing, it provides some basis for achieving consistency. It helps the administrator to distinguish the variety and degree of skills necessary for preparation of different publications.

Yet the structure is far from satisfying. For one thing, we might accept the fact that some reviews are "higher" and some "lower" in intellectual level. This assumption has obvious validity in the case of the bibliographic versus the critical. Yet in the middle range one would be hard pressed to place the analytical, interpretive, or synthetic on a higher or lower point of the continuum.

An alternative and perhaps more useful structure would be to consider types of reviews on a time scale relating to users' information requirements. The scale would begin with the germination of an idea or the identification of a problem and end with public use of an adopted technique or technology. Early reviews would deal with literature designed to bring the problem into focus for administrators and legislators and to convince them to allocate resources. These reviews may range from simple administrative memoranda to formal research proposals or even to popular reviews appearing in the mass media. If commitments are made to research the problem, state-of-the-art reviews are prepared. As the work proceeds into the research phase there may be need for the methodological and analytical and the synthetic reviews. As the research phase ends and adoption begins there is again need for reviews to interpret the research in language understandable to administrators. Following this second type of commitment, practitioners (teachers, engineers, etc.) require interpretations that allow them to use the results of research as developed techniques or technology. Finally, popular reviews might be prepared that enable the general population to appreciate the advantages and disadvantages of what has been accomplished so they can support, reject or use it.

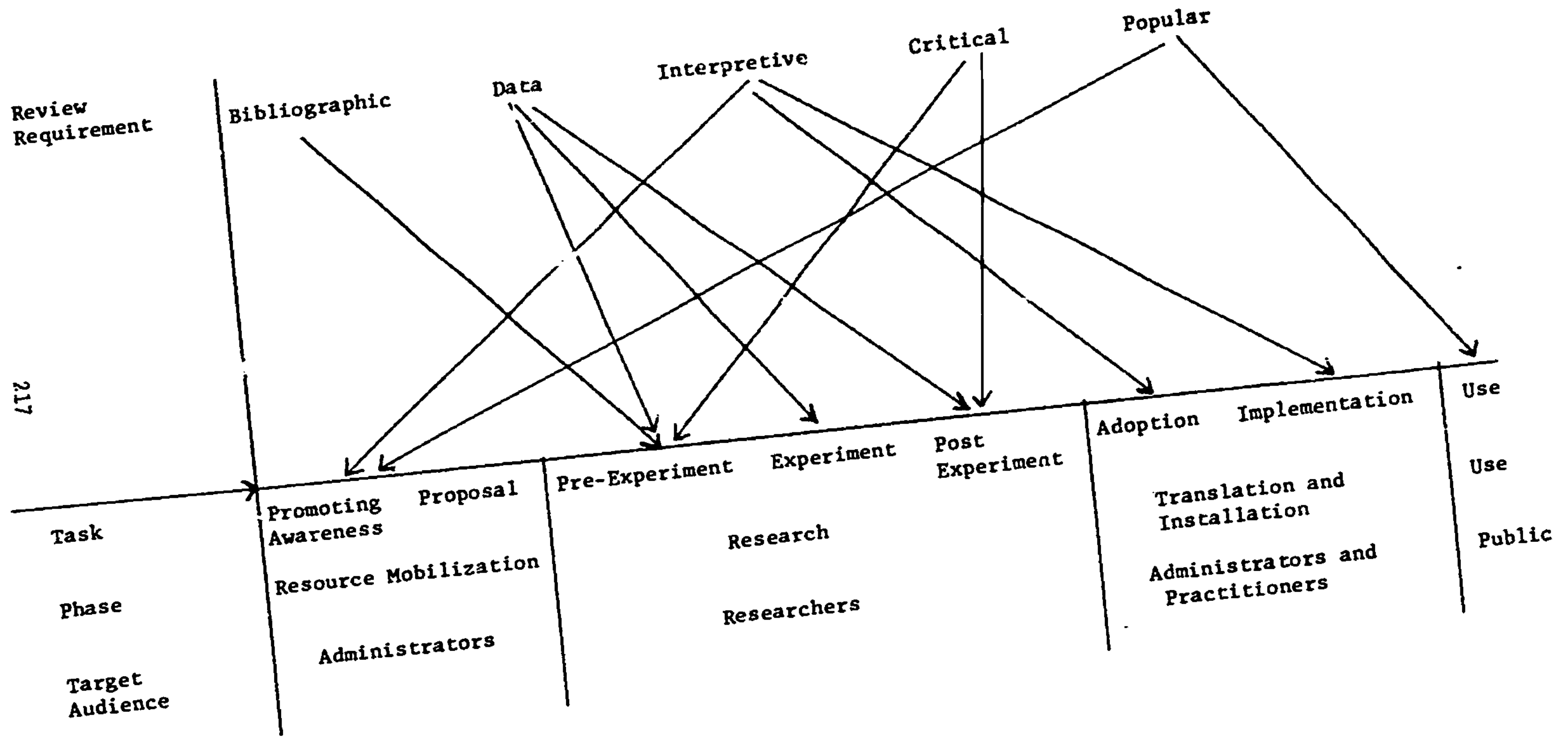
Figure 2 provides an example of such a structure showing how different types of reviews are required for tasks necessary to the development of an innovation. Typical audiences are also indicated. Because of space limitations only five types of reviews are shown, and there are, of course, many more. It should also be recognized that not all tasks and phases are listed, for example a development phase and a maintenance task might also be included.

This developmental model with variations, "loops", and fill-ins has been found useful by those interested in "technology transfer" beyond the research phase. Various types of reviews have always been used within the research stage, interestingly enough in a less systematic way than by those who plan the transfer of techniques and technology.

Both of the foregoing models, the critical and the developmental, suffer from an important defect. They only describe one review dimension. They do not reveal characteristics of reviews other than the intellectual. Yet it is from user requirements in an additional dimension, requirements that reviews be timely, comprehensive, authoritative, readable and so on that those charged with specific planning or preparation derive some of their most important criteria. Something must be added to these models.

Figure 2

Needs for Five Types of Reviews as an Innovation Develops



III Information Analysis

Perhaps insights into other dimensions and further specification of the intellectual dimension can be arrived at by examining users' purposes in consulting or reading reviews. If we knew their goals perhaps we can design vehicles to reach them.

Purpose of the User

From a review of the deductions of previous authors 1/ and the little empirical evidence available, it would seem that reviews perform four major functions. These functions are similar to those afforded by: (1) textbooks (2) alerting services or current bibliographies (3) reference books and "finding devices" and (4) inspirational works.

The textbook function accommodates those who need to achieve an understanding of an aspect of their own field with which they have limited familiarity. Similarly it accommodates those who need orientation in a new or peripheral field.

As there are a variety of textbooks available on a single topic, designed for different audiences, so there are a variety of reviews that perform the same function. Some translate from one technical language to another. Others interpret material at a popular level. In these ways, reviews provide a means to achieve understanding regardless of the nature of the original material. On this point, Scott Adams noting the institutionalization of this effort in medicine, made the interesting observation that in medicine, the Advances characterize the research level, Yearbooks the practice level, and reviews of the Scientific Monthly type the popular level. 2/

Reviews are not only used to aid in understanding new or peripheral areas or material on a different technical level. They may also be utilized in the same way a student will use a variety of textbooks, that is, to "see" difficult material from a different angle or to make it "sink-in."

In all these ways reviews serve those who are not able or do not have the time to perform the intellectual processing necessary to make original papers usable for their purposes.

A second important function is the current bibliographic. Users, whether they be practitioners, researchers or teachers must maintain current awareness in their field if they are not to become "obsolescent." In some cases, particularly with researchers, they must have assurance that they have not missed recent significant works in the course of their regular reading. When reviews are used to serve these purposes they fulfill the same functions as alerting services or current bibliographies although they more often go beyond the mere provision of references and often supply actual information sufficient to make further consultation of the literature unnecessary. There is some evidence that this continuing education or current awareness function is the most important one served by reviews.

Indexing and abstracting services, data compilations, directories, etc. provide data, information or reference. Reviews are often used for the same purposes. It is true that the review will often provide information in a more timely way or in a context that includes caveats, suggestions or alternatives. Despite this the basic purpose of use is for reference.

Finally, these publications may be used as a source of inspiration, a way of maintaining flagging interest, of stimulating work in new directions. They are a source of new ideas and a "creative bed" for their development.

Actual data on users' purposes in reading or consulting reviews is sparse. Usually they can be found only in bits and pieces within broader user requirement studies. Recently, however, two research studies 3/, 4/ have appeared that provide relatively comprehensive data on purposes served by two specific review publications.

Using both a critical incident technique and a structured questionnaire to study the Annual Review of Information Science and Technology (ARIST) Carlos Cuadra arrived at results that tend to confirm the validity of the four categories. At the same time he provides data on relative use within each category. 5/ Some of his results are presented in Table 1 where they are rearranged to show how the categories he discovers "fit" into the four described above.

Percent of Readers Finding ARIST Useful in Serving Various Goals

Category and Goal	Percent
Textbook	
Learning about an area not in reader's specialty	22
Current bibliographic	
Keeping up with own area	33
Keeping up with peripheral area	35
Reading the original literature more selectively	15
Reference	
Checking on particular project	19
Checking on individuals	5
Inspirational	
Identifying areas that require further research	14
Allocating R&D funds	2
Other	4

Table 1

Similarly when the critical incident technique was used responses were highest in the keeping current and reference areas and in responses related to learning.

Another investigator, Linda Harris, found that the major purposes of use of another review publication (the readership of the Review of Educational Research) were also for learning, keeping current and reference. 6/ In this structured study she included an additional category for "to make sure I hadn't missed important literature." This category demonstrates another aspect of the current bibliography approach emphasizing appropriate selectivity and comprehensiveness rather than currency.

III Information Analysis

If it is difficult to find data on the purposes of the user, it is almost impossible to find them on the purposes of the producer. Two years ago in "breaking ground" for another project, the author surveyed 20 directors of information centers. The directors were asked in a very open-ended way to indicate their purposes in producing a total of 200 review papers. Responses seemed to fall naturally into five general classes: one literature oriented and four topic oriented. They are tabulated in Table 2.

<u>Producers Purpose in Publishing Reviews</u>

Category	Percent of responses
Aid user to use the literature	40
Provide analysis and evaluation	32
Aid in putting research into practice	16
Educate and stimulate	8
Answer repetitive inquiries	4

<u>Table 2</u>

These replies do not relate to a particular type of review as in the case of the periodic type in the Cuadra and Harris studies since the directors were free to produce any type of review they wished including the periodic.

The results are probably skewed since the centers were at that time only recently organized so that one would expect an over-emphasis on bibliographic reviews while a "data base" was being established. Fewer topical and critical reviews would be expected in information centers than in other centers.

Yet the users and producers do show considerable correspondence in their purposes. Aid to the user in using literature is composed of such replies as "update the user," "update a publication," "provide a comprehensive compilation." It corresponds rather strongly to the reference and current bibliography approach. "Analysis and synthesis" relates to the textbook approach as does "aid in putting research into practice," since they all involve teaching or the preliminaries thereto. The "educate and stimulate" purpose finds its counterpart in the textbook and inspirational types of use.

It is certainly encouraging to find that the purposes of these publications seem to match the purposes of their users. Thus, if a practitioner needs to have a summation of research translated into language he can understand (textbook approach), it is essential that producers match this requirement. If a community of researchers finds it necessary to keep current in a field peripheral to their own (current bibliography approach), it is well that producers fill this need. The categorization proposed thus provides general criteria for whether a type of review is to be produced. It also provides some guidance for the monitor of review programs and a starting point for the student of reviews. The categorization also gives the administrator a tool for structuring the requirements of the marketplace and possible for devising strategies.

Yet the model is not quite satisfactory to the student of reviews. The student needs a way of analyzing their finer structure and of quantifying review parameters and user needs in order to formulate and test hypotheses. The review designer obviously does not need the detail required by the scholar yet he does have a need for finer detail than has been provided so far.

The following section is devoted to how this can be done; how criteria for planning and evaluating reviews can be formed from three review dimensions: intellectual content, characteristics and substance. It will be shown how selection of criteria for design and subsequent evaluation takes place through consideration of user requirements in these areas and how specification and quantification of these criteria may be accomplished in some cases. Scholars will find implicit in the discussions suggestions for further study.

Dimensions and Criteria

In 1959 Isabella Leitch characterized reviews as being of seven types: the periodical review, the occasional, the analytical, those that review data, those that review concepts, the interpretive and the creative. 7/ The periodical is distinguished by its limited and regular time coverage, usually restricted to a year. The occasional review is not distinguished by time but by the expertise or authoritativeness of the writer. It is intended to be selective and not comprehensive. The analytical review was taken in part to be an "... inquiry into the deductions that may be drawn from an accumulation of results treated as a new whole." 8/ Data and concept reviews were obviously distinguished by the things they reviewed. By interpretive review, Leitch did not intend to describe one that translated data or theory into terms understandable to the reader but rather referred to the interpretation and deduction necessary on the part of the writer. The creative review possibly corresponds in part to the synthetic "... the highest and rarest which takes data from more than one field and shows that they are related and what the relation is." 9/ Here she was writing of the generation of new hypotheses and not merely of putting a piece of research in its context.

It is not important at this point that we agree on the definitions and formulations proposed by Leitch. At a future date we can come to an agreement on the meaning of "analysis," "synthesis," "interpretation" and so on. What is important is to see that she was really describing reviews across three dimensions. One dimension is the type and degree of intellectual processing that has been brought to bear. The resulting content constitutes an intellectual dimension. This dimension includes the condensation and transformation that has been applied to information or data to convert it into a form useful to a particular audience or discipline. Some reviews are almost lacking in this dimension in that they merely describe or give references to the material they cover.

Others are rich in that they not only survey the field, identify the trends and pinpoint the significant, but they actually create new knowledge by formulating new hypotheses.

A second dimension is apparent in her description of periodic reviews and occasional reviews where characteristics such as timeliness, periodicity and selectivity are mentioned. This describes a second dimension, the characteristic dimension. Substance and other content constitutes the third dimension and in the above case it is represented by data and concept.

Since 1959, there have been other attempts to describe types of reviews. On occasion, new types have been uncovered. More often other names are provided for what is really the same concept or new characteristics are announced. Despite this, what we really find in these descriptions are really combinations and reformulations of and discoveries within the three dimensions.

To give one of many examples, in 1961 Scott Adams described the Jahresberichte or discipline review as "... a comprehensive, descriptive record of annual contributions... not created to be critical or evaluative but to provide a systematized running record, year by year, of the contributions made within the disciplines." 10/ Here we can see parts of all three dimensions identified and then combined to conceptualize an important type of review - and all in one simple statement!

There are other indications of the validity and usefulness of these three dimensions. For one thing the few studies that have been published tend to group data according to the above categories. Menzel, 11/ for example, provides a table in which he groups the responses according to characteristics and intellectual content. Cuadra includes a table where percent of use is given for substance. 12/ Also it should be pointed out that the dimensions coincide with the major universal facets proposed by many information scientists, namely substance, process and property. While additional facets are often advanced depending on the discipline, these three are usually proposed as the coordinates or dimensions by which any concept can be described or analyzed.

As discussed above, if we consider the various ways by which reviews have been designated, we find that they are merely expressions of one or more aspects of the three dimensions. The following table illustrates this point.

Types of Reviews

Dimension	Common Designations of Reviews
A. Intellectual content	Evaluative
	Analytical
	Interpretive
	Bibliographic
	Critical
	Creative
B. Characteristic	Comprehensive
	Periodic
	Authoritative
	Popular
	Occasional
	Current
C. Substance and Content	Substantive designation
	Data reviews
	Review of concepts
	Methodological
	Interdisciplinary

Table 3

As described previously, these designations contain the factors that may be used as design and evaluation criteria. Before they can be used in this way they must be listed as illustrated in Table 4.

Specification of Dimensions to Form First Level Criteria

Intellectual content	Characteristics	Substance
Analysis	Authoritativeness	Data
Pin-pointing significant	Comprehensiveness	Theory
Discovering shortcomings	Currency	Methodology
Discovering trends and patterns	Periodicity	Ideas
Synthesis	Degree of compression	Citations
Evaluation	Readability	Interdisciplinary subject
Interpretation	Ease of Use	Substantive designation

Table 4

The items listed within each dimension may now be considered to be first-level criteria. This transformation of dimensions into criteria is a normal progression. Of course, to be considered as true criteria they must be progressively specified and ultimately quantified. Even in the form presented, however, they become a shopping list useful in the planning and evaluation of the publications. Figure 3 shows how this may be done.

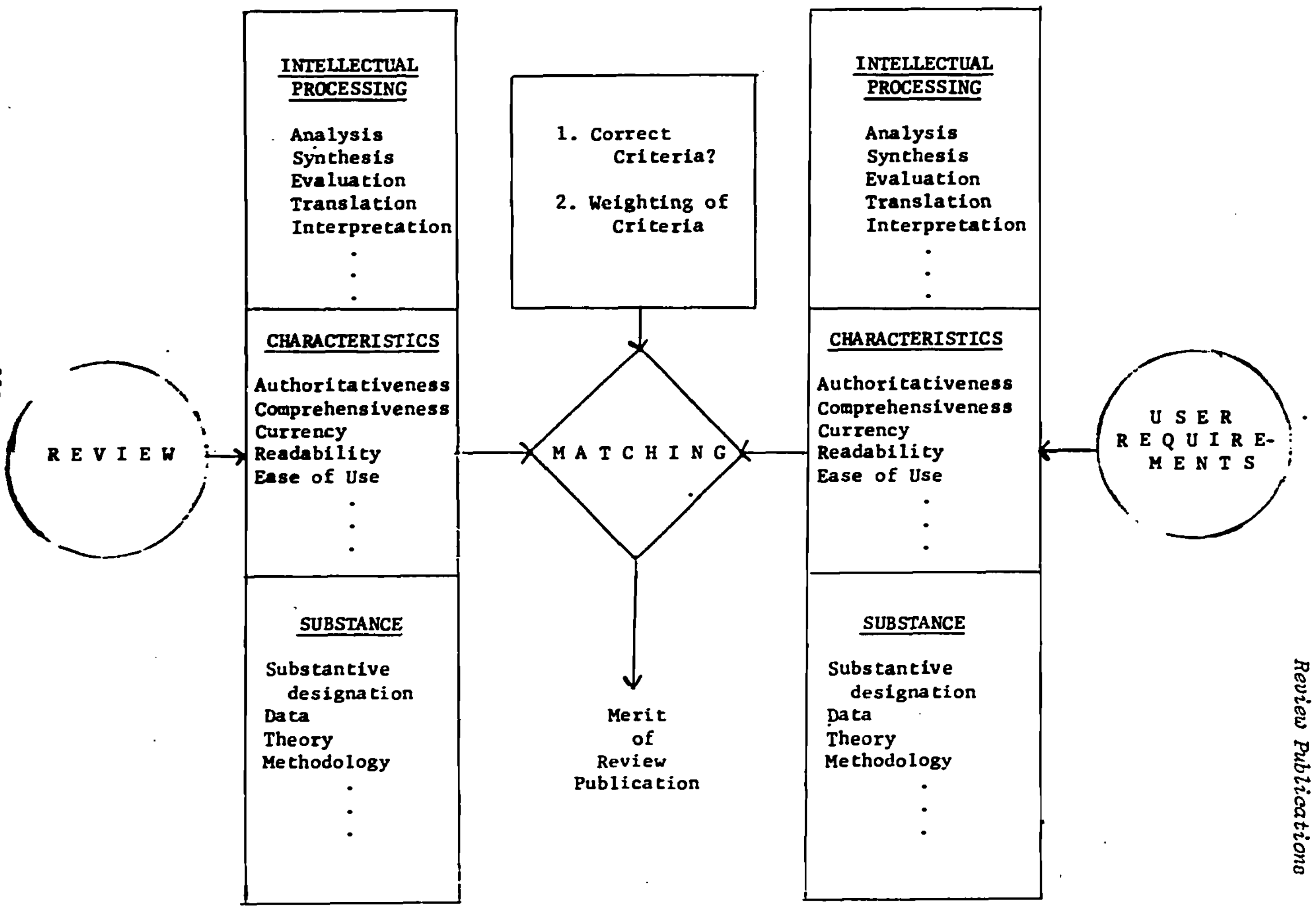
Here, the users' requirements which may be rather amorphous originally are first defined and specified as first-level criteria within all three dimensions. Thus, a general request for information on audio-visual methods in education would be more substantively defined. The other two dimensions would also be specified. Within the intellectual dimension users may or may not require that the material be interpreted or evaluated. In the characteristic dimension, they may require periodic updating on the latest developments in the field or the potential audience may be composed mostly of practitioners so that readability may be important.

Given these requirements, an information analysis center may decide to produce a review for the target audience. The same criteria (in this case readability, currency, etc.) that were derived from an analysis of user needs may now be used in the design of the review. Later evaluation would then consist of measuring the characteristics of the resulting publication against user criteria.

Evaluation criteria developed in recent studies of user requirements may appear to differ from those presented here. Impact and some effectiveness studies seemingly deal in an entirely different coin than that of timeliness, currency and so on. A typical measure of merit in these studies would be that the user's reading of the review did or did not result in an effort to obtain publications that were cited.13/

Figure 3

Design or Evaluation of Reviews



Such measures are useful and necessary but are preparatory to the development of the criteria listed here. To include impact studies this can be shown as in Figure 4 by titling the user circle "User State A" and adding another circle for "User State B." This state is arrived at as a result of the user having read the review. In the case where ultimate effectiveness, e.g. resulting innovations, changes of direction of research projects and so on are considered, an additional circle would have to be drawn for "goal achievement."(Figure 4)

As in all good system work the development of criteria would be a backward progression from the terminal point. Thus a group of users may not attempt to obtain publications cited in a review they have read. In order to either evaluate or design this symptomatic measure must be translated back into the fundamental criteria. That is, it would have to be determined that the review was not current or comprehensive enough or that the format and style discouraged the reader from thoroughly examining the publication.

In addition to design criteria, there are a large group of "producers' criteria" which are outside the scope of this paper. These cannot be considered to be other forms or alternative groupings of the design criteria. They are mostly management criteria derived from the objectives of the producer's organization, its resources and the constraints imposed on it. Additional criteria are derived from the literature environment, mostly the quantity and quality of existing publications and potential duplication or near duplication. Therefore, the decision to produce a review and the selection and emphasis placed on particular design criteria are actually derived from a variety of sources and not just user requirements. This is illustrated in Figure 5.

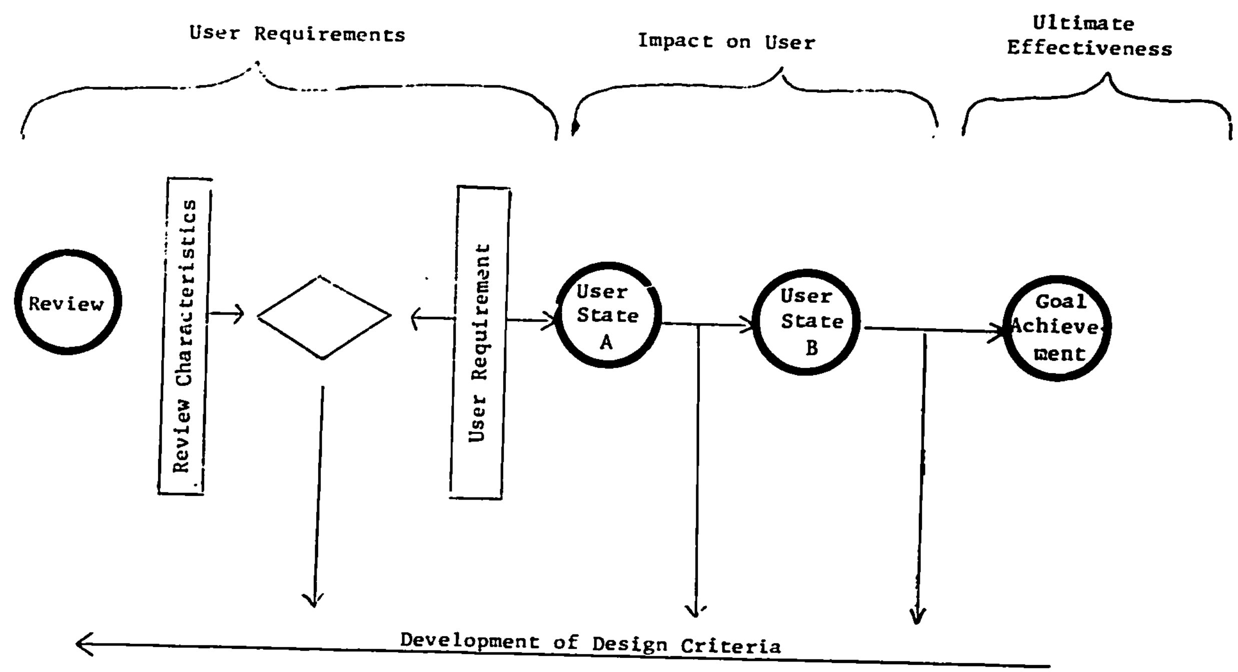
Criteria from the various sources interact. To illustrate, the need for a current evaluation of a particular topic may have been expressed within a subarea of education, and an information center may have received a number of reference questions on that topic. In deciding whether to publish the review the manager may consider that he does not have a staff member who is expert enough to prepare an evaluative review, nor can he find an outside specialist to do the work. Despite this he may judge that the topic is so significant and the amount of literature so large that, using less specialized personnel, he will publish a non-evaluative review emphasizing currency and comprehensiveness. He notes that this product will be of some immediate use and that, if the bibliographic spade work is done, an outside specialist may be more willing to produce a review. Here we have an interaction between user requirements, resources and constraints resulting in the selection of design criteria.

Returning to Figure 2 one can see then that in a judgment of this particular publication a reasonable selection of criteria would include its comprehensiveness and currency but not the degree of evaluation exercised. The director or manager can possibly be criticized for not electing to produce an evaluative review. The publication itself, however, should not be criticized for being insufficiently evaluative since this feature was not selected as a planning criteria.

The foregoing merely provides a structure and a shopping list of design and first-level evaluation criteria that may be used in planning and evaluating reviews. Designers and evaluators need more.

Figure 4

Three Types of User Studies



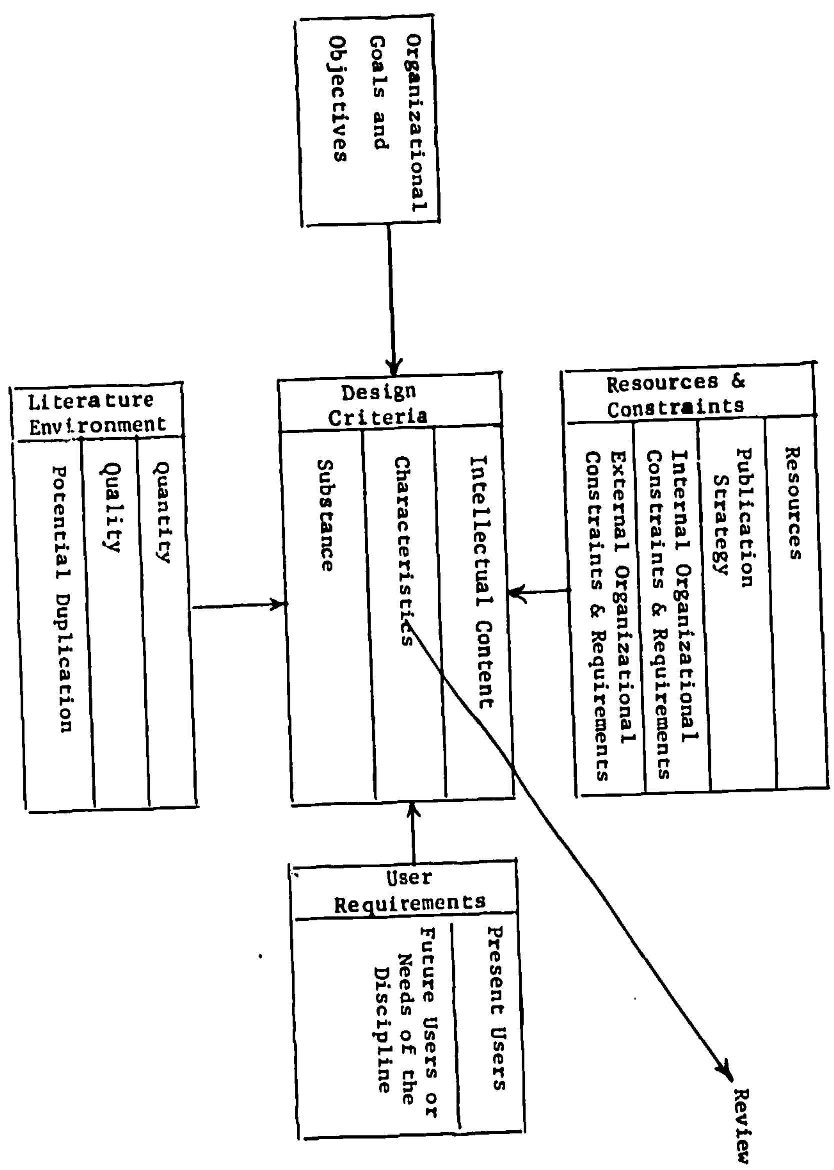


Figure 5 Sources of Criteria in Planning Reviews

Opportunities do exist for making the list of criteria more complete, for defining and specifying them and, at least within the two non-substantive dimensions, for quantifying them.

Benjamin S. Bloom's Taxonomy of Educational Objectives, Handbook I: Cognitive Domain14/ provides definitions for most of the criteria included in the intellectual dimension. Definitions useful to our purposes are provided for terms such as analysis, synthesis, interpretation, translation and others. Also an additional level of specification is provided for each category. For example, "analysis" is defined and divided into three categories (analysis of elements, relationships and organizational principles). This is not to say that these divisions should necessarily be accepted as they stand. For illustration an indication of an alternative type of division is included under "analysis" in Table 4.

There are, of course, alternative models which might be used. J. P. Guilford's familiar "structure-of-the-intellect" model is an excellent possibility.15/

This model has been used in the field of classification in connection with the development of "relational operators" where Farradane basis the basic mental processes he adopts on Guilford's findings.16/ Guilford himself incidentally took some issue with the Bloom taxonomy which he felt contained "an enormous amount of redundancy" in four of six major categories.17/

The characteristic dimension provides ready opportunity not only for specification but also for quantification. Rather sophisticated indexes of readability that involve sentence structure, content, number of syllables and word length have been developed and tested.18/

Attempts have been made to measure currency. A useful way to judge the currency of a publication is simply to consult experts to determine whether recent published studies are included or whether undue reliance was placed on older or even out-of-date publications. This, of course, tells little about the objective age of the material included in the review. Some quantification can be achieved through tabulating the age of the citations in current bibliographies and comparing them with the average age of material included in the review. Similarly, "half-life" studies have been conducted in many fields and subfields. They provide a way of comparing the age of material within a publication with those in the discipline in which its topic falls.

Another opportunity for quantification of characteristic criteria may come about through compression ratios mentioned by Herring19/ and Cottrell.20/ Cottrell proposes the use of a compression ratio (reference pages/state-of-the-art pages) as an evaluation criteria. Herring proposes a similar measure. With reviews of the bibliographic type, recall/precision types of tests might provide a measure of merit although it would be difficult to "fix" the universe from which the publications were drawn.

The most difficult dimension to specify is the substantive. As suggested previously, evaluation in this area may best be left to the subject specialist. Yet even here there are opportunities to develop criteria for design and evaluation. While little can be done to specify actual substance, close criteria and even specification can be established for format, and statements can be made that alternative solutions to the problem will be included as will appropriate cautions, caveats and constraints. A number of guides have appeared for authors of annual review chapters, review journals and government-sponsored publications that contain specifications for simple content. There are also guidelines for referees that contain criteria for evaluation. These guidelines which are a principal vehicle for presenting authors with all design criteria can be improved as knowledge of reviews grows.

Summary

Four independent approaches to the formulation of a taxonomy of review publications and the consequent development of criteria useful to their planning and evaluation have been presented. One approach considered reviews on a continuum of criticality ranging from the bibliographic review through the interpretive to the critical. Another considered types of reviews as information requirements necessary to the development of innovations or the solution of problems. A third alternative characterized reviews according to the purpose of the producers and the consumers. A final approach showed how criteria may be derived from (1) the type of intellectual processing applied to the preparation of the review (2) its characteristics (readability, currency, etc.) and (3) its substance. It was shown how evaluation criteria may be developed through the specification of review dimensions.

Regardless which approach is taken, the process of evaluation is based mainly on the "match" between user requirements and review characteristics. This is made possible by conceptualizing them in the same way.

An underlying aim of this paper has been to initiate and encourage the development of reviewing as a field in its own right. While each of the four approaches has its special usefulness, the last mentioned seems to offer the most promise for the academic development of the field. From this standpoint efforts should be first directed toward definition of the intellectual processing that takes place in review preparation and requirements for this processing. The works of Guilford and Bloom offer a starting point in this regard.

It is hoped that interested persons will be encouraged to begin work in this area so unexplored yet so rich in opportunity.

References

1. A recent review of users purposes appears in Virgo, Julie. The Review Article: Its Characteristics and Problems, University of Chicago, Graduate Library School, Chicago, Illinois, March 1969, pp. 3-6.

2. Adams, Scott. "The Review Literature of Medicine." Bibliography of Medical Reviews, Vol. 6, 1961, unpaged.

3. Cuadra, Carlos A. et al. Impact Study of the Annual Review of Information Science and Technology, System Development Corporation, Santa Monica, California, 15 November, 1968, 121 p.

4. Harris, Linda S. Readership Study of the Review of Educational Research, Institute for Communication Research, Stanford, California, February 3, 1969, 88 p.

5. Cuadra, op. cit., pp. 64.

6. Harris, op. cit., pp. 19-22.

7. Leitch, Isabella. The Place of Analytical and Critical Reviews in any Growing Biological Science and the Service They May Render to Research, International Conference on Scientific Information, Vol. 1, 1959, pp. 571-588.

8. Ibid., p. 580.

9. Ibid., p. 584.

10. Adams, Scott, op. cit.

11. Menzel, Herbert et al. Review of Studies in the Flow of Information Among Scientists, Office of Science Information Service, National Science Foundation, Washington, Vol. II, January 1960, p. 143.

12. Cuadra, op. cit., p. 68.

13. Harris, op. cit., p. 1, 23.

14. Bloom, Benjamin, ed. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. Longmans, Green and Co., New York, 1956, 207 p.

15. Guilford, J. P. The Nature of Human Intelligence. McGraw-Hill, New York, 1967, 538 p.

16. "Farradane's Relational Operators," Journal of Documentation, Vol. 24, No. 4, December 1968, pp. 275-277.

17. Guilford, op. cit., p. 67.

18. Klare, George R. The Measurement of Readability. Iowa State University Press, Ames, Iowa, 1963, 311 p.

19. Herring, Conyers, "Critical Reviews: The User's Point of View," Journal of Chemical Documentation, Vol. 8, No. 4, November 1968, pp. 232-236.

20. Cottrell, W. B. " Evaluation and Compression of Scientific and Technical Information at the Nuclear Safety Information Center," American Documentation, Vol. 19, No. 4, October 1968, pp. 377-378.

RETHINKING IN CLASSIFIED CONCEPT COORDINATION INDEXING

S.Sirajul Husain
(Regional Research Laboratory, Hyderabad-9, India)

SYNOPSIS

A "Classified-Concept-Coordination" Model has been built by first mapping a classificatory plan of a Subject and then by treating the classified isolates, after assigning an ordinal notation, as independent semantic entities for free coordination, according to a predesigned algorithm of fundamental categories for indexing and retrieval operations.

The invention of "wheel" by any standard has been and will remain to be a major breakthrough in imparting a tremendous acceleration to the industrial and technological advancement of all times. In the field of organisation of knowledge the discovery of Facets in a subject field has also been an equally great breakthrough. Ever since the evolution of the concept of facet and its most explicit application by Ranganathan, there has been a growing interest in shifting from the purely enumerative to freely faceted classification schemes. The Categories (facets) according to Vickery (3) are real and important feature of the subject matter we index. He further reaffirms (4), "At the very least it (facet analysis) has the virtue of being an explicit technique that can be described, communicated, taught, analysed, criticised, amended and improved. It is not left to uncommunicable and inconsistent intuition. This is the chief reason why the technique merits attention."

Another significant breakthrough in the field of indexing has been the "Concept Coordination" technique which, using certain constraints, tries to transend the traditional subject boundaries and the rigid barriers of classification practices. The technique essentially works by identifying "Concepts" in a subject or subjects as independent semantic entities, to be coordinated in several combinations, within or without a framework of syntactical constraints like "Role Indicators".

Yet another greatly important breakthrough has been the propounding of the Fundamental Categories, analogous broadly to the role indicators, for the formulation of an algorithm for contriving a most helpfully unequivocal syntactical sequence in the indexing elements — the concepts — of a subject matter.

The impact of the above three discoveries on the organisation of micro-interdisciplinary literature has been profound, such that we have two very distinct and well established indexing models available to us; one, the freely-faceted model; two, the concept-coordination model. The freely-faceted model proceeds by a minute, classified scheduling of all the subjects manifested in the literature to be indexed, with the ultimate object of displaying a classified order both in the document file and in the index. The indexing elements being tied up with the classified schedule of individual subjects, the ultimate class number obtained is therefore not only checkered but also tends to be unwieldily long. The Concept Coordination model on the other hand is a bold departure from a formal classification scheme and thrives on the eventual combination of the detached indexing elements of a subject matter. Though the system provides varying degrees of specificity in indexing, and very large possibilities of cross indexing, it woefully lacks the capacity to display a classified sequence either in the document file or the index itself.

An efficient model for indexing therefore should necessarily possess the following basic features: (i) the indexing elements should come from a classificatory plan, (ii) they should behave like highly independent entities for free coordination, (iii) the coordination should be a linear sequence offering a helpful order in the indexing elements, and thus, finally, (iv) when translated into notation, should result in a compact, distinct and elegant class number.

An exploration into such a theoretical model offers ample scope for research in the following areas:

1). The nature and pattern of growth of knowledge and the interaction amongits branches.

2). The nature and scope of facets and facet-analysis of a subject.

3). The treatment of a multitude of isolates emerging from interacting subjects, but irrevocably bound with the core-subject for indexing.

4). The treatment of subject isolates emerging from certain subject-facets themselves due to their probable seminal overlapping.

5). The classified scheduling of interacting subjects; the advantages or disadvantages in indexing.

6). The role of fundamental categories at the stage of designing a classificatory schedule under each subject-facet.

7). The scope of the fundamental categories, how general these should be without being elusive and vague.

8). The optimal sequence in the fundamental categories.

9). Notation.

The need for research has been, from the beginning, pointed out by Ranganathan. He says, "Even without waiting for all these models to get outmoded, it is open to pure models"(6). Considering the weakness of the present position in the information retrieval systems, Farradane, with concern, asks, "Despite the many suggestions made, the variety of systems developed .... has there emerged any understanding of the basic facts on which one can rely and build? Has there been any coherent substantiated theory (available)?"(7).

The Model: As a result of research conducted in the above areas at the Operations Research and Technical Information Division, Regional Research Laboratory, Hyderabad (India), it has been possible to evolve a "Classified-Concept-Coordination" model for indexing. The model offered here embodies certain important constraints along with those which essentially belong to the freely-faceted and the Concept-Coordination models.

1. Universe of knowledge is a dynamic continuum and so are the various constituent subjects.

2. A subject can best be identified, demarcated and divided by the facets included in it.

3. There are as many facets in a subject as there are number of characteristics applied for the analysis of the subject.

4. Each facet of a subject is composed of a system of isolates at various levels, from general to specific, etc.

5. The subject facets and the respective isolates can be arranged in a classificatory plan on the basis of system of inclusion, genera and species, part and whole, etc., or merely

by enumeration, as the case may be.

6. Two or more facets of a subject need not be absolutely mutually exclusive; they may have certain portions of their areas overlapping with each other, thus giving rise to isolates, which either conceptually or by the usage, are common to two or more than two facets, and probably to an alien subject too.

7. There is no benefit in classifying the interacting, alien subjects, and constructing elaborate facet-isolate schedules for each of them. It is, however, of definite advantage to treat the alien isolates as free tallies. Thus, all the isolates belonging to the alien subjects, and some of the subject-isolates belonging to the common areas of subject-facets put together, can be taken to constitute a "Common-Facet" comprising the "Common-Isolates".

8. The index as well as the document file should necessarily display a classified order.

9. An algorithm for indexing and retrieval can be most effectively devised by the conscious use of certain Fundamental Categories which should necessarily be of general nature to encompass the highly probabilistic functional roles of the concepts — Isolates — specifying a subject matter.

10. The number of fundamental categories and their respective positioning in the indexing and retrieval algorithm should be in accordance with the complexity of a particular subject field. The general guiding principle being to descend from the phenomental level near to the seminal level.

11. Identification of subject isolates — concepts — as preconceived manifestations of one or the other of the fundamental categories at the stage of mapping of the classificatory plan itself is undesirable and it only induces rigidity and unhelpfulness in indexing and retrieval operations.

12. Only a linear sequencing of the subject isolates is feasible in practice. Occasionally, more than one linear sequencing, if required, increases the recall factor.

13. All subject-facets and subject-isolates can be assigned an ordinal value, from a heirarchical system of notation, either alphabetical or alphanumeral.

14. The common-isolates can be assigned a purely sequential, non-ordinal, arbitrary notation, again alphabetical or alphanumeral.

15. The common-isolates are necessarily arranged in a purely alphabetical sequence owing to their inherent arbitrary character.

16. The subject-facets and the subject-isolates are arranged in a classified sequence.

17. The subject-facets, the subject-isolates and the common-isolates can all form a single system of concepts - a thesaurus - arranged alphabetically for the purpose of free coordination in any manner for expressing the thought content of a document, irrespective of their classified position and the corresponding ordinal value in the schedule.

Terminology:

Main subject: A subject or a group of subjects, in which indexing is required, constitutes a main subject or a core subject.

Facet: The totality of the isolates on the basis of a single train of characteristic is a facet, in the idea and the verbal planes.

Isolate (Idea): An idea belonging to a facet. In association with a basic class, it creates a class of smaller extension than the basic class.

In the classified-concept-coordination system isolates are considered as synonyms of keywords, descriptors, concepts.

Fundamental category: Concepts of high generality and wide application, fabricated by the mind with direct or indirect reference to the experiential world, and employed by the mind in the interpolation of that world.

Example: As a practical example, a composite subject field comprising Operations Research, Industrial Research Management and Statistics, has been considered because the literature in this field is not only of specialised kind but also it is probably the best example of an interdisciplinary subject.

Subject Analysis: The subject study, for the purpose of facet-analysis and isolate-analysis, was conducted from two angles. One, a thorough study of standard books, monographs, seminar proceedings, etc. was made; characteristics for facet-analysis were identified, and under each facet the subject isolates were separated. Two, from journals and periodicals, the literature trend was studied, and facets and isolates were identified and collected as above. This dual approach afforded a better understanding and consistency in arriving at the characteristics, the consequent facets, and the resultant isolates.

The first kind of subject analysis provided, apart from the subject-facets and isolates, certain other isolates which were found to be common to two or more than two facets and thus were general in scope of their meaning and application. Some of these general isolates could be placed at a position just above the respective facets in the classificatory plan, while some others being much more general in scope were clubbed together as the common isolates. Also, the second kind of subject analysis resulted, in addition to the subject-facets and isolates, a host of such isolates which did not obviously belong to the main subject field. All these isolates, actually emerging from alien subjects and tied up with the main subject matter, and the above common-isolates were pooled together under a common facet. The common-isolates, in this case, belonged to several subjects like chemistry, economics, humanities, computer and chemical engineering, as also, there were proper names of persons and places, and techniques(Figure 1 and Figure 2).

Notation: The classified subject-facets and isolates were assigned an essentially hierarchical, alphabetical notation, using capital as well as small letters in combination. Using this notation for this particular example, it was observed that not only the base provided by the Aa-Za length was sufficient but also all interpolations and extensions, in the sub-divisions, received an accurate or nearly accurate ordinal value in the classified perspective.

The common-isolates, on the basis of their total strength and further probable additions, were all assigned a single, arbitrarily sequential, non-ordinal, two-digital, small-letter notation, aa to zz, each symbol being enclosed in parentheses.

Schedule: The subject-facets and -isolates with their respective notation assigned, and arranged in their classified sequence, constituted the classified schedule. This schedule formed a basis for evaluation and development of the Classified-Concept-Coordination (CCC) system(Figure 2).

Thesaurus: The subject-facets, the subject-isolates and all the common-isolates, with their respective notation assigned, and arranged in a single alphabetical sequence, constituted the thesaurus. This thesaurus was the main machinery for indexing and retrieval purposes.

Classified Concept Coordination Scheme

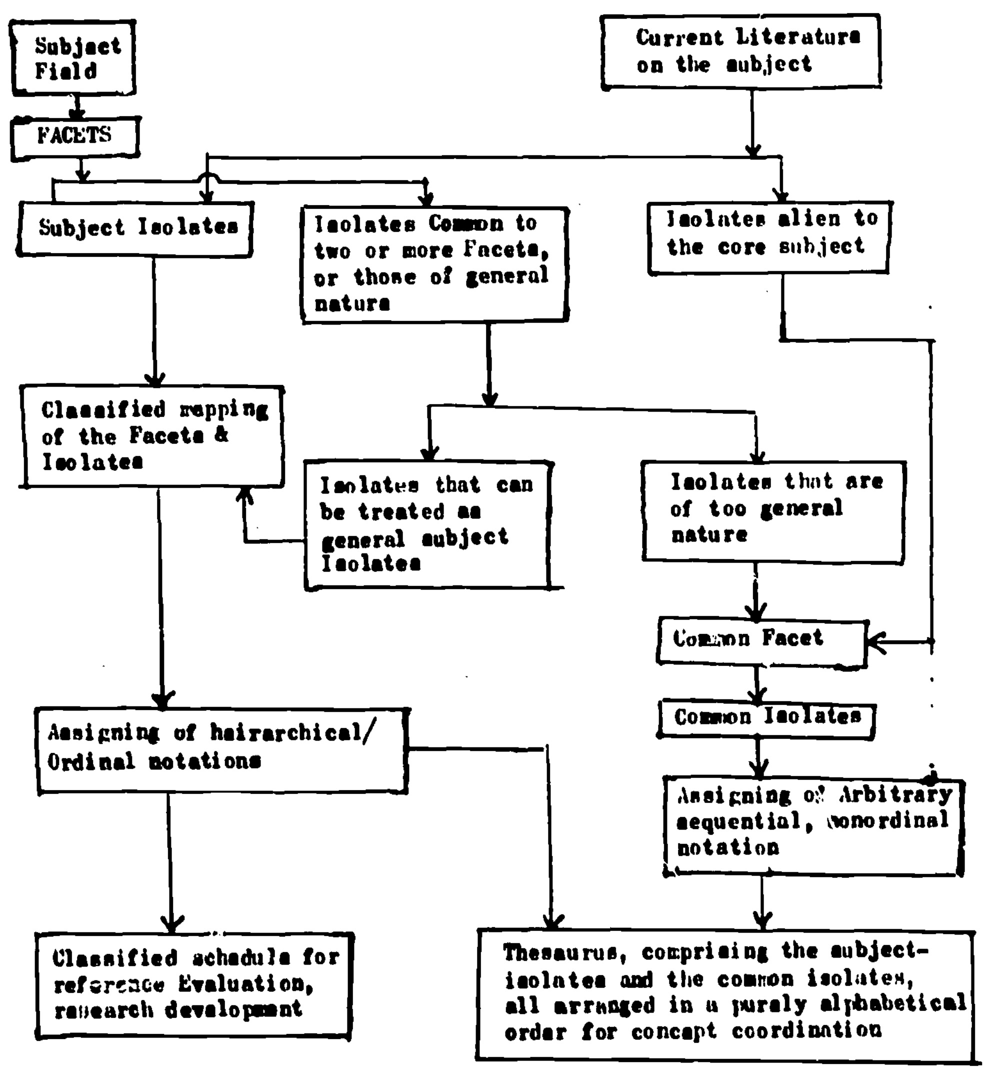


Figure 1

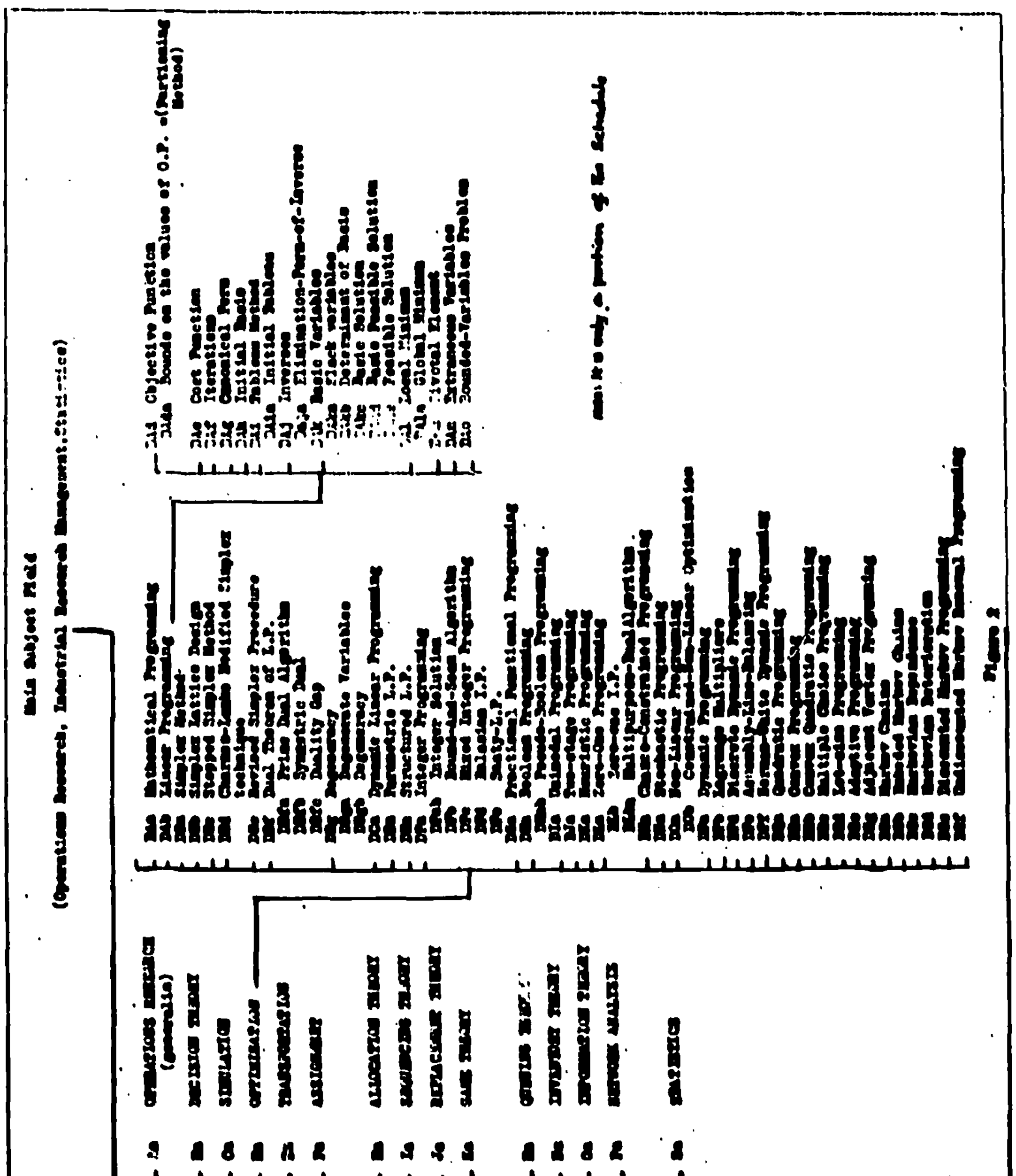


Figure 2

Common Isolates

(ac) Additive Algorithm
(gx) Algorithm
(ai) Automated; Automation
(al) Break-Down
(an) Certainty-Equivalents
(ap) Characteristics
(ar) Chemical Plant
(au) Combinatorial Problem
(aw) Complex
(ax) Computation
(ba) Contract-Research
(bc) Controlled Variability
(bi) Data
(bj) Data-Processing
(bk) Dead-Line
(bl) Decentralised Units/Cost
(hi) Delivery
(bn) Derivation
(bo) Design
(bp) Detached Coefficient
(bq) Detection
(bu) Differential Equations
(bv) Digital Computer
(bw) Discrete-Time
(cd) Distribution
(bx) Economy; Economics; Economical
(by) Electronic
(bz) Enumeration
(cf) Extrapolation
(ch) Feasibility
(ci) Feed Back
(cj) Feed-Mix
(cm) Forecast
(cq) Graph-Theoretic Methods
(cs) Idle-Time
(cy) Information Retrieval
(cz) Information Storage
(da) Information Systems
(db) Input
(di) Iron & Steel Production
(ds) Man-Machine Systems
(du) Mathematical Basis
(ed) Mixed System
(em) Obsolescence
(eq) Oil Production
(eo) Optimum; Optimal
(ep) Output
(er) Parameter
(fb) Production
(fd) Project
(fe) Protein
(fk) Real-Time Systems
(fm) Relative-Location Facilities
(fp) Research
(fr) Returns
(fw) Safety
(fx) Scheduling
(fy) Science Policy
(gb) Sensitivity
(gh) Solution
(gj) Squares
(gk) Strategy
(gp) Survey
(gq) System
(gr) Tangential-Approximation
(gs) Target
(gw) Time-Standards
(hc) Uncertainty
(hf) Variety

Note: It is only a representative sample of the common isolates file

Figure 3

Fundamental categories: From an examination of a large cross-section of the literature appearing in the field, it was found adequate to use four fundamental categories in the following sequence, namely,

Main subject: The overall, broad subject under which a search was most likely to be made.

Focus of study: Specific subject of interest, which was more often found to be an extension of the main subject.

Means: Methodology, technique, procedure, tool or equipment employed.

Conditions: The conditions under which the study is performed.

Any possible attributive isolates with respect to any of the fundamental categories, were placed at the immediately right side of the fundamental category to be qualified.

Connecting symbols: There was practically no need for any connecting symbols between any two fundamental categories, as each notation for an isolate began with either a capital letter or a parenthesis, except when a certain intermediate fundamental category was not manifested occasionally, it was marked by a hyphen, and when an intermediate isolate started with the same notation, its first capital letter was dropped and a point was used in its place, for brevity.

Indexing and Retrieval Strategy: A document was scanned for the concepts that describe the subject matter completely. Each concept was searched in the thesaurus for its notational value. The concepts were fitted into the fundamental category algorithm, and then their respective notations were substituted. That formed the class number for the document. Sometimes depending on the complexity of the subject matter and probable alternative search approaches, a document was assigned alternative class numbers by means of suitable combinations of the search elements — the concepts. The documents as well as the catalogue cards were filed separately in a classified sequence.

Similar procedure was employed for retrieval. A querry was analysed into its concepts and their notational values were noted from the thesaurus, and substituted in the fundamental-category algorithm. The resulting class number (or numbers) was matched with the catalogue and the retrieval was effected.

Examples:

1. DaCc
Ablow, CM., Brigham, G; An analog solution of programming problems; J.Opns.Res.Soc.Am; 3(4),1955; P-388.

2. Ma(fb)-(ar)
Ackoff,RL; Production and Inventory Control in a chemical process; J.Opns.Res.Soc.Am; 3(3),1955;P-319.

3. DJa- -Si(hc)
Agisy,MEL; Two-stage programming under uncertainty with discrete distribution function; J.Opns.Res.Soc.Am; 15(1),1967;P-55.

4. DbAb(eq)
Arnofsky,JS; The use of linear programming and mathematical models in underground oil production; Management Sci; B(4), Jul.1952; P-394.

5. MBj.Ag
Avi-Itshak,B; Preemptive repeat priority queues as a special case of the multipurpose server problem; J.Opns.Res.Soc.Am; 11(4),1963; P-597.

6. DPa.m.Pg
Bellmann,R; On a dynamic programming approach to the caterer problem; Management Sci; 3(3),Apr. 1957; P-270.

7. DAb.Zb(ax)
Bellmann,R; On the computational solution of a linear programming problem involving Almost-Block-Diagnol Matrices; Management Sci; 3(4),Jul.1957; P-403.

8. Pe.j-.ka(az)
Berman,EB; Resource Allocation in a PERT network under continuous time cost function; Management Sci; 10(4),Jul.1964;P-734.

9. Nda.aaCab
Berman,EB; A model for the Monte Carlo determination and optimal stock distribution in a spacial inventory system; J.Opns.Res.Soc.Am; 10(4),1962; P-500.

10. Mo.i.i.
Bhat,UN; On single-server bulk-queuing process with binomial input; J.Opns.Res.Soc.Am; 12(4),1964;P-527.

11. Da.bi.bf
Boot,JCG; On trivial and binding constraints in programming problem; Management Sci; 8(4),Jul.1962; P-419.

12. DPa(gx).Sb-(eu cx)
Came,JS; A dynamic programming Algorithm for embedded Markov Chains when the planning horizon is at infinity; Management Sci;10(4),Jul,1964;P-716.

13. Da(dj).Afc
Charnes,A; On representations of semi-infinite programmes which have no duality gaps; Management Sci; 12(1),Sep.1965; P-113.

14. Pea-DMa.Na
Charnes,A; Critical path analysis via chance cons;J.Opns.Res. Assoc.Am;12(3),1964;P-460.

REFERENCES:

1. Ranganathan, SR; Colon Classification;Ed.4;Madras(India); Madras Library Association, 1952.

2. Ranganathan,SR; Prolegomena to Library Science; Ed.2;London; The Library Association,1957.

3. Vickery,BC; Classification and Indexing in Science; second edition (enlarged),London; Butterworths Scientific Publications; 1959, P-9 and P-12.

4. Vickery, BC; Faceted Classification Schemes;(Rutgers Series on Systems for the intellectual organisation of information 5); 1966; P-26.

5. Ranganathan,SR; Hidden Roots of Classification; LIB Sc V4,N1; 1967 Mar; P-19.

6. Ranganathan, SR; Areas for research in library science; LIB Sc V4,N4; 1967 Dec; P-310.

7. Farradane, JEL; Analysis and Organisation of knowledge for retrieval; Aslip Proceedings; 22(12),Dec. 1970, P-609.

MANAGING IDENTIFIERS

Lawrence Papier

Program Officer, Research and Program Development, Division of Library Programs, Bureau of Libraries and Educational Technology, U.S. Office of Education, Washington, D.C.

SYNOPSIS

A description of identifiers is presented and their significance is set forth. Current practice in managing these terms in several large United States information agencies is outlined, including the use of standards and guidelines. These measures are designed to minimize effort in handling them and to increase their usefulness. Available statistics on their use are provided.

The value of specificity in indexing has long been recognized. A common rule in many indexing systems is to use terms that are as specific as the document or subject they represent. It is commonly accepted that proper use of such terms will reduce the number of nonrelevant documents retrieved when conducting searches.

The purpose of this paper is to describe what is perhaps the most specific type of term that can be used in language-oriented systems. Although literature about these terms is practically non-existent, in many systems they are called identifiers. Their current status will be described and it will also be shown how they may be managed so as to increase their effectiveness and at the same time to minimize effort expended in handling them.

Identifiers are terms of a specialized nature, almost always proper names, which tend to uniquely describe their subject. Identifiers are used to represent organizations, acronyms, coined terminology, equipment and so on. They are not usually subject to lexicographic analysis, structured or entered into the thesaurus itself until, as sometimes happens, they become descriptors. These terms together with descriptors and identifiers constitute the three basic types used in indexing report literature in many systems.

The designation "open-ended terms" has been coined to describe candidate descriptors, candidate identifiers and miscellaneous terms which have not been subjected to lexicographic analysis or structuring. They occur simply because the indexer does not have the time to structure them, or feels that the procedure is unwarranted or that the decision is premature. Yet he feels that it is necessary that the term be made immediately available for indexing and search. The indexer uses it but does not include the open-ended term in the thesaurus pending further analysis.

Although some former identifiers may be included, thesauri are composed of descriptors. Descriptors are terms representing generic concepts which have been structured by providing reference to narrower terms or lower generic classes, broader terms or higher generic classes, related terms, synonyms and near synonyms. Scope notes are often included with the term in the thesaurus proper, and definitions are often separately maintained.

Identifiers, like open-ended terms, are usually excluded from the thesaurus but for somewhat different reasons. They are large in number and, because of their specificity, structuring is less necessary. Since most identifiers cannot be assigned narrower terms or lower generic classes, less is to be gained by this procedure. Perhaps more importantly, because of their large number (note, for example, geographical terms) their inclusion could make the thesaurus's use cumbersome if not unmanageable.

A term that is an identifier in one system may be a descriptor in another. An important factor is the discipline or mission served by the information system. For example, in a transportation system, the term "10 ton trucks" might well be an authorized descriptor. In a medical system, where there might be some interest in their exhaust emmissions, the term might be used as an identifier but would probably never become a descriptor.

Not all indexing and abstracting services distinguish between descriptors and identifiers. In many systems, all terms are organized into the thesaurus as they occur. A case in point would be the information system of the National Aeronautics and Space Administration (NASA). This system, while recognizing the existence of identifiers, treats them in exactly the same way as descriptors. Terms are either structured as descriptors or they are excluded. This procedure acts to prevent the proliferation of terms and insures that all terms used by indexers receive the advantages of structuring. It also allows the use of one authority list instead of two.

There are several disadvantages that accrue under the either-or approach. Since descriptors are generic terms there is a tendency to merely use one of them rather than to go through the effort of structuring an identifier as a new descriptor. Such actions can preclude search specificity and result in retrieval of an excess of unwanted documents. The advantages of generating a systematic list of identifiers that might become descriptors are also lost. When properly used such a list can provide a valuable aid in establishing broader terms under which identifiers can be grouped. These authority lists may be used as shopping lists of potential descriptors and when frequency figures on their use are kept they supply information necessary to decide on the advantages of converting identifiers to descriptors.

Finally, in some systems large numbers of identifiers have been used--in some more than 100,000. To structure 100,000 descriptors could well consume 100,000 professional man-hours. Even were a system able to support such a burden, it would be extremely difficult to use a structured vocabulary of this size in indexing or search.

Use and Usefulness of Identifiers

Figures are difficult to obtain on the use of identifiers in conducting searches, and it is doubtful that they have ever been compiled in any system. However, figures have been published on their total number and on the number of times they have been used in indexing (posted) in the Educational Resources Information Center (ERIC) system.[1] Comparable figures were also obtained from the Defence Documentation Center (DDC) and the National Technical Information Service (NTIS).

In the ERIC system, as of July, 1970, 11,000 terms designated as identifiers had been authorized. These were posted 37,000 times or an average of three times per identifier. In the same period, 4,600 descriptors were authorized and structured. These were used 370,000 times or an average of 80 times. Thus, identifiers were used in indexing with 1/10th the frequency of descriptors and average usage is even less.

The situation at DDC is a bit more complicated. Establishment of authorized identifiers ended in 1966. At that time there were 7,400. As of December, 1970, these had been used in indexing 204,000 times or an average of 28 times per identifier. During the same period, DDC had used 7,300 descriptors 6.5 million times or an average of 895 times per descriptor. Unfortunately, these figures do not tell the whole story since DDC has used 154,000 open-ended terms for which usage figures are not available. Most of these terms are more analogous to identifiers than to descriptors. At NTIS postings of descriptors exceed identifiers by 12 to 1.

Even in the absence of unambigious, documented figures one conclusion seems clear. It is common for a larger number of identifiers to be incorporated within a system than descriptors and yet these identifiers are used far less in indexing than descriptors. In addition, although published figures are not available, based on the impressions of several searchers, use of identifiers in search constitutes no more than 10 percent of the use of descriptors.

Despite these shortcomings two factors make identifiers almost indispensable. From an indexer's standpoint, they often provide the quickest and most suitable means of indexing certain documents. A document about the statistical validity of the "Ginzberg-Landau Theory" can be most efficiently indexed with the identifier "Ginzberg-Landau Theory," regardless of additional descriptors that may be used. Identifiers are also of considerable value when conducting searches. This is particularly so when the user is searching for specific information in contrast to general background information on a topic. One can envision retrieving all and only those documents on the "Ginzberg-Landau Theory" in cases where it is not necessary to use other descriptors.

In view of their large number, their relatively infrequent use in indexing and their usefulness in both indexing and search, a sensible stance to adopt toward them is to provide means for their effective and consistent handling and at the same time to limit the effort expended on individual terms. This is in fact commonly done in many systems. Such means are of two types: The use of guidelines and standards for inclusion and the use of guidelines and standards for construction. Additional measures are establishment of indexing rules and the maintenance of usage lists.

Guidelines and Standards for Inclusion

Standards that have been adopted by various systems include: lists of permissible categories, published reference compilations and authority lists of previously used terms.

In many systems, a number of identifier categories may be provided. Use of these categories may be obligatory or permissive. "Obligatory" means that only identifiers within the prescribed categories may be used.

"Permissive" means that the categories are only used as suggestions and to prescribe format. In most systems, the use of categories is permissive. Those who use obligatory categories argue that to allow the use of identifiers outside of the categories after it is clear that few, if any, additional ones will be added is unnecessary. They feel that the use of obligatory categories does not unduly restrict the indexer and not to do so can defeat a principal tool for preventing unmanageable proliferation.

Identifier categories are highly dependent on the subjects and disciplines covered by the indexing system and after a time tend to become fixed. When the sponsoring organization adds or substracts from the discipline covered or has a major change in mission, categories can be added or substracted. With regard to "discipline-dependence" it is interesting to compare identifier categories generated by DDC and ERIC as shown in Table I.

Table I

Lists of Identifier Categories in Two Organizations

DDC	ERIC
1. Aircraft	1. Acronyms
2. Electronic Equipment	2. Organizations
3. Computer Programs	3. Geographic Locations
4. Engines and Motors	4. Tests and Testing Programs
5. Forms	5. Textbooks
6. Geographic Names	6. Trade Names
7. Launch Vehicles	7. Legislation and Assistance Programs
8. Military and Organization names	8. Methods and Theories
9. Missiles and Rockets	9. Names of People
10. Ordance Items	10. College Courses
11. Personal Names	11. Indian Tribes
12. Programs, Projects, etc.	12. Conference and Meetings
13. Satellites	13. Projects
14. Ships and Boats	
15. Storms	
16. System Programs	
17. Test Explosions	

Only four of the categories are common to both lists, providing evidence that outside of general categories each indexing service must develop its own list.

Another way of providing a standard for inclusion is to maintain an authority list of identifiers that have been used and to add new terms as they occur. Since identifiers are freely added and the authority lists are not usually structured, even with "see" references, the maintenance of such an aid may be questioned. Sometimes, in the case where a term is a "cliff-hanger," it might play a part. For example, if "ABC Radio" has previously been established (possibly because it was the major subject of the document) the indexer may decide to include the related "ABD Radio" even though he might not ordinarily do so. Also, if a tally of the number of times a term is used is kept, the value of the list in deciding on inclusions is enhanced.

Published lists of trade names, catalogs, and specialized dictionaries are sometimes authorized as external authority lists. A rule is made that any term appearing in a designated list may be used in indexing.

Rules applying to identifiers usually govern their form rather than their inclusion and appear within indexing guidelines rather than as separate publications. An example of an indexing rule governing inclusion would be, "If an organization is treated as the subject of a document enter (establish if necessary) the name of the organization as an identifier." A principal aid to the indexer then in deciding on whether to establish a particular identifier is a list of permissable categories. Internal authority lists of identifiers are less important for this purpose, since a principal intent of using these terms is to provide greater latitude to the indexer. The list should nevertheless be maintained as a list of candidate descriptors. As will be seen below both internal and external lists serve as aids in determining form of entry.

Guidelines and Standards for Construction

Both ERIC and DDC use guidelines and external authority lists in establishing form of entry for identifiers. NASA, which treats identifiers in exactly the same way as descriptors, has a few guidelines for "identifier-like descriptors." ERIC has published their identifier guidelines as part of their Rules for Thesaurus Preparation.[2] In addition, ERIC has published a complete listing of identifiers together with frequency of use for each.

The ERIC identifier guidelines provide general rules for abbreviations, capitalization and allowable number of characters, followed by rules within the specific categories displayed in Table I. The major aid provided by the guideline is the use of examples for each category as is shown on the sample in Figure 1.

Figure 1

Example of ERIC Rules for Identifiers[3]

3.2.6 <u>Legislation and Assistance Programs</u>

3.2.6.1 Legislation referred to by proper name is entered:

McAteer Act
Taft Hartley Act
Elementary and Secondary Education Act

When legislation is indicated as an acronym, a double entry is made:

ESEA Title III
Elementary Secondary Education Act Title III

3.2.6.2 Assistance programs that are designated by some authority at the Federal, State or local level are entered:

California Assembly Bill 1967

When assistance programs are indicated as acronyms, double entries are made:

ESEA Title I Program
Elementary Secondary Education Act Title I Program

3.2.7 <u>Other Identifiers</u>

3.2.7.1 Specific names of methods and theories are entered:

Montessori Method
Logogen System
Chomsky Competency Model

3.2.7.2 Names of people used as the subject of a document are entered:

Ernest Hemingway
Stephen Crane
Plutarch
Shakespeare
J.P. Jones

3.2.7.2.1 Do not use a period following the authors' initials; use a space.

General DDC identifier guidelines are an integral part of their guidelines for open-ended terms. Open-ended terms are considered to be of two types, " ... terms which could eventually be authorized as descriptors"[4] and those which might become identifiers.

A major general rule relates to parallelism, that is, that open-ended terms analogous to descriptors should be similar in form to descriptors and open-ended terms analogous to identifiers should be similar in form to identifiers. Another rule, relating to hierarchy, provides that, when an identifier is used, an additional descriptor or broader identifier should also be used to allow for hierarchical retrieval. Other general guidelines deal with adjectival qualifiers, term length, numerals, punctuation, abbreviations and size. A typical rule is "To indicate a size class of items follow the size with a qualifier for the item."[5] . This rule for composed identifiers seemingly reintroduces a hierarchy problem since the indexer has the choice, for example, of using the terms 10-ton vehicle instead of 10-ton truck. However, if the indexer follows the common indexing rule term, that the term be on the same level of specificity as the document, and in addition if he "covers" the term with a parallel descriptor or identifier the problem is diminished.

The second half of the guideline gives specific guidance on the forms of entry within the 17 categories mentioned previously. In some cases these guidelines are relatively terse while in others they are detailed.

The rule for programs, projects, operations, etc. states simply, "Reverse the usual order of the parts of the name with the qualifier. For example, write Project Quick Fix as Quick Fix Project ... "[6] . Guidance for geographic names on the other hand is rather detailed. Here, sub-rules are provided for abbreviations, natural features, political units and geographic qualifiers.

In addition to these rules, internal and external authority lists act as standards to establish and insure consistency in form of entry. DDC incorporates references to authority lists into their rules. Thus the rule for Engines and Motors states in part, "Use the military designation, if available .... "[7] and then cites Jane's All the World's Aircraft as the authority. ERIC lists external aids as references in its Rules for Thesaurus Preparation. When an entry can be found in an authorized external list it goes farther toward insuring consistency than even the most carefully devised rules.

Published directories and compilations are also used as authority lists. Some are general enough to be applicable to almost any system. The United States Government Organization Manual used by ERIC is an example of a general tool. Portions of Jane's Fighting Ships constitutes a specialized tool used by DDC as does Buros' Mental Measurements Yearbook, referenced by ERIC.

As mentioned previously a difficulty in using external listings is that they soon get out of date and it is important, when there is a choice, to select one that will be regularly updated. A new item can be indexed in a variety of ways within the system until a preferred term is established through the sanction of the external thesaurus. Similarly, terms may also become scattered for want of the cross-referencing often provided in an up-to-date external listings.

Usage

As is the case with descriptors, figures are usually maintained on the number of times each identifier has been used in indexing. These figures have a number of applications. For one thing, they tell the searcher how many documents he can expect to retrieve using that term. From this information the searcher can then make an estimate of whether to choose a different identifier or descriptor or whether he must use a coordinate descriptor or identifier.

Frequency-of-use statistics are also important on the input side when deciding on whether and how an identifier should be structured as a descriptor. An identifier that has only been used once or twice would probably not be made a descriptor. An identifier that has been used a large number of times might (if possible) be provided with a number of narrower subdivisions when structured. Usage figures do not seem to be systematically maintained and, in any case, are not available for the number of times identifiers have been used in searches. This is unfortunate. Such figures, and when possible, figures on the number of times they have been used in successful searches would provide additional evidence not only on whether a descriptor should become an identifier but also on the merits of identifiers themselves.

Caveat and Conclusion

The reader should not infer, from the foregoing, a more orderly picture of the status of identifiers than is actually the case. While many organizations would agree with the definition of identifiers presented here others do not. NASA considers identifiers to be only those subject terms that include a numerical or alphabetical designation. [8] . Lists of "identifiers" often contain open-ended terms. There is considerable variation in practice in the use of standards and guidelines and, of course, some do not use any at all. There is the overriding issue of whether identifiers should be treated differently than descriptors. Even within the same organization there is sometimes disagreement as to whether identifiers will or should become descriptors.

This situation is not surprising in view of the fact that identifiers have never been systematically studied and many facts about them are lacking. Additional studies and the dissemination of results are necessary to more efficient and effective management of these important terms.

1. ERIC Processing and Reference Facility, ERIC Identifiers. Leasco Systems and Research Corporation, Bethesda, Maryland, June 1970.
2. Educational Resources Information Center. Rules for Thesaurus Preparation. Washington, D.C. 1969. 534 p.
3. Ibid. p. 17.
4. Defense Documentation Center. Guidelines for Forming Open-Ended Terms. Cameron Station, Virginia, 1968. 13 p.
5. Ibid. p. 5.
6. Ibid. p. 11.
7. Ibid. p. 7.
8. National Aeronautics and Space Administration NASA Thesaurus. Vol. I Washington, D.C., December 1967.

USER-BASED AND USER-ORIENTED CLASSIFICATION SYSTEM: SEQUENCE OF SUBJECTS AND OF COMPONENTS IN A SUBJECT

A. Neelameghan and M.A. Gopinath
Documentation Research and Training Center,
Indian Statistical Institute

0 Introduction

01 Scope of the Paper
This paper considers the essential features of the theoretical basis for the design of a system for the classification of subjects suitable for use in a document finding system. The system recognizes a more or less stable pattern in the organization of the components in subjects, preferred by a majority of intellectuals. At the same time it provides facility of selection of documents on the particular subject of interest-at-the-moment of each reader irrespective of the name of the component idea in a subject which the reader brings up in using the system.

02 Terminology
To facilitate discussion on the subject of this paper a few terms are defined below. The definitions are taken from the Prolegomena[2].

021 Subject
A subject is an organized or systematized body of ideas, whose extension and intension are likely to fall coherently within the field of interest and comfortably within the intellectual competence and the field of inevitable specialization of a normal person.

022 Isolate Idea
An Isolate Idea is an idea or idea-complex fit to form a component of a compound subject other than the Basic Subject in it.

023 Basic Subject
A Basic Subject is a subject without any isolate idea as a component.

024 Compound Subject
A Compound Subject is a subject with a Basic Subject with one or more isolate ideas as components.

Example: "Raman spectroscopy of ultraviolet light in physics" is a Compound Subject. The component ideas in the subject are "Raman spectroscopy", "Ultraviolet light" and "Physics".

According to Colon Classification,
"Physics" is the Basic Subject; "Ultraviolet" and "Raman spectroscopy" is each an isolate idea.

025 Facet
The term 'Facet" denotes any component -- be it a Basic Subject or an isolate idea -- of a compound subject.

026 Classification of Subject
The classification of a subject consists essentially of

1 Recognizing the component ideas in the subject;

2 Determining the appropriate degree of interrelation among the component ideas; and

3 Arranging the component ideas in a linear sequence according to their mutual filiation as determined by the degree of interrelation among them and preferred by a majority of the specialists in the subject.

1 UBIQUITOUS ROLE OF CLASSIFICATION OF SUBJECT
The classification of subject as defined in 026 above occurs in several areas of library work and service[5].

For example, in
1 Classifying documents;

2 Preparing library tools, such as

21 Designing and developing schemes for the classification of subjects,

22 Preparing subject headings,

23 Preparing a thesaurus for a subject; and

3 Reference Service[16], as in
31 Specifying the subject of the reader's query through reader-librarian dialogue, and

32 Classifying the subject of reader's query.

11 Synergistic Effect
Thus, classification is incident in one way or other, in several stages in the design, development, and use of a document finding system. Therefore, the efficiency of the document finding system depends, in a large measure, on the efficiency of the scheme for classification used in the system. Any improvement in the methodology of classification would effect improvements in the efficiency of the document finding system as a whole several fold. In the design, development and use of a document finding system this ubiquitous and important role of classification has to be kept in view.

2 EFFICIENCY OF DOCUMENT FINDING SYSTEM
21 Objective
An efficient document finding system should enable the pin-pointed, exhaustive and expeditious selection of documents on the specific subject of interest to the reader at the moment -- that is, providing a service satisfying the Laws of Library Science[9].

22 Influencing Factors
The factors that affect and bear upon the design of an efficient document finding system are associated with the attributes of

1 The universe of subjects; and

2 The psychology of readers.

Some of these attributes are briefly mentioned in the succeeding sections.

221 <u>Universe of Subjects</u>
1 The universe of subjects is ever growing. The rate of growth is accelerating.

2 New ideas and subjects are being thrown forth incessantly and at an increasingly faster rate.

3 Each new idea or a new subject may fit in a particular position in the universe of subjects according to its filiation to the existing ideas and subjects.

4 The universe of subjects is multidimensional -- that is, several characteristics may have to be used to individualise any one idea from all others in it.

5 The interrelation among the existing subjects may be disturbed due to the emergence of new subjects.

6 There may be new modes of formation of subjects -- that is, modes of combination of ideas.

222 <u>Psychology of Reader</u>
1 The facility to communicate his exact requirement at the moment varies from one reader to another and even with the same reader in different contexts.

2 An idea emerging in the wavefront of research may not even have a specific name and the reader may have to use devious ways and the inadequate existing terminology to express his subject interest at the moment.

3 Specialization concentrates attention on a very narrow subject -- that is, a subject of great intension and small extension. The number of components necessary to coextensively specify such a subject may be large. The reader may not recall all the components at the same time. Further, at the moment he may be most intensively concerned with one or only a few of the component ideas. And he may recall only the names of these at the same time of using the document finding system.

223 <u>Dynamic Context</u>
Thus, it is multidimensional, ever-growing, turbulently dynamic universe of subjects that the document finding system has to deal with. Further, the dynamic feature of the psychology of reader in the search for information mentioned above has also to be taken into account.

3 METHOD OF APPROACH
One approach to achieving the objective mentioned in section 21 is to develop a document finding system having the features mentioned in the succeeding subsections.

31 Coextensive Classification
The scheme for classification used should enable the <u>coextensive</u> classification

of each of the subjects of the documents selected as input to the document finding system. Coextensiveness consists in

1 Recognizing each of the component ideas in a subject and the appropriate degree of interrelation between them, and

2 Arranging and representing the component ideas in a linear sequence preferred by a majority of the specialists in the subject so as to explicitly display the variation in the degree of interrelation among the components.

Coextensiveness is to be achieved irrespective of the intension and extension of the subject; and not only in classifying the subjects of the past and of the present but, at least, those of the near future also. This will help in minimising noise and leakage in document search and selection in response to the requirement of the reader.

32 Similarity of Pattern
It is helpful if in arranging the components of subjects there is a similarity of pattern everywhere -- that is, similarity in pattern of arrangement of

1 Components in one and the same compound subject recurring again and again;

2 Compound subjects going with the same Basic Subject; and

3 Compound subjects going with different Basic Subjects.

33 Resilience
Coextensive classification of subjects of the future requires that the scheme should have built-in capacity for helpfully accomodating new ideas, new subjects, and new interrelation among component ideas and among subjects, thrown forth by the universe of subjects from time to time, without the basic pattern of arrangement of ideas in a subject and the sequence among subjects being appreciably affected. Further, the document finding system should be capable of accomodating and meeting the variety of approaches of a variety of readers in searching for information about subject.

4 PROBLEM OF TRANSFORMATION

41 Linear Sequence
Library classification requires the arrangement of the components of compound subjects in a linear sequence. The number of compound subjects going with even one and the same Basic Subject is far too large to be arranged helpfully and consistently without the aid of explicitly stated guiding principles. For, classifying is equivalent to transforming the n-dimensional configuration of the universe of compound subjects into a linear configuration. The difficult problem of "invariant" arises in such a transformation.

42 APUPA Pattern
An arrangement of the compound in a sequence helpful to a majority of the specialists and thus satisfying the Laws of Library Science gives an APUPA pattern -- that is, Alien-, Penumbral-, Umbral-, Penumbral-, Alien subjects. But it is impossible to secure an APUPA pattern everywhere in the entire range of compound subjects. For, it requires keeping invariant every immediate-neighbourhood-relation among all the subjects while transforming the n-dimensional universe of compound subjects in a line. Add to this the problems of a turbulently

dynamic growth, tendency to become infinite and tendency to become a continuum, of the universe of subjects.

43 Consistency of Pattern of Sequence
Thus, the classificationist can keep invariant one and only one of the many immediate-neighbourhood-relations found in the n-dimensional configuration of compound subjects. The question is the immediate-neighbourhood-relation of which point should be kept invariant, and which should be kept as successive removes[1,2,3]: Conjecture by classificationists as to what is helpful to the majority of readers may not yield consistent results -- that is, one and the same pattern in the sequence of compound subjects. But consistency in the pattern of sequence of compound subjects in the context of all Basic Subjects would indeed be more helpful to all concerned -- classificationist, classifier, and reader.

5 POSTULATIONAL APPROACH

51 Near-Seminal Level
It is felt that the desirable invariant-complex cannot be seized by searching for it in the phenomenal level of the universe of compound subjects. A dive to a near-seminal level has yielded something practical, thereby bypassing the confusing picture presented by the phenomenal level. The design of a scheme for classification and classifying could be based on a set of Postulates for the identification and separation of component facets of a subject and a set of guiding principles for determining the helpful sequence of all such facets. These have reinforced the principles already formulated to arrive systematically at a helpful sequence of isolates in an array and of subjects in general. The Postulates are acceptable so long as they lead to a helpful sequence of subjects in the universe of subjects and a sequence of component ideas in a subject. This helpfulness implies that the right invariant-complex is preserved. This process of arriving at a helpful sequence of the component ideas in a compound subject is known as 'Facet Analysis'.

52 Hierarchy of Guiding Principles
Consistent and helpful development of classification cannot take place in isolation. An integrated approach is necessary. The General Theory of Library Classification developed in India provides this approach. It is based on a hierarchy of about a hundred explicitly stated guiding principles[13]. These are:

1 General Normative Principles applicable to many disciplines;

2 The Five Laws of Library Science applicable to the Library Science and Library as a whole;

3 Canons for Classification; and

4 Principles -- such as the Principles for Sequence of Isolates in an Array -- applicable to a particular piece of work in designing schemes for classification.

53 General Theory of Library Classification
The General Theory of Library Classification developed in India and its application have been detailed elsewhere[3,6,11]. The main features of this theory are mentioned below:

1 The pursuit of library classification in three distinctive planes -- idea

plane, verbal plane, and notational plane. In this functional division of the work, the idea plane is assigned paramountcy.

2 Working at the near-seminal level, bypassing the confusing picture presented by the phenomenal level of isolate ideas and the variety of interrelation among them so as to grasp something more stable and practical.

3 Provision of a set of explicitly stated postulates and guiding principles to secure efficiency and consistency in each of the three planes of work in designing a scheme for classification and in classifying.

4 Securing a built-in dynamism in the theory by basing the latter on the findings of continuous research and study of the attributes of the universe of subjects, the psychology or readers, and the practice of classification in libraries and in other contexts involving the arrangement of ideas in a helpful sequence.

5 Formulation of a rigorous terminology for the subject in order to minimize the possible aberration in thinking and in communication of ideas.

54 Reduction of Work

Thus, the almost impossible task of enumerating and arranging a very large number of subjects -- almost tending to infinity -- in a helpful sequence in the universe of subjects, is reduced by the General Theory of Library Classification to the enumeration and arrangement of a small number of

1 Main Subjects;

2 Basic subjects going with one or the other of the Main Subjects; and

3 Isolates in each array in the schedule of isolates. In these cases, arrangement in a helpful sequence can be determined by a few objective principles instead of by flair.

55 Restricted Scope of the Paper

With this background in mind, we shall now consider:

1 The pattern of sequence of component ideas in a subject; and

2 The general pattern of arrangement of subjects.

We shall consider these factors

1 From the angle of the prescriptions of the General Theory of Library Classification;

2 The helpfulness of the resulting sequence to a majority of the readers; and

3 Overall design of a document finding system to meet the requirement of a majority of readers in general as well as the specific interest-at-the moment of a particular reader.

6 BASIS FOR FACET SYNTAX

61 Absolute Syntax of Ideas

The sequence in which the component ideas of compound subjects with a Basic Subject usually arrange themselves in the mind of the majority of normal intellectuals may be called Absolute Syntax of Ideas among intellectuals. It is conjectured that such a syntax of ideas exists. It may not coincide with the linguistic syntax -- that is the syntax of words in all languages. Investigation by a team consisting of reference librarians, classificationists, specialists in linguistics, epistemology, psychology, anthropology, and statistics has been suggested to examine the validity of the conjecture[10]. This is yet to happen. In the meanwhile, however, several helpful results in the theory and practice of classification have been derived on the basis of this postulate about Absolute Syntax. The possible basis for the conjecture is mentioned in the suceeding section.

611 Biological Basis
An idea is generally a product of intellectual activity. Intellectual activity is known to be controlled by the brain. There is considerable similarity in the structure and, therefore, the functioning of the brain in most of the normal human beings. In other words, a majority of the normal human beings have more or less a similar mode of thinking and learning -- that is, in forming ideas and in combining them to build subjects.

It is further believed that biologically man has not changed to any appreciable extent since he became Homo sapiens many thousands of years ago. The educationist William H. Kilpatrick writes: "It is generally believed that man has not developed biologically in any significant manner or degree since he became Homo sapiens; culturally, yes -- greatly so; biologically as regards mind, no, not so far as we can tell"[2]. Or again, as Bertrand Russell writes: "There had been a time when there were biological improvements in brain capacity with corresponding advance in genetic capacity. But that time ceased some 500,000 years ago"[15].

612 Inference
From the above discussion we may infer that:

1 The mode of thinking/learning among normal intellectuals is similar;

2 The mode of thinking/learning among normal intellectuals has remained basically in the same pattern for several thousands of years; and for this reason;

3 The probability of a sudden change -- that is, a mutation -- in the mode of thinking/learning of a majority of normal intellectuals in the near future is quite low.

62 Facet Sequence
It has been suggested that the sequence of the component ideas in compound subjects -- that is, the facet syntax -- should parallel the Absolute Syntax of Ideas. For, if we can secure a facet syntax parallel to the Absolute Syntax then the pattern of sequence of component ideas in compound subjects will be

1 Helpful to a majority of normal intellectuals; and

2 Consistent and more or less stable and hence continue to be helpful to a majority of intellectuals at least in the near future.

The sequence of component ideas thus derived is reader-requirement-based. Incidentally, the aberration likely to arise by using the verbal plane in structuring and naming subjects, due to the variation in linguistic syntax, is reduced to a minimum.

63 In Practice
A number of principles for securing a sequence of ideas deemed to be helpful to

a majority of readers have been formulated and explicitly stated. These have been used for several years now in designing schemes for classification of subjects, in classifying, and in other contexts[14]. One of these Principles of Helpful Sequence is the Wall-Picture Principle. It states: "If two facets A and B of a subject are such that the concept behind B will not be operative unless the concept behind A is conceded, even as a mural picture is not possible unless the wall exists to draw upon, the facet A should precede the facet B"[8]. It has been found that the Wall-Picture Principle

1 Is the most ubiquitous in its application; and

2 Secures a sequence of the component ideas in subject that is helpful to a majority of specialists.

It may, therefore, be said that the facet syntax resulting from the application of the Wall-Picture Principle parallels the Absolute Syntax of Ideas. It has been further found that the other Principles for Helpful Sequence are corollaries of the Wall-Picture Principle. Therefore, the sequence of ideas resulting from the application of any of the other Principles for Helpful Sequence will conform to that derivable by application of the Wall-Picture Principle itself. In a particular context, the use of one of the other principles may be more convenient than to work with the Wall-Picture Principle.

64 Example
Given below are examples of names of subjects to demonstrate the variation in the sequence of component ideas in a compound subject, viewed from the angle of Linguistic Syntax and Facet Syntax arrived at by Facet Analysis using the postulational method.

641 Example 1

English: Antibiotic Treatment of Bacterial Diseases of Lungs in Children.

French: Traitement antibiotique pour les maladies bacteriennes du poumeau des enfants.

German: Antibiotische Behandlung von Bakterialkrankheiten der Lunge der Kinder.

Facet Syntax: Medicine. Child. Lung. Disease-Bacteria. Treatment. Antibiotic.

642 Example 2

English: The Sociology of Alcoholism among the Middle Class people in Developing Countries.

French: La Sociologie de l'Alcoolisme parmi la Bourgeoisie dans les Pays en voie de Developpement.

German: Die Soziologie des Alkoholismus in den Mittelstand der Entwicklungslander.

Facet Syntax: Sociology. Middle class people. Alcoholism. Developing countries.

7 OVERALL SEQUENCE OF SUBJECTS
It is generally found convenient to divide the universe of subjects, as a first step, into a few mutually exclusive and totally exhaustive chunks called Main Basic Subjects or Main Subjects for short. These are not too many in number. Therefore, most of the Schemes for Classification succeed in arranging them on

the basis of some principles, whether explicitly stated or not. The resulting sequence is more or less helpful in most cases. Therefore, we may take the Main Subjects and their sequence to be postulated by the Scheme prescribing them. It is presumed that no main subject can be expressed in terms of the other Main Subjects in the schedule. It is generally found convenient to divide further some of the Main Subjects as a second step, into a few mutually exclusive and totally exhaustive Non-Main Basic Subjects. In a Main Subject these are not too many in number. Most schemes for classification succeed in arranging the Non-Main Basic Subjects going with a Main Basic Subject on the basis of some principles, whether explicitly stated or not. The resulting sequence is more or less helpful in most cases. Therefore, we may take the Non-Main Basic Subjects and their sequence to be postulated by the scheme prescribing them. Colon Classification recognizes Non-Main Basic Subjects of the following varieties: Canonical, Specials, Environmental, and Systems. The term 'Basic subject' is a generic name denoting either a Main Basic Subject or a Non-Main Basic Subject. According to the Postulate of Basic Subject, each subject is deemed to go with one or the other of the Basic Subjects. Therefore, the subjects get grouped, in the first instance according to the Main Subjects. Some of the features of the Main Subjects and their sequence prescribed by Colon Classification may be worth examining.

71 Basic Feature

1 The universe of subjects as a whole is broadly divided into three groups: Natural Sciences, Humanities, and Social Sciences, and arranged in that sequence.

<u>Note</u>.- When we consider a particular Main Subject we take along with it, its Adjunct Main Subjects -- such as, Biophysics and Biochemistry taken along with Biology.

2 Among the Main Subjects in the Natural Sciences group, an applied subject follows the pure subject on which the former is dominantly based. For example, "D Engineering" follows "C Physics", "F Chemical technology" follows "E Chemistry", "J Agriculture", follows "I Botany", and "KX Animal husbandry: and "L Medicine" follow "K Zoology". In the Social Sciences group also a similar sequence of "Pure-followed by-Applied subjects sequence" is discernible. For example, "YX Social Work" follows "Y Sociology".

3 The pure disciplines in the Natural Sciences group are arranged among themselves in the increasing sequence of "concreteness" from B Mathematics to Δ Spiritual Experience and Mysticism.

B	Mathematics	I	Botany
C	Physics	K	Zoology
E	Chemistry	Δ	Spiritual experience and mysticism
G	Biology		
H	Geology		

4 The remaining Main Subjects covering the Humanities and the Social Sciences, from Δ Spiritual Experience and Mysticism to Z Law are arranged in the sequence of increasing artificiality.

5 It will be noted that, looked at from one angle Δ Spiritual Experience and Mysticism is the most concrete subject and looked at from another angle, it is the most natural.

6 The Partial Comprehensions are interpolated in the schedule of Main Subjects in the appropriate positions -- immediately preceeding the first Main Subject it comprehends. For example, "Social Sciences" comprehends "T Education", "U Geography", "V History", "W Political Science", "X Economics", "Y Sociology", "Z Law". Therefore, the Partial Comprehension "SZ Social Sciences" is placed immediately earlier to "T Education".

72 Accomodation of New Main Subjects

There is a group of Main Subjects placed between "z Generalia" and "A Natural Sciences". A number of these subjects have been given the status of Main Subject in CC, Ed 7 (in preparation) only. Heretofore, some of these new Main Subjects have been isolate ideas in subjects going with one or the other of the Main Subjects. Some of them have been ideas denoting practice-in-action in different subjects. Because of the emergence of a generalisation or a theory out of such practice-in-action to constitute a field of specialization by itself and the consequent literary warrant, it has been found convenient to deem them as independent disciplines. Some of these subjects can occur as isolate idea in a large number of subjects going with different Basic Subjects. Some others, such as Library Science and Book Science may have any of the other subjects as component ideas in forming Compound Subjects. Thus, in one way or other, these Main Subjects can be associated with subjects going with different Basic Subjects in the fields of Natural Sciences, Humanities, and Social Sciences. Interpolating a Main Subject of this kind between any two of the Main Subjects in the Natural Sciences, Humanities, or Social Sciences would violate the Principles of Helpful sequence already used in arranging the Main Subjects. For, it is difficult to decide the relative concreteness, or artificiality as the case may be, of any one of the new Main Subjects with respect to the existing Main Subjects. On the basis of these considerations, these new Main Subjects have been placed earlier to the Natural Sciences and after Generalia.

Thus the sequence among the subjects going with different Basic Subjects in CC is more or less helpful and is based on certain principles. Therefore, when interpolating any new Main Subject in the future the pattern will be retained without violent dislocations.

8 READER'S INTEREST AT THE MOMENT

81 Building up from Fragments

In Sec 6 we have discussed mainly the basis of the postulate regarding the sequence of components of Compound Subjects and the principles used to secure a general sequence of the components acceptable to a majority of normal intellectuals. In practice, however, it may often be found that a specialist reader seeking information about a subject of his interest at the moment does not coextensively and precisely express his requirement. He may mention only one or two of the component ideas. In the design of a document finding system this feature of reader's psychology should be taken into account. The system should enable the formulation and expression of the subject coextensively, given only these fragments[4].

82 Attention Confined to Specific Idea

Without going into the details of the psychology of the reader, we can consider three varieties of situations:

1 Among the component ideas in a subject, a specialist may, at the moment, be working intensively on one or only a few of them. As a result of such specialization-at-the moment, in his search for information on the subject of his interest, he may frequently use the name of the component idea he is working on at the moment. Different readers may be specializing-at-the moment on different facets of even one and the same subject. In their approach to finding documents on the subject of their respective interests, the name of the facet of interest-at-the moment will feature predominantly. Such a concentration of attention may shift from one facet to another from time to time even with the same specialist. But, generally the totality of interests of a specialist is likely to be confined to ideas going with a particular Basic Subject.

2 A specialist may continuously -- and not temporarily -- concentrate attention or a particular facet of a subject. For example, "Raman spectroscopy" in Physics, "Antibiotic treatment" of disease in human beings and "Methods of treating alcoholic addiction". Here again, the totality of interest of the specialist is likely to be confined to the ideas going with a particular Basic Subject. The reader is not likely to approach the document finding system with the name of the facet he is interested in.

3 Area studies constitute another variety of specialization on a particular component idea in a subject. This may be a permanent interest of an individual. The totality of his interests may not be confined to the ideas going with a particular Basic Subject. For instance, a specialist on China (Mainland) may want to have information on the state of the different sciences, the arts, education, geography, history, politics, economics, sociology, and law in China. Naturally, in his use of the document finding system 'China' will be predominantly featured. In fact, subjects not connected with China in one way or other may not interest him.

83 No change in Mode of Thinking

The way in which a reader expresses a query about a subject may not coincide with the pattern of the facet syntax. This may be due to

1 Differences in the linguistic syntax, as has already been mentioned (<u>See</u> Sec 6) <u>or</u>

2 The variety of specializations on a specific component idea in a subject as pointed out in the preceding section.

These variations do not imply any basic deviation from the Absolute Syntax of Ideas in the mode of thinking of the specialists.

What is needed is a mechanism built into the document finding system which will

1 Enable the reader to enter the system via the name of the component idea he brings up;

2 Recall to his mind other component ideas he may be interested in; and

3 Bring together within a small range all the entries for the documents on the subject of his interest at the moment.

An Available Model
A system for document finding that takes into consideration the normal mode of thinking among a majority of intellectuals (that is, Absolute Syntax of Ideas), the varied approaches of readers arising from their respective specialization-at-the moment, and other features of the psychology of readers while searching for information on a subject has been developed and refined as and when found necessary on the basis of practical experience. The system consists of a catalogue/documentation list divided into two parts,

1 Classified Part; and

2 Alphabetical Part.

Classified Part
In the classified part the entries for documents forming the input to the system are arranged in a classified sequence -- that is, according to the ordinal value of the digits in the Class Number. The subjects of the documents are classified according to a freely-faceted scheme for classification based on explicitly stated postulates, and other guiding principles[13]. Each entry is fitted with an adequate feature heading -- that is, representation of the subject in the Verbal Plane, structured as in the Class Number.

Alphabetical Part
The entries in the alphabetical index to the subjects of the documents listed in the classified part are derived on the basis of a systematic procedure such as Chain Procedure and its modified versions.

Search and Selection
A reader interested in a specific subject at the moment consults the alphabetical index using the name of any one of the component ideas in the subject. The different subjects in the system having this idea as a component are brought together and displayed in the alphabetical part. The reader selects the entry representing his subject interest. Using the Class Number or Serial Number given against the entry, he is landed in a spot in the classified part wherin the entries for documents on subjects of likely interest to him -- subjects in the penumbral and umbral region -- are arranged in a helpful sequence displaying their mutual filiation, within the range of a small number of entries[4]. Browsing among the entries -- the feature heading, title of document, abstract, etc -- the reader gets a feedback, speedily taking him to the entries for documents of interest to him at the moment. In the browsing process, he is enabled to recall almost all the component ideas in the subject of interest at the moment, although he may not remember them when the search was started. The search and selection of documents on the subject of interest to the reader at the moment is made pinpointed (that is, minimum of noise or irrelevancy), exhaustive (that is, minimum of leakage), and expeditious.

Area Study-Type Interest
An area study-type interest may require collecting together all subjects which have the area or facet as a component. It has been pointed out that for this purpose it is more advantageous to uphold the uniqueness of the structure of subject as represented by an expressive class number and meet the demands of the Laws of Library Science with the aid of the catalogue and administrative method[7].

87 Computer-Aided Document Finding System

Experience during the last two years in the development of a computer-aided document finding system based on the model mentioned in Sec 84, shows that all the advantages of the system discussed in this paper can be fully realised. Obviously, in the computer-aided system, the classified part of catalogue/documentation list alone would be sufficient. The unique structuring of each subject helps in reducing the chances in the selection of irrelevant documents in response to a query. Further, the computer-memory can be fed with the schedules of a Scheme for Classification whose design is guided by a set of explicitly stated guiding principles. With the aid of this the computer can arrange the components of a subject in a helpful sequence or synthesise an expressive class number for the subject, irrespective of the sequence in which the Kernel Terms of the subject, are fed into the computer[1].

9 CONCLUSION

The General Theory of Library Classification, some particular aspects of which have been discussed in this paper, provides a model for a Freely-faceted Analytico-Synthetic Scheme for Library Classification based on explicitly stated guiding principles and postulates. The built-in internal dynamism of the theory enables the methodology for design of the scheme to keep up with the relevant findings about the attributes of the universe of subjects and psychology of readers. The same theory guides the process of classifying such that the findings of the idea plane about the subject of each document can be implemented and co-extensively represented in the notational plane. This is the postulational method of classifying.

91 BIBLIOGRAPHICAL REFERENCES

1 Sec 87 GUPTA, B.S.S. Program-package for a system for document finding. (Library science with a slant to documentation. 7:1971: Paper H). [This paper gives references to earlier reports of the work done at DRTC on a computer-aided system for document finding].

2 Sec 611 KILPATRICK, W.H. Philosophy of education. 1951. New York, Macmillan company. pp 237-47. [Also published in Park (Joe), Ed. Selected readings in the philosophy of education. Ed. 3, 1968. New York, Macmillan company. pp 107-17].

3 Sec 53 NEELAMEGHAN, A. Classification, Theory of. (Encyclopedia of Library and Information Science. 5:1971).

4 Sec 81, 85 --. Integrated approach of India to the design and development of document retrieval systems. (International Forum on Informatics (Moscow) (1968). [Papers]. 1969. pp. 1114-47.

5 Sec 1 --. Theoretical foundation for UDC: Its need and formulation. (Paper contributed to the International Symposium on UDC in relation to other indexing languages. (Herceg Novi) (1971).

6 Sec 53 -- and Gopinath, M.A. Research in library classification. (Library science with a slant to documentation. 4:1967: Paper R).

7 Sec 86 RANGANATHAN, S.R. Conflict in classification for document retrieval. (Library science with a slant to documentation. 2:1965:Paper K).

8	Sec 63	—. Elements of library classification. Ed 3. 1962. Bombay, Asia Publishing House. Sec N32.
9	Sec 21	—. Five laws of library science. Ed. 2. 1957.
10	Sec 61	—. Hidden roots of classification. (Information storage and retrieval. 3:1967:399-410:specially Sec 7.3 to 7.5).
11	Sec 53	—. Prolegomena to library classification, Ed. 3. Assisted by M.A. Gopinath. 1967. Bombay, Asia Publishing House.
12	Sec 02	—. —. Chap. CR
13	Sec 52 f642	—. —. Chap. DA
14	Sec 63	—. —. Part F
15	Sec 611	RUSSELL,Bertrand. Has man a future. 1961. (Penguin special). p. 8.
16	Sec 1	SPECIFICATION OF reader's query. (Seminar on Reference Service, (DRTC) (1971). Working Paper. Chapter BG).

Session Three - Discussions

INFORMATION ANALYSIS

Chairman: Prof. H. Arntz (Germany)

MR. F. LIEBESNY (UK): Languages in themselves are not really the topic of my talk; rather, the meaning of languages - semantics. There is the more recent science of linguistics called "semiotics" which refers to the meaning of signs and which is just as important in the communications scene as semantics.

The best definition of semiotics was given to me by a Hungarian colleague when we were discussing the facilities provided at a certain airport and he asked me if I had been in that particular place. I said, "Oh, that place!", and he said, "You have said enough with one gesture of yours because you have made it quite clear that the facilities provided there are not up to the five-star standard of the Sheraton Hotel in Tel Aviv."

The language problem, whether semantics, semiotics or indeed any other aspects, is an extremely great one. It is not enough to know a language. You have to know the cultural background of the language. You have to understand, particularly in our field of scientific and technical documentation, the subject matter with which you are dealing.

In our discipline we are not concerned with the 3,117 languages quoted in my paper, but with about 6-10 major languages. We know that something like 1.4 or 1.5 million articles are published every year in the field of science and technology, but we must bear in mind that about half of those - between 1/2 and 3/4 of a million - are written in languages other than English, and recent surveys have shown that the next most important language in our field is Russian.

It is important to realise that the people whom we are to serve need our help in the language field as much as they do in many other fields. I have tried to indicate some ways of overcoming the language barrier.

One way is with the help of our dear old friend the computer. An eminent professor at the Hebrew University, Prof. Yehuda Bar Hillel, said some years ago the classical phrase: "Mechanical translation is dead," and he may perhaps be right, but I feel that there are some possible uses for MT and one of them is to use it in the compilation of dictionaries. Work is being done in this field of electronic or computer dictionaries, particularly at the European Coal and Steel Community, the AEA in the UK, and in several other places.

This is one promising use of the computer in the language field. The other one is a combination of auto-abstracting (abstracting by the computer) and machine translation, whereby, for example, the vast amount of Chinese literature which to many of us is inaccessible could be digested in abstract form by the computer and then perhaps be translated.

Basically, in documentation the language problem is not so great, but we must bear in mind that users' knowledge of other languages is normally not as great as that of the provider of information.

Being a translator myself, I have obviously concentrated on the translation aspect n linguistics, but, of course, there are other aspects such as interpreting, which is completely different discipline. Interpreting, however, does not really come into ocumentation to the same extent as translation.

The whole problem of languages is not new; we have been faced with this problem ince the destruction of the Tower of Babel. I would like to quote the eminent French ovelist Rabelais who about 150 years ago defined translation as: "le traduction est omme une femme, si elle est belle, elle n'est pas fidele, si elle est fidele, elle 'est pas belle." The very subject of translation will be apparent if you try to ranslate that phrase.

ROF. D. KRALLMAN (Germany): I would like to discuss the construction of a "cumula-ive dictionary" (CD) which is flexible in size and in which there exists no problem f unknown words because the dictionary does not deal with words. It is not construc-ed especially for application in information systems, but in general, to manipulate erbal language data.

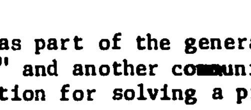
The dictionary is to be seen as part of the general communication model which onsists of human communicator "M" and another communicator "C" - the computer. etween them there exists cooperation for solving a problem.

The computer plays the role of the communicator. To enable it to perform this unction, it is necessary to program at least some parts of human work behavior. You an't assume that an individual has a complete dictionary in his brain, but rather a echanism that can generate and construct ideas, words, phrases, sentences and the elations between them. The CD is based on the assumption that the vocabulary of a anguage is reducible to a limited number of semantically-oriented units called formatives" which are classified as basic concepts, left segments and right segments. e assume that each word has at least one basic concept. This basic concept is de-ined as a word formative which determines the semantic conception of the words.

If we analyze a complete German dictionary consisting of 120,000 lexicon entries, he analysis of the German entries in this dictionary (68,000 entries) results in a eduction to 4,500 basic concepts and about 200 left and right segments and 1,300 on-inflectable words. Each word fed into the computer is analyzed and the basic oncepts of left and right segments are classified. In the generation routine, the asic concept of a word is generated by the statements and directions of the human ser.

There are a lot of ambiguities with this routine. For example, if you have the ord "KANTINE", you can segment it into KANT, IN, and E, and you have a basic concept f KANT and two right segments, female IN and E. If you start with the basic concept hich determines the semantic conception of the word, KANTINE must be the female form f the philosopher Kant.

As to the relation between words in this dictionary, we assumed that they can be enerated by the relations between concepts. Concepts are parts of semantically-riented networks, e.g.



. These relations bring with them many problems and these problems are ambiguities in respect to their semantic interpretations.

For example, in the German word for computers, there are several possibilities for segmentation.

AUT/OMA/T/EN or AU/TOMAT/EN

We have programmed some routines for the analysis and generation of words and we have analyzed the reduction from 68,000 words to 4,500 basic concepts and left and right segments.

MR. B. VICKERY (UK): I want to explain why I wrote my paper and what I am trying to say. I called it "The Many Uses and Forms of Subject Representation" and by this I mean the subject tags that we put onto documents in order to help us retrieve references in retrieval systems.

Traditionally this whole area of assigning subject tags to documents has been a battleground. In the old days it used to be,"should they be alphabetical terms or should they be class numbers?" Today the battleground is, "should it be natural language or should it be a controlled language?" In these battles over the last ten years we have had a number of evaluation tests that have tried to adjudicate between different claims of the various types of language. But judging from current practice where all different types of tags and newly invented types keep on appearing, it seems that these evaluation tests do not give enough guidance. We must ask ourselves why.

When designing a manual retrieval system, like the typical library card index, the system is designed to do one job - to provide for specific topical search on a particular subject. In designing mechanized systems, one first thinks about creating a bibliographic record, key-punching into machine-readable form, and then thinks of all the possible ways this record can be used because it is an extensive keyboard and one wants to get good value for money and use the record to produce a number of services.

In two systems I have recently been connected with, we wanted to produce from the machine record a specific index printed out for visual search similar to a card index. We wanted entries that could be searched mechanically, either for specific subjects or for general subjects; we wanted the possibility of selective dissemination from the magnetic tape; we wanted the possibility of grouping entries into bulletins by broad topic. There was a whole series of things that the subject keys had to do.

If you begin to think of the information services as a whole, subject representations have a variety of functions and it seems more and more that the great variety of structures of subject keys, whether alphabetical, classified, natural language, controlled language, coded or whatever, are related to the fact that they are expected to perform different functions.

In my paper I discussed a number of functions that subject keys can perform. Topical search - retrieving information on a specific topical subject; generic search - pulling out material on a much broader subject; group dissemination - pulling out material for a bulletin; selective dissemination; providing printed index entries for visual search; and another aspect which is generally becoming more important, providing the means for transferring data from one system to another.

All these are legitimate and necessary functions found separately or together in information systems and their variety is such as to make one doubt whether any single subject representation, such as a natural language title or abstract, is going to be able to do all of them. So what I have tried to do is to take these functions one by one, and on the basis of some examples to show various ways in which people have designed subject keys and indexing languages, to try and meet this particular function and so build up a picture of the way in which the structure of the subject keys varies according to the sort of function that they have to perform.

We can see the use of natural languages in titles or abstracts; the tagging of some of these words as "significant", either positively by tagging words in the title, or negatively by using a stop list; the truncation of words to shorten them by chopping off letters in the front or the back of the word to help in searching for more general topics; the use of standardized, controlled index terms to improve search performance; the introduction of generic relations between terms; introduction of group terms; standard sets of group terms; linking natural language words into themes in order to produce structured index entries that can be printed out and used in the manner of a conventional index; and so on.

When one looks at all this variety of structure and compares it and relates it to the functions that are performed, it is clear that the structural variety is only a reflection of the functional variety. ASLIB people asked us, which is best, natural language or controlled language? Which should we choose? And our immediate reaction was, what do you want to do with the language?

This question of the purpose for which one is designing retrieval languages is something which needs to be clearly thought out before one embarks on the design. Very often it seems to me that people design the language and then say, what can we do with it?

The arguments that we have had about alphabetical versus classified, natural versus controlled and so on, are something like a carpenter arguing as to which tool he should buy in order to do all the work he has to do - a chisel or hammer or saw or plane or whatever. The question is obviously an unsuitable one for a carpenter and a similar sort of question is equally unsuitable in the retrieval field. We have a lot of different jobs to do with our subject keys. We must choose at different times different tools to do these jobs.

PROF. R. HELLMAN (US): There are three kinds of people in the total information world. First there is the manufacturer of entirely new information. He is also a searcher of publications essentially for the little bit of information which is entirely new. The second person is the reader for inventory. He is not creating anything new. He may be a college student who is doing a term paper. He wants to read what is available on his subject and essentially this is an inventory process. The third type of person is what I call the information banker. He catalogues what is written and makes it available to the researchers and inventories.

My contribution here is on the subject of newness. Every now and then we encounter a book which contains much that is new and this is a treasure, but in general there is very little new because each researcher builds on what was done before him. The tremendous amount of writing which contains nothing new creates a tremendous problem for the manufacturer of new data. The only feasible solution to this is for information bankers to compel an author routinely to list not only what he is writing about in his subject index, but also to isolate his new contribution. This would be a tremendous help in rationalizing the exponential explosion of knowledge which, before long, will make it an absolute necessity that this be done, otherwise computers and everything else will become useless due to inability to select.

I think our problem more and more is selectivity, and selectivity turns on what is new, what I call in my paper "delta new". The researcher doesn't research according to some established system of priorities, but according to what interests him, and this may, or may not, be high on the scale of need. I am not trying to get into the problem of basic research. Nevertheless, I see no reason why a government or institutional agency or group of institutions cannot get together every year or five years and set up a priority list of research needs and make recommendations. This does not mean that the researcher has to conform to it, he is still free. But this would simplify the problem of finding out what is new and we wouldn't have a tremendous amount of literature which makes no new contribution but much of which is written simply to enhance the professional reputation or promotion of the writer. I know my own profession is full of it to the point of being scandalous.

In 1968, a Committee of Congress directed my Agency of Small Business Administration to make a study on why small business in the US was not getting a larger share of the $15-16 billion of R&D money expended by the Federal Government. We called it an Inter-Agency Committee on the Anatomy of the R&D Process and the Economics of Scale translated in the small business role. Participating on the Committee were the Department of Defense, the Atomic Energy Committee, the Space Agency, as well as the Social Science groups.

In the US, small business and manufacturing is defined by law or by regulation as having fewer than 500 employees, so that I don't think there are more than one or two firms in Israel which are not "small". We picked the seven top researchers in the R&D process and gave them small contracts on study design and finally we developed a solution which we wanted to test involving the information processes which I would like to outline because it brings together many of the problems which have been discussed, and introduces the time factor, the timeliness factor.

We found that the government researchers were spending billions of dollars a year for research, but by the time 3 or 4 million small businesses advertised for an R&D proposal, they were out of the picture, since the government researchers had already consulted one or two private firms.

The problem was how could these people find out what is being considered in the R&D field before work on it has started. We proposed a "functional catalogue". The team would break their problem into functional areas which could be given out in pieces, so that a billion dollar research project might be broken down into 30 pieces and one piece might be $.5 million which would suit small business. All these pieces would be classified and described and put into a computer which is essentially a "functional catalogue." The small business man in turn would classify what he does and the amount of resources he is capable of putting in, according to the same "functional catalogue." He would go to a terminal which would be in almost any big city in the US and query the terminal by his categories to see what is being planned. He then could get a list from the computer and can say, "I am not interested in this." "This is too much for me, but this is interesting."

When a team is dreaming up a research project, they can go to the computer, put in their functional categories and pick out what companies, big or small; universities; or institutions know about or are working on, and what competence they have. It can work both ways and much duplication is avoided.

The problem is to develop the functional categories. A tremendous amount of work has been done on research in, for example, sewage disposal on space vehicles and many of these same problems exist today in the sewage disposal of Tel Aviv and New York.

Today there is almost no way to link the two. The purpose of the functional catalogue is to interlink them all. The first step is to develop 200 to 500 pilot functional categories and test them out in the field against actual contracts and actual businesses. If it looks feasible, we shall go on to develop a functional catalogue.

CHAIRMAN (Prof. Arntz): Prof. Hellman, thank you. I found this a very nice idea that the amount of literature - if we were to do what Prof. Hellman proposed - would decrease enormously because many researchers compelled to state in their publication what they are doing would be ashamed to write nothing at all and so perhaps renounce writing the paper.

DR. W.E. BATTEN (UK): (Comments to Liebesny paper). I want to sound a word of sympathy for the good people who are doing the analysis of the structure and nature of language and to say that I hope that they are not fighting a losing battle against the appalling erosion in language standards which is going on all around them.

I hope that they are not analyzing a dead and academic and pure and very elegant form of the language which seems to have little to do with the way in which words are used today. I refer particularly to the pernicious practice of writers today, popular writers, perhaps newspaper writers more than most, of forcing nouns to do the job of adjectives and sometimes adverbs, and generally using groups of letters without any recognition of the fact that there is such a thing as a part of speech. They say that they are straining after some shade of meaning which formal grammar will not permit them to formulate. I don't believe anything of the kind. I think they are throwing words together because they have not thought out what they really mean. This is a terrible handicap for the analysts.

PROF. ARNTZ: Mr. Liebesny mentioned about ten major languages, but we know that, for example, in a chemical abstract journal, literature from more than 60 languages is abstracted. It is thought, for example, that scientific literature in English is understood in India, but this is not so at the engineering level. Exporters in industry cannot deliver descriptions of equipment which goes to India in English, as the users will be engineers or foremen. Such material must be in Marathi or Gujarati for the Bombay region or in Hindi for the New Delhi region.

Until its independence, Ceylon, for example, had been in its upper stratum an English-speaking country, but in 1956 English was eliminated and Singalese introduced as the state language and the language of the publications of the country. Another example is Burmese.

DR. M. KESSLER (US): To Mr. Liebesny. The use of language as an information transfer instrument is only a minor utilization of this instrument. Language is more commonly used for its emotional content; the use of language as purely an information tool is of rather recent origin. So we shouldn't be surprised that it has its limitations. The attempt on our part to shape the language into an efficient instrument of information may be blocked by the other influences that shape the language in the artistic, aesthetic, emotional and just the daily necessities of life.

Perhaps an effort could be initiated to revive the notion of the universal language which is purposely designed to be independent and devoid of all colorful characteristics of daily language - something like Esperanto, which was started way ahead of it's time and failed because 80 years ago it did not have any useful function.

The scientists, engineers and commercial people get along on a very limited level of interaction. They don't have to love each other, they don't have to write poetry for each other - all they have to do is to perform the simple matter of transferring a little bit of information.

I was wondering if some experiments ought to be done in the matter of the universal, neutral, non-emotional language devoid of any aesthetic values and emotional content, to be used just for information purposes. One example of this very interesting possibility of running into trouble with languages took place when people at MIT got interested in ecology. We searched a well-known data bank for titles with ecology and we found hundreds of papers on "gynecology."

MR. LIEBESNY: To Dr. Kessler. I would very much agree with you that we must treat language as an instrument. It is not a very precise instrument and most of us working with languages are only too familiar with the shortcomings and the inadequacies of this instrument and we try hard to reduce these inadequacies by reducing ambiguities, omissions, etc.

But language is - and I am glad about it - a living thing, and I would, therefore, deplore your plea for a universal non-emotional language because the very import of a phrase, of a meaning or even of a raised eyebrow can tell you so very much. I have referred in my paper to the usefulness of pictorial dictionaries since I believe in the old Chinese proverb that one picture is better than 10,000 words and I think that Esperanto has been found by experience to be a dead language. People speak it, but they think in their own language and translate into Esperanto. It lacks communicativeness and flexibility.

DR. H. BUNTROCK (Luxembourg): To Mr. Liebesny. In your paper, you quoted some percentages of the distribution of languages. We did the same for secondary literature and it is interesting to compare the distribution of languages in primary and in secondary periodicals. You quoted a publication of Lloyd's which stated that 44% of primary literature is published in English, 15% in German, 13% in French and 8% in Russian. We found in 1969 for secondary publications, 63% in English, 12% in German, 11% in Russian and 11% in French. Of course all these percentages are only magnitudes and cannot be taken literally.

MR. LIEBESNY: To Dr. Buntrock. Your figures for 1969 for secondary literature are very interesting. It certainly confirms that English is the most important language in science and technology and that German, Russian and French rank second. The frequency of the language ranking varies considerably from discipline to discipline and, for example, in physics, English occupies 87% in the secondary literature. In a survey of other fields English is less than 50%.

MR. H. FANGMEYER (Italy): To Mr. Liebesny. You say that mechanical translation is dead. Is it only so for the "belle and not fidele" translation or is it for the "fidele and not the belle" also? There exists an MT system used by Oak Ridge and by us, and the scientists are very glad to use it and are able to understand what the machines are translating.

MR. LIEBESNY: To Mr. Fangmeyer. I was very glad to hear about the work that is being done in MT in various establishments. Unfortunately we seem not to know very much about the successes, but hear mostly about failures of MT. In answer to your question if MT is dead - it is dead as far as "la belle traducion" is concerned, but as far as "la fidele traducion" is concerned, I think it has got a great amount of good work to do.

MRS. M. ALONI (Israel): *To Mr. Liebesny.* Are there any standards for translators or any examinations or training centers? I understand that the professional simultaneous translations have reached a very high standard so would it not be possible to use simultaneous translators to translate papers instead of having them edited by a computer?

MR. LIEBESNY: *To Mrs. Aloni.* In the UK we have a professional body called the "Institute of Linguistics" which holds professional examinations and has qualifications. We also have two university schools for translators, one at Salford and one at the University of Bath where we train translators who can become professionals. We have a standard code of ethics, we have a standard minimum rate and I would be pleased to give further details in private discussion.

I am afraid I cannot agree with the suggestion that a simultaneous translator should translate papers, because the simultaneous translator - whom I call an interpreter - although he does do translating, has completely different abilities, mentalities, attitudes, from a translator. I work as both so I know a little about it. If I interpret at a Conference I cannot remember afterwards what any of the speakers have said because the moment I have said something I have forgotten it. When I work as an interpreter, I must turn myself into a machine and if the speaker says, for example, this blackboard is red; then I must automatically translate it into whatever language I am using, and not think, "by God, the man is color-blind." An interpreter must turn himself into a machine while a translator must live with the thing; and therefore I don't think this suggestion can work.

MRS. ALONI: *To Prof. Krallman.* Have there been any similar attempts such as that described in your paper made in other languages? Does a book similar to Roget's Thesaurus exist in other languages? There has been a limited attempt at this in Hebrew. Is the cummulative dictionary made mainly by computer or will it eventually take the form of a book?

PROF. KRALLMAN: *To Mrs. Aloni.* I know of a similar work done at Rand Corp. for the Russian language and the results were published recently. Our results are available on tape. They are done on 7090 and on 370, also IBM 360 computer and the results we have are available from our Institute.

MR. M. KELLER (US): *To Prof. Heilman.* I was sorry that as a solution for a very real problem which I appreciate, the author should be required to state what is new. I think that every author will find in his paper something new, because he may not know that it isn't and in any event, since he must get his paper published, many authors will learn to stretch their imaginations; therefore, in my opinion the solution really is for the information scientists to move a step forward and to become themselves the experts and judges of what is new.

Now there are limits to the capacities of information scientists to do this in many fields, but they can, by associating themselves with the right institutions where multidisciplinary faculties are available, use expert advice. They can at the very least learn when to be suspicious and when to ask questions - is this new or is this the same thing only slightly disguised? They then can treat the material on the basis of this expertise as new or not. I think this has to become one of the functions of the information scientist. I would not rely on the authors - I know too many of them.

PROF. HELLMAN: To Mr. Keller. There is no substitute for the author pointing out what is new in his paper because if I, as a researcher, see what the author thinks is new, I certainly can evaluate it very quickly. What I want to know is what he thinks is new. If you leave it to the information bankers you will not get the kind of person you need to pick out what is new and evaluate it. That kind of a person is going to do his own research. He is not going to be a reviewer. There is a tremendous advantage in having the author himself pick out what is new. In many instances he himself will be sufficiently embarrassed either not to publish at all or to do more research to develop something that is really new before publishing.

MRS. L. VILENTCHUK (Israel): To Prof. Hellman. It is a very interesting idea, but it will certainly not help to limit the amount of publications because people publish not because they have something new to say but because they have to advance in their academic professions. Very unfortunately, nobody asks: "are you a good teacher, do you have personality, do you know something?" You are asked - "what have you published?" And not even what but how much have you published. We don't primarily have an information explosion, but a publication explosion and it seems to me that this is the right forum to make a recommendation that the evaluation and advancement of professionals not be based solely on the amount they publish, but that some other standards be offered, and this will probably bring down the amount of publications.

With regard to applied research there is the perennial complaint that it is very difficult to get good people into applied research. This to a great extent is because in applied research, they don't have the easy means of advancement by publishing the results of their work. They often can't publish because what they have done is the property of the organization for which they are working and is restricted. They have no way to prove their value and, therefore, ambitious people don't go into applied research, but into basic research. This creates the great gap between the quality and quantity of basic research and applied research which is so detrimental to the development of industry.

MRS. A.M. DE BUSTAMANTE (Mexico): We have people who want to work as reviewers, for example, for Applied Mechanics Reviews. The professional societies could act as referees and send publications to specialists in different areas who could devote a little time to saying what is new and what is not new.

DR. S. SCHWARZ (Sweden): I think this discussion about newness and truth is quite interesting, but it is not as simple as it appears. It is not easy to recognize what is new at a specific stage and perhaps one asks too much if one tries to define it.

The social system of science is linked to the publication systems and the suggestion that one should change this system might cause inbalance in the system. You can't really say that people publish too much; this somewhat highbrow opinion of the information scientists that too much is published and there is too little information in it does not consider the real advance of science.

DR. BUNDROCK: It is not always as easy to recognize what is new as it might seem. For example, in 1860, Gregor Mendel discovered the Mendel rules and it was 50 years later that two men had to rediscover the same rules. Punch cards were developed by Hollerith in connection with the US census. You can see a 2,000 year-old voting machine in the Agora Museum in Athens which is based on the same principle.

SESSION FOUR

economics of information systems

Authors of Papers

THE IMPACT OF ECONOMICS ON SYSTEM DESIGN

W.E. Batten, Director
**United Kingdom Chemical Information Service,
University of Nottingham, University Park,
Nottingham NG7 2RD, England.**

SYNOPSIS

The total organisation of scientific literature into searchable 'packages' is likely to be achieved by merging and partition, at various levels. Since 'packages' require a market, the partitioning pattern is likely to be a compromise. Market volume and user taste will interact to require a variety of information techniques.

It will be obvious to this audience that to design a system merely to produce a prescribed technical result, without measuring the likely cost of achieving that result, is something that does not belong outside the research department. It will also be obvious that all operations have a greater or lesser element of 'fixed cost' in them and that therefore the product of an operation must have enough 'appeal' in the market to spread that fixed cost over a number of customers. This leads on to such notions as condensing and compacting the file record so far as possible, and to arranging it in such an order that frequently consulted data is easiest to access. Many other rather obvious devices will occur to any competent systems designer. The same system principles apply even in non-computerised systems.

I do not propose to occupy your time with such technical desiderata, important though they are. There are far more difficult and involved problems attaching to the economic organisation of information as a whole - all the way from the primary literature down to the ultimate sub-disciplinary distillate.

As I see it, this is essentially a distribution problem rather than a technical one; a complex distribution problem in which the very technique of distribution is itself fundamentally affected by market needs. Whereas manufactured goods at least emanate from a single source, and are usually distributed through wholesale and retail networks, literature emanates from a multitude of individual sources and must be first associated into saleable packages before the process of distribution can redirect it to a multitude of individual recipients. This process of grouping has, traditionally, been carried out by learned and professional societies; only recently have purely commercial influences entered this part of the network.

The historical and contemporary picture is, therefore, that the primary journals tend to be the co-current contributions of members and fellows - ie. it is their society fellowship and membership that brings their contributions together between covers rather than a close communion of personal pursuits. Of course, this pattern has become modified as the volume of the technical literature has grown. Many famous journals have adopted the technique of sectionalisation so that, at least, "analysis" is separated from "physical chemistry". Even so, one wonders whether it should be!

It is not cynical but purely realistic to conclude, therefore, that primary papers become grouped for publication in such a way that a journal can find enough paying subscribers to keep it alive. It is a matter of:- "if x% of the contents interest me, I will buy it". And very rational too!

But we have reached the situation where the average scientist needs to have his quota from not one but from each of (say) 20 journals. His total requirements could very well constitute a "journal" in its own right - if only there were enough other people with exactly the same pattern of technical appetite. But, usually, there are not enough such people. So we remain with the situation that everyone must buy (or at any rate consult) primary journals that contain a lot that does not interest them and must use many primary journals (all subject to the same limitations) to collect up their total intellectual feedstock.

Systems people have been called into the scientific arena to try to overcome these difficulties. Their contribution, so far, has been remedial rather than prophylactic. It is, possibly, an absorbing evening pastime to speculate what they could suggest if they were invited to set aside the acute problems in favour of tackling the chronic ones! But, left to themselves, I would not be hopeful about the outcome. I don't doubt that it would be logical and efficient. But that little problem of 'fixed cost' would worry me.

However, I am not calling upon the information world to crusade for some total and revolutionary change in the arts and practices of publication. Change can only be evolutionary, unless there were some sudden and quite exceptional degree of agreement between nations and between benevolently committed interests. So we, like so many other crafts, must do our best with our raw materials as they arise. Their forms, of course, change slowly with time, and we shall have to adjust our own approaches accordingly. We are not likely to find the perfect solution for all time anymore than any other technology does - which is perhaps good for your employment prospects!

Since a speaker in this field must draw the line somewhere (unless he is prepared to talk all day), I am going to concentrate on what I believe to be our critical area - which is not to say that many other areas may not be equally important from the users point of view. But the critical operation - basic to all other things - is the technique of selection: selection of desired items from a collection assembled not for homogeneity, but for saleability.

We have already noted that an author's paper is not deliberately directed 'at birth' to the relatively few and scattered people who benefit from it. Some of those people can be identified, of course, and an author will often serve them on a basis of 'personal courtesy'. But his fringe readers can only be reached via the inevitable 'mixing' of papers that makes primary and secondary publication an economic possibility. Hence, the reader is in need of a simple, well-organised, cheap un-mixing or unscrambling operation. It is this unscrambling operation that tends to dominate all my thinking about today's information problems - although I know there are problems to be solved in the area of physical retrieval as well as in intellectual retrieval.

You will not have overlooked the fact that I said that it must be 'cheap', and you will fairly ask what I mean by 'cheap'. The only answer to this is 'as cheaply as the reader could do it himself, using printed indexes'. This, of course, is no answer. The situation of individuals varies widely; some have ready access to large libraries with generous indexes, some have not. Some have physical assistance in the shape of helpers, or pre-screened collections - all of which ought to be costed into any valid comparison. But information work is as poorly costed as any activity in the world. It is meagrely financed because most administrators cannot rationalise a conviction to drop it altogether or rationalise the courage to support it with measured generosity. One can cost the retrieval operation, but it is very difficult to put a value on the product. The result of all this is that information activities are not rigorously costed. So the answer to "cheap" must, at the present time, be "as cheap as possible, but without being too nasty". It must be cheap enough to attract an initial volume of use sufficient to pay the costs. From that basis it is reasonable to hope that increasing use will enable prices to be lowered and/or quality and coverage to be increased.

Cheapness may be sought in many ways - even within my prescribed operation of 'selection'. Much choice is in the hands of the local information worker, but possibly more weighty choices lie with others who operate at an earlier stage in the distribution network. Let us look again at the total picture.

With few exceptions scientific literature is organised and distributed in accordance with the traditional disciplines - and we have seen why this is so. Again, with few exceptions, the major secondary sources are similarly organised, although there is already fairly abundant cross-referencing and a good deal of overlap, duplication and omission. Manufacturing 'for the market' has promoted reasonable viability without rationality - a rather intriguing paradox!

The individual reader's needs do not closely match the packaging. As I see it, those needs are of at least two kinds viz a) philosophical nourishment: for which he will turn to almost any outstanding work within his own sub-discipline and b) facts and techniques from many disciplines and technologies - for which he will turn to anything relevant - if only he can find it. Different users need a) and b) in different proportions, but probably no user can manage solely with a) or b).

The present packages go quite a long way towards meeting (a) in several disciplines. True, the coverage is not 100%, and none are yet truly retrospective (except manually) but these are improvements that may be expected to evolve. Needs of the (b) type, however, can only be met by costly improvisation at present. Many sources must be searched or consulted to bring together what is required, ie. too many, and too large, haystacks must be forked over to find the needles.

At what point, then, in the distribution network should re-arrangement and re-packaging begin? And can it ever lead to a single re-packaging design that will provide as cheaply as possible for everybody?

It would seem at least that the 'disciplinary' and the 'interdisciplinary' have to be treated as separate user requirements of equal importance. It would be as inconvenient to extract disciplinary material from an interdisciplinary file as vice-versa. It is also questionable whether a common minimal item representation will effectively serve both purposes - although there might be no objection to a single intellectual digestion which produced various representations.

It is to be noted, in passing, that interdisciplinary needs are not co-terminous with interdisciplinary items. Interdisciplinary needs often demand disciplinary papers. Disciplinary needs often demand items from other disciplines, but less frequently demand intrinsically interdisciplinary items.

We may surmise, I think, that the contemporary disciplinary secondary publications (and data-bases) will be our dominant raw material for a long time to come. At what level, then, should the creation of interdisciplinary or mixed collections start, and by what processes? If time were no object, mixed abstract journals could be created (by copyright arrangement) out of the traditional secondaries. Searchable data-bases could then be prepared from these. But how would the proper content of the mixed secondary publications be identified? If these contents could be satisfactorily defined, there would be little need for mixed secondaries. Furthermore, in the name of cheapness, why print twice over - and months later?

No, we are driven back to making the best use we can of whatever secondaries and data-bases achieve market viability. It is then for some experts(s) to use those bases as skilfully as possible in meeting every kind of demand for item selection. We may expect, I think, that there will eventually be so many secondaries and data-bases that few enterprises who merely use information (ie. they do not trade in it) will be able to afford the resources to do their own total selection from the basic secondaries. Users will need some intermediate service that will repackage the secondaries into mixed collections of various kinds. This may well be one of the chief social functions of the emerging national information centres.

If this be the likely pattern of evolution, then, for all our well-meaning intentions, these re-packagers will themselves be thrown back against the same market influences that have constrained the activities of the primary and secondary publishers. Packages must command markets - and even with improved computer technology this repackaging will be a costly overhead that has to be spread. The interdisciplinary packages will, therefore, have to be fairly large: they will have to embrace agglomerates of industrial activities, rather than identifiable trades and processes - although all things are possible at a price!

No doubt, as in the past, these developments will start tentatively in the highly developed and so-called prosperous industrial societies where demand can (up to a point) be measured - although it might be near the truth to observe that it is often clamour that is measured rather than need. In the course of these developments, there will be the same degrees of overlap, omission and incompatibility as we have seen in the past - moderated, no doubt, by such agreements and conventions as may be adopted for machine-readable records, file formats, and so forth. But in the absence of active communications and urgent logical discourse between all concerned, events and expediency will overtake rationalisation as they have done before.

But it is very difficult to see how a trial and error approach can be avoided. Such classifications as there are of trades and industries were not, in the main, drawn up for the canalisation of technical information and they will provide little guidance towards economic and effective re-packaging. But even trial and error can be pursued in an enlightened manner. Before too many resources are committed to active repackaging it would be valuable if some public and international assessment could be made of the pattern of partitioning that is required.

So far, I have spoken of things that ought - for economic reasons - to be planned at communal level (albeit by enterprises working for the community) rather than at the level of a particular organisation. How will such things affect the economic position of the information manager who has to use them to fulfill corporate and individual needs? There will now be twice as many data-bases as there were - all of them roughly as expensive as to-day's data-bases. The poor manager seems further away than ever from feeding his flock with his paltry housekeeping allowance! This is where the so-called information centre (be it 'national' or otherwise) has a further social duty. It is now the probable custodian of both disciplinary and 'mixed' data-bases. It must have organised those bases for fast and cheap searching at levels which may extend from the information manager who needs a large searchable sub-collection regularly up-dated, down to the individual who needs a one-off retrospective search on demand. I think it may be inescapable that the larger centres will be involved in both 'wholesale' and 'retail' business for a long time to come - unless sub-centres emerge, based possibly on research associations or other co-operative bodies.

What should viably subtend from the activities of the repackaging centres must depend upon a fine interplay between the forces of classification and the forces of the market. I have postulated 'large' interdisciplinary files, to be tapped by organisations and by individuals. It will always be for continuous study what degree of sub-packaging is warranted in anticipation of a volume of smaller and individual enquirers. We have seen these considerations at work in the discipline of chemistry - without, perhaps, any very clear cut evidence whether mere partitioning of a collection is truly viable or not.

I am afraid that this has all been rather laborious. One should, perhaps, never try to rationalise one's dreams. But whilst the eventual reality may be quite unlike my prediction, it will, I am convinced, be predicated by what is considered to be the viable level and pattern of pre-arrangement and pre-partition. Costly 'arrangement' must produce an array that permits ready retrieval from all foreseeable viewpoints - otherwise it is just a theorist's arrangement. Costly pre-partition must produce a sub-set that is either used or sold to a sufficient extent - otherwise it is a theorist's separation.

The choice between a)elaborate 'labelling' of large collections and b) merging and then partitioning into non-noisy sub collections is likely to engage our attention for a long time to come.

How then should these larger issues affect the direction of development at the processing level?

In my view they seem to point further and further away from any universal and uniform technique for the handling of literature references, and more and more to the need for a variety of proved techniques appropriate to tasks and to environments. Quite apart from the very real differences between information shops - differences of size, of skills, of resources - there are intrinsic differences in the modes of sorting and sifting and screening required. One could, I think, postulate at least three recognisably distinct situations - although in practice they would be no more distinct and discrete than the 'stops' on a camera lens: and, incidentally, I tend to recall that each 'stop' has its own depth of focus. Perhaps there is a useful analogy in that!

These three situations might be described as follows:-

1) Systems to enable discipline-based collections to contribute to large interdisciplinary collections.

2) Systems to enable large interdisciplinary collections to contribute to task-oriented collections.

3) Systems to enable task-oriented collections to contribute to personal collections (notebooks).

The relative neatness of this array is however 'spoiled' by the near certainty that the personal collector will wish to be able to tap 1 and 2 directly as well as through 3. This puts a considerable strain upon the design of 1 and 2 and imposes a requirement for a kind of double-standard, as we shall see.

The above three process contributions towards effective discrimination for, and collection at, the individual level have each to make their own decisions in at least one fundamental area viz:- the mode of item representation. Their subsequent processing operations will be constrained - but not determined in detail - by that decision. Mode of item representation is, possibly, the most debatable and sensitive area in the arts of information for we all know - and we are inhibited by the thought - that much work may be done on an item that is never retrieved.

To a great extent, situation 1 is having its fate determined for it. The discipline-based data-bases all have a conscientiously prepared abstract somewhere in the background, although it may or may not appear on the searchable data-base. The pattern seems to be to compile a searchable record from the title, the bibliographic detail, a few keywords or 'high-spots' (eg. chemical compounds) and a very rough 'classification' into classical sub-areas of the discipline. Most of these characterisers are the inevitable spin-off from the processes of publication and therefore represent a relatively low overhead on the retrieval process. Any attempt to add de novo 'retrievability' will add a specific and considerable overhead no matter whether this attempt be in the direction of adding the whole abstract (free text approach) or of more exhaustive keywording (co-ordinate indexing). The paradox is that the <u>individual</u> searcher would benefit by such additions, but cannot afford them, whereas the situation 2 information shop could afford them, but probably does not need them. I would surmise, then, that situation 1 will settle down with an embellished title and a 'free-text' search technique - with, no doubt, fairly sophisticated aids to assist semantic rapport with the 'free-text'. This should be quite adequate for the selection of situation 2 collections, and has already proved useable, if not excellent, for the individual searcher.

In situation 2, several demanding requirements come into play. In the first place the situation 2 collection will be a 'merge' of items picked up via word-lists and concepts that are valid currency in situation 1 but which are only incompletely articulate in the ears of the situation 2 user. Also, the several situation 1 glossaries will not be consistent or uniform amongst themselves (eg. bridge-chemistry, bridge-physics, bridge-engineering). The intended concepts may be clearly distinguishable by readers, but they will not be distinguishable by computers. Furthermore, none of these may be in the terms that the situation 2 enquirer would himself customarily use for those concepts: he may have his own vernacular use for the word 'bridge' (eg. bridge - dentistry, bridge - banking) and prefer some more specific descriptor for a concept that is outside his daily discourse.

It seems unlikely therefore that situation 2 can use the situation 1 representation for anything more than the gathering of its input. When it comes to output, it needs a representation that is more compatible with the discourse language of its user community. This might mean a) specific re-analysis of the items being collected or b) intense organisation and correlation of the 'inherited' keywords with the language of the user community. One feels intuitively that b) alone will not suffice; all interdisciplinary activities breed a convenient jargon which ought to find direct representation in its search glossary. Even if the total disciplinary abstract were carried through to the situation 2 collection, I feel that it might not serve the situation 2 searcher. It would depend, I suppose, upon whether he conceded that he was searching outside his inter-discipline. I am inclined to believe, therefore, that the merged cross-collection is going to require some supplementary intellectual examination of its constituent items. And since, by definition, the conceptual elements looked for will not be explicit in the situation 1 representation, there is going to be a choice between a) interpretive additions in textual form or b) additional access points in the shape of parochial keywords. Since the purely technical choice between these is still an open one, we are likely to see both approaches.

In situation 3, we are catering for the ultimate user - a person who is probably familiar with the terminology of one of the constituent disciplines, but is no more than acquainted with other disciplinary terminologies. And on top of these he will have grown into a craft-terminology - so that he accepts the malpractice of calling a radio receiver a 'transistor' even though he knows that this is semantic anarchy.

Here we seem entitled to cash in on the thought that the essential glossary of a truly common endeavour can be quite small and yet highly specific. We have reached the stage where the purpose of the terms or keywords is little more than parochially mnemonic: they do not have to describe themselves to the world outside. Here then may be the proper field for the controlled (but not static) vocabulary and thesaurus. Here is a situation where the essential concepts of the user community may be identified and coded in a notation that is free from the confusion brought in by twisting (however benevolently) the basic meaning of words. Who need care if the concept code "347" can only be explained by the use of 20-30 words? It has only to be done once: it is not used in that form and will not therefore give rise to the false drops that might be associated with the corresponding free-text term (eg. "quartz iodine lamp") Situation 3 seems to me to be the natural hunting ground for the thesaural enthusiast. The community is small enough to be identified and consulted; its specific vocabulary is small enough to be articulated and coded.

Where does all this lead, if anywhere? Possibly to the comforting conclusion that there is so much room for versatility in improvement and innovation that no one working in information science need feel that he is wasting his time. But that view would be altogether too cosy! I have tried to paint a picture of the progressive distribution of items through processing phases that I have called 'situations'. The important thing is, I think, to understand which of these situations one is working in, or is nearest to. To become deeply committed to the techniques of the wrong situation is to waste ones own time - even though one may solve someone else's problem. There is little comfort in the latter thought however, for the retrieval of information about information retrieval is so poor that the solution will probably not reach the problem!

Consistency is, of course, the refuge of fools, but I like to think that my current thinking has proceeded from my observation of many years ago that "a thesaurus is for a community: not for God!" Communities are of many sizes and many kinds, but the truth still remains that they discourse and they question in a communal vernacular that they have evolved because of its precision and economy for them. The search system need not answer back in the vernacular: but it must be able to respond to a vernacular enquiry.

For all its omissions, commissions and archaeological stratifications, we might do worse than re-appraise UDC and vote it the funds to put new flesh on to its excellent bone-structure: and then keep its tissues properly restored.

A COST-EFFECTIVENESS ANALYSIS OF SEE-REFERENCE STRUCTURE IN DIRECTORIES*

Manfred Kochen
Professor of Information Sciences, Urban/Regional Planning
and Research Mathematician
Mental Health Research Institute
University of Michigan

SYNOPSIS

Is there a most cost-effective number of index terms in an index which point, not to the items they index, but to other index terms? We found no optimal number between the maximum the language will allow at one extreme and no cross-references at the other, for the models we employed. The analysis was as valuable for what it adds to techniques of mathematical analysis in information science as for the insights it gave us into some fundamental properties of an index.

1. Introduction

Modern libraries, universities, research centers, government agencies and other service organizations are becoming more complex. The size of large libraries, for example, has been doubling every 16 years during the past few decades, and the number of services they provide increased too. This complexity may be increasing faster than does our ability to cope with it. At least two factors contribute to organizational complexity:
(1) great variety in the services performed; and
(2) the sheer number of services, agents, resources, and organizational component
Important tools for coping with both size and variety are directories that reflect (or create) order and organization among the organizational components, and they guide users according to this order.

An example is the telephone directory. It serves as a directory to the various resources in a city, much as a university or library catalog serves to direct its users to its resources. The "classified" directory (Yellow Pages, subject catalog has greater organizing power than a name-address directory if the number of servic types (variety; e.g. plumbers, physicians, etc.) is less than the number of servic tokens (e.g., United Plumbing, All-Plumb, D.G. Gish, M.D., A. Jones, M.D., etc.). Consider another, hypothetical, example of a directory to medical cases by patient name with a directory by syndromes. There are, in general, fewer syndromes than

* The collaboration and help of R. Tagliacozzo and P. Roosen-Runge in the early discussions of this work contributed to some of the ideas in this paper.

cases. It is, of course, not necessary that the entries in a directory form a classification system in the rigorous sense: a subject-heading or keyword list also provides organizing power.

We shall confine our analysis only to such subject-directories because of their importance in prior work toward new types of information services. We shall focus on the benefits and costs of using indirect references, or cross-reference pointers from one subject term to another term in the same directory because of the additional organization such a structure is believed to provide. In this work, we build upon earlier results about cross-reference structures (2). We confine it to only one kind of cross-reference, the "see"-reference, because of its widespread and long-standing use and the simplification it brings to our analysis.

While there are published discussions (5)(4) prescribing how see-references should be constructed, there are to my knowledge, no objective analyses of the benefits and costs involved. There has been a tendency on the part of catalogers, indexers, authority-structure compilers, and even system planners to perpetuate time-honored practices without critical appraisal and examination. At the same time the penalties resulting from insufficiently rationalized systems has greatly increased with the use of more powerful new technologies for storage, retrieval, and search.

It is not atypical for a computerized directory to be viewed merely as a "conversion" of its "manual" predecessor. Existing catalog structures and cataloging rules are thus frequently perpetuated, perhaps to the detriment of their users. This danger arises from failure to distinguish between the function of a directory and the means for implementing that function, and also, from failure to realize fully the increase in the number of alternatives for implementation afforded by the flexibility of modern information technology.

Mathematical models of various means of designing a see-reference structure, though over-simplified and not corresponding to actual implementations, are of great value in eliminating the failures mentioned above. They focus on the most essential relevant variables. They provide insight into the logical structure of the concepts underlying the idea of a see-reference structure, the design parameters, the key performance variables used to evaluate a design, and the connections among these.

2. Formulation of Mathematical Problems

Visualize a directory as a table with a partition down through it. The left part is a vertical list, ordered in some way, such as lexicographic, of what we shall call index terms. Each index term is a possible access point to the directory. The right half of the directory is a list of what we shall call profiles, one associated with each index term. Each profile is, in turn, a list of terms. It may be ordered or not. Each term may be an index term or an address. By an address we mean a unique identifier or label for a record, resource, agent, facility, etc. to which the directory ultimately directs its users.

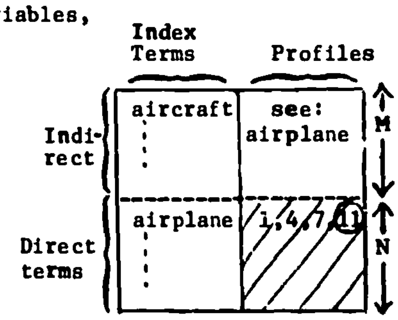
We now partition the index terms into direct index terms, which point only to addresses, and indirect index terms, which point, via a see-reference, to other index terms. Why should there be any indirect index terms in a directory at all? If the total number of index terms is fixed, why should not every term be direct? Suppose, for example, that the direct index term "airplane" points to document labels 1, 4, 7,

11. Suppose that "aircraft" is another of the fixed number of index terms. We could make it indirect by having it point to "see: airplane", or direct by having it, redundantly, point to 1, 4, 7, 11. Doing the latter leads the user more quickly to the needed source than going through a see-reference, and uses up only 10 characters of storage (counting spaces and commas) vs. 13 characters for the see-reference, counting "see: ".

The see-references give the directory a logical structure that could be of educational value. But, in this paper, we shall ignore this potential benefit and focus only on the benefits from decreased storage and updating time, less the costs of increased search time.

To analyze this problem, define the following basic variables, interpreted in figure 1.

N: number of direct index terms, or rows in the bottom half of figure 1 ($\sim 10^4$).
M: number of indirect index terms, to be found.
N+M: total number of index terms.
n_d: average number of items per direct index term.
d: average number of direct index terms per document (~ 10).
D: total number of documents or items to which the direct terms can point ($\sim 10^5$).



Directory: Figure 1

If d_j is the number of index terms assigned to the j^{th} of the D documents, then $\sum_{j=1}^{D} d_j$ is the total number of term-item pairs or the total number of items' id. numbers entered in the lower right (shaded) quadrant of Figure 1. Then $\frac{1}{D}\sum_{j=1}^{D} d_j$ is the average number of item id. numbers in one of the N rows; this is what we have called d. But $\sum_{j=1}^{D} d_j$ is also N times the average number of items per direct index term. Hence, $dD = \sum_{j=1}^{D} d_j = Nn_d$, or $dD = Nn_d$.

To estimate the total cost of storing the directory of figure 1, let
c_s: the cost, in \$/bit/month ($\sim$.00005) (or 250 millicents/bit, purchase)
b_d: average nr. of bits per item identification number, $> \log_2 D$ (~ 20)
b_i: average nr. of bits to store an index term, direct or indirect (~ 200)
r : average nr. of direct index terms per indirect term. (~ 1)

We assume here that no indirect term points to another indirect term. It is not difficult to drop this assumption; this leads to interesting conditions for being trapped in a run-around cycle, but it diverts from our main line of argument. It follows that if r_j is the number of direct index terms to which the j^{th} indirect term points, then $\sum_{j=1}^{M} r_j$ is the total number of pairs of the form: (indirect term $\rightarrow$ direct term). Clearly, $r = \frac{1}{M}\sum_{j=1}^{M} r_j$. (Also, $\sum_{j=1}^{M} r_j$ is the average number of indirect terms pointing to one of the direct forms $\times$ N; the average number of indirect terms per term is thus $\Sigma\, r_j/N = \frac{rM}{N}$).

To store the left "half" of the directory in figure 1 requires $b_i N$ bits. To store the upper right quadrant of $\sum_{j=1}^{M} r_j$ direct index terms requires $b_i(rM)$ bits. To store the lower right quadrant of $\sum_{j=1}^{D} d_j$ item id. nrs. requires $b_d N n_d$ bits. The total storage cost, C_s in \$/month, is:

$$C_s = c_s(b_i(N+M) + b_i rM + n_d b_d N) = c_s(b_i(N + (r+1)M) + dDb_d) \qquad \text{Eq.(2)}$$

Evidently, increasing M can never result in a saving of storage costs.

To estimate the cost of updating the directory, let:

D' : the number of new documents (item id. nrs.) per month at which the collection grows (1% of D or $\sim 10^3$)

τ : average time, in seconds, to access and compare a randomly chosen index term of the directory with a given search term. ($\sim 10^{-5}$)

We now assume that a new document is indexed with d (on the average) of the N direct index terms already in the directory, and that the list of N terms is ordered, say lexicographically, so that it can be searched for the given search term in a binary fashion. The average number of necessary comparisons is, then, $\log_2 N$. It will take $\tau \log_2 N$ seconds to find a direct term in the directory if it is there. Suppose next that the n_d item identification numbers posted next to a direct index term are recorded in the order in which they arrived. Let

t_w : time, in seconds, to record (write) an item id. nr., and ($\sim 10^{-5}$)

c_t : cost, in \$/second, to write or read into or out of the stored file. ($\sim .05$)

The average updating cost, C_u , in \$/month, is thus:

$$C_u = c_t \cdot D'd(\tau \log_2 N + t_w) \qquad \text{Eq.(3)}$$

Next, let:

p_i : the probability that a randomly chosen search term is one of the M indirect index terms. Assume it to be kM, where $k < \frac{1}{M+N}$.

p_d : the probability that a randomly chosen search term is one of the N direct search terms. Assume it to be kN.

p_o : the probability that a randomly chosen search term isn't an index term in the directory at all $= 1 - p_i - p_d = 1 - k(N+M)$.

The average time to search a (lexicographically) ordered directory for the presence or absence of a specified search term, direct or indirect, is $\tau \log_2 N$ seconds. This is the total time per query if the search term isn't listed, an event of probability p_o. <u>We assume that each query is specified by exactly one search term.</u> If the search term is listed, and is one of the direct terms - an event of probability p_d - then the total time per query is $\tau \log_2(M+N) + n_d t_w$ seconds because n_d items have to be retrieved and the n_d id. nrs. have to be temporarily stored. If the search term is one of the indirect terms, the total time per query is $\tau \log_2(M+N)$ plus τ times the total time for each direct term this indirect term points to. This is $\tau \log_2(M+N) + r(\tau \log_2(M+N) + n_d t_w)$ seconds per query.

The average total read-write time per query is about:

$$p_o \tau \log_2(M+N) + p_d(\tau \log_2(M+N) + n_d t_w) + p_i(\tau \log_2(M+N) + r(\tau \log_2(M+N) + n_d t_w))$$

seconds per query. The query-processing cost, due to read-write operations, in \$/month, is with the formulas for n_d and for p_o, p_i and p_d and with R = the number of queries/month, each consisting of just one search term:

$$C_p = c_t R \quad (1-k(M+N))\,\tau \log_2(M+N) + kN\,\tau \log_2(M+N) + kdDt_w$$

$$+ kM(\tau \log_2(M+N) + r(\tau \log_2(M+N) + \frac{dD}{N} t_w)) \qquad \text{Eq.(4)}$$

$$= c_t R \quad \tau \log_2(M+N) + kdDt_w + kMr(\tau \log_2(M+N) + \frac{dD}{N} t_w) \quad .$$

$$C = C_s + C_u + C_p$$

$$\doteq c_s \quad b_i(N + (r+1)M) + dDb_d$$

$$+ c_t D' d(\tau \log_2(M+N) + t_w) \qquad \text{Eq.(5)}$$

$$+ c_t R \quad (1 + kMr)(\log_2(M+N)) + kdDt_w(1 + \frac{Mr}{N})$$

3. When Does It Pay to Repeat the Reference Profile?

By the reference profile of a direct index term t we mean the list of item identification numbers to which t points. Since the average list is n_d or $\frac{dD}{N}$ items long, with b_d bits per items, it will require $\frac{dD}{N} b_d$ bits per direct term.

An indirect term points to r, usually r=1, direct term and requires only rb_i bits of storage. If we were to repeat the reference profile under an indirect term rather than using a see-reference, we would require an additional $\frac{dD}{N} b_d - rb_i$ bits per indirect term. Altogether, $M(\frac{dD}{N} b_d - rb_i)$ more bits would be required.

In updating, an id. nr. has to be posted on an additional $\frac{rM}{N}$ indirect terms per direct term, making the total number of id. nrs. to be written $d + \frac{rM}{N}$ rather than d id. nrs. per new document. This takes an additional $\frac{rM}{N}(\tau \log_2(M+N) + t_w)$ seconds per new document.

So far, repeating the document id. nrs. under what could be indirect index terms has incurred an added cost of $c_s M(\frac{dD}{N} b_d - rb_i) + c_t D' \frac{rM}{N} (\tau \log_2(M+N) + t_w)$ dollars per month. What is gained? It should be possible to save time in searching. The search time is now just $(1-p_o)\tau \log_2(M+N)$ seconds per query, where p_o is the probability that the term isn't in the index at all. We assume, as before, that $p_o + p_i + p_d = 1$ and $p_i = kM$, $p_d = kN$, so that $1 - p_o = k(M+N)$. This saves $\tau \log_2(M+N) + kMr\tau \log_2(M+N) + kdDt_w(1 + \frac{Mr}{N}) - k\tau M\log_2(M+N) - k\tau N\log_2(M+N)$ or $kM\tau(r-1)\tau \log(M+N) + \tau(1-kN)\log_2(M+N) + kdDt_w(1 + \frac{Mr}{N})$ seconds per query or $C_t R\, \tau(kM(r-1) + (1-kN))\log_2 (M+N) + kdDt_w(1 + \frac{Mr}{N})$ \$/month.

We can now ask when it pays to repeat the reference profile in the indirect terms. This is to ask when the savings in search cost exceeds the increased storage and updating costs. Let us simplify by first taking r=1. The search savings are:

$$S_s = c_t R \left[(1-kN) \log_2(M+N) + kdDt_w(1 + M/N) \right] \quad \$/\text{month}.$$

The extra costs are: $C' = c_s M(\frac{dD}{N} b_d - b_i) + c_t D' \frac{M}{N} (\tau \log_2(M+N) + t_w)$ \$/month. We simplify further by fixing values for some of the variables. Suppose that $N=10^4$ direct index terms, M=4N indirect index terms, d=10 index terms/document, $D=10^5$ documents, $k=10^{-5}$, so that $p_o = k(M+N) = 5Nk = 5 \cdot 10^4 \cdot 10^{-5} = .05$, or 1 chance in 20 of hitting no index term at all. $D' = 10^3$ documents/month or 1% of D. $b_d = 20$ bits/doc. id. nr. $b_i = 200$ bits/index term. We next consider two cases, corresponding to a I computerized and II a manual system.

	CASE I		CASE II
c_t (\$/second)	.05	computer time	.001 for human labor at \$4/hr.
c_s (\$/bit/month)	5×10^{-5}	disc storage	10^{-8} for book storage in libraries
τ (seconds to match)	10^{-5}	CPU compare time	1 human compare time
t_w (seconds to write)	10^{-5}	CPU write time	2 human write time
S_s (\$/month)	$10^{-5}R$		.023 R
C' (\$/month)	4000		61
Least value of R for which $S_s \geq C'$	4×10^8	requests/month	2650 req./mo.

Neither case is very plausible, because even 2650 requests per month are too high.

Let us repeat the above calculation with only M changed. Suppose R=1000 and M=aN, with a to be determined such that $S_s = C'$. Case I: $S_s \doteq .2a + a(15+\log_2(1+a))$

$C' = \log_2(1+a) + 15 + 2a$. A value of a slightly exceeding 1 satisfies this condition.

4. Benefit-Cost Maximization

Let v be the utility of obtaining a match in the directory for a given query. The expected benefit, in dollars per month, for successful uses of the directory, is $R(1-p_o)v$, or $Rk(M+N)v$, with the assumptions made before. If we let $M=aN$, the net utility, or benefit less cost is:

$$\begin{aligned} U = {} & RkN(1+a)v - c_s\, b_i N(1+a(r+1)) + dDb_d \\ & - c_t D'd(\tau \log_2 N(1+a) + t_w) \\ & - c_t R\,(1+akrN)(\tau \log_2 N(1+a)) + kdDt_w(1+ar) \end{aligned} \qquad \text{Eq. (6)}$$

Suppose first that all variables except a are specified as in Cases I and II of the previous section. Which value of a maximizes U? Assuming U to be a differentiable function of a, we find a such that $dU/da=0$. We have:

$$dU/da = \frac{A}{1+a} - B \log_2(1+a) + C, \text{ where, for} \qquad \text{Eq. (7)}$$

case I, with R=1000, and v unspecified, we have:

$$A = -\frac{.005}{\ell n2} \qquad B = 5 \times 10^{-6} \qquad C = 10v - 200$$

where

$$A = \frac{c_t\tau}{\ell n2}\, R(krN-1) - D'd$$

$$B = c_t \tau RkrN$$

$$C = RkrN\left(v - \frac{c_t\tau}{\ell n2}\right) + kdDrt_w - N(c_s b_i(r+1) + c_t Rk\tau r \log_2 N)$$

If a is a small positive number, then $\log_2(1+a) \doteq a$ and the middle term is negligible; the value of a which makes $-\frac{.005}{\ell n2} \cdot \frac{1}{1+a} = 200 - 10v$ is then $a \doteq \frac{.007}{10v - 200} - 1$; this is positive only if $.007 > 10v - 200 > 0$, which implies that v is between 20 and 20.0007. Thus, a depends with very great sensitivity on the precise value of v in this very narrow range. This is a typical result in this kind of optimization, and casts doubt on the model generating Eq.(6). It is, of course, unrealistic to assume that the value of a successful search could be known with such precision. Does this mean that there is great inherent uncertainty in the number of M of indirect references to use?

Before we conclude this let us examine what happens if a is large. How large? Large enough for $\log_2(1+a)$ to equal $200 - 10v$ unless again we force v to be about 20.

For $\log_2(1+a)$ to be, say 10, a has to be 1000, and a thousand indirect terms for each direct term is quite unrealistic. So the above conclusion holds. What, then, is wrong?

Let us try numerical values for case II and R=1000. Now A = −16 B = .01 $C \doteq 10v + 2$. A value of a for which $\frac{16}{1+a} + .01 \log_2(1+a) = 10v + 2$ must be quite large if v is large for any v. Even if v=0, <u>a</u> must be at least 7, for then $\frac{16}{1+7} + .01 \log_2 8 = 2.03$ vs. 2. This is more reasonable, though an upper limit of 4 for a would be more realistic since there may not be many more than 4 words in English that could serve as indirect references to the same see-reference.

We could, of course, conclude that it is never cost-effective to use see-references. To explore this, we should not regard N as fixed. The larger N and M, the greater the probability of success in searching, which varies as k(N+M). The expected penalty of failure is −k(N+M)vR \$/month. If this exceeds the costs, then it should be desirable to make both N and M as large as possible. Let us ask which values of M and N jointly maximize U(N,M) = k(M+N)vR − C. We thus seek M and N such that $\frac{\partial U}{\partial M} = 0$ and $\frac{\partial U}{\partial N} = 0$. We have: $\frac{\partial U}{\partial M} = -A \cdot \frac{1}{M+N} - B \cdot \frac{M}{M+N} - B' \log_2(M+N) - D \cdot \frac{1}{N} + C$ and

$\frac{\partial U}{\partial N} = -A \cdot \frac{1}{M+N} - B \frac{M}{M+N} + D \frac{M}{N^2} + C'$, where $A = \frac{c_t \tau}{\ln 2}(D'd+R)$, $B = \frac{c_t R \tau k r}{\ln 2}$

$B' = B \ln 2$ $D = c_t RkdDt_w r$ $C = kvR - c_s b_i(r+1)$ $C' = C + c_s b_i r$. To simplify, let us estimate the values of the constants for case I

$$\frac{.007}{M+N} + 7 \times 10^{-10} \frac{M}{M+N} + 5 \times 10^{-10} \log_2(M+N) + \frac{.0005}{N} = \frac{v}{1000} - .02$$

$$\frac{.007}{M+N} + 7 \times 10^{-10} \frac{M}{M+N} - .0005 \frac{M}{N^2} = \frac{v}{1000} - .01$$

If we subtract the second equation from the first, we have

$$5 \times 10^{-10} \log_2(M+N) + \frac{.0005}{N} + .0005 \frac{M}{N^2} = -.01 .$$

Since M, M have to be positive, this condition cannot be met.

Note that if kvR is large while $c_s b_i$ and $c_t Rk$ are small, increasing M and M will increase U. We can thus defend the following general conclusion: If the value of a successful query and the number of requests/month are sufficiently high, then as many direct and indirect index terms as possible should be used. The ratio of indirect to direct terms should be the maximum the language allows. This maximizes the benefit-cost difference.

5. Conclusions

We can learn several lessons from the preceding analyses. Before discussing them, let us reflect about their nature. They reflect a kind of thinking which is not widely practiced by designers and analysts of libraries and information centers. It is mathematical thinking. It follows in the tradition of Morse (3) and others who have done operations research in this area. If more information systems analysts would embrace this kind of thinking, information science would perhaps advance much more rapidly and at a higher level.

A. Lessons for the Information Scientist.

1. There is still a great need for precise conceptualization, for explicating basic notions underlying most of information science. We can view the problem of this chapter as perhaps one of the most elementary ones in the field. Yet, we have not really solved it in a satisfactory way. The analyses are more instructive in showing us what we still do not know than in applying what we know. We do not know, for example:

   (a) how to measure the benefits of using a directory. A directory to be used to find an ambulance service or to the fire department at a moment of crisis, can be quite valuable. In design, we should obviously make sure that every conceivable term under which any directory user might look for an ambulance is entered - and we would be far wiser to repeat the document id. nr. (or telephone number) under each variant than to use "see-references." But getting this obvious conclusion to be a rigorous implication of a model requires a mathematically more sophisticated model than the one we have analyzed here.

   (b) the user's time and effort needs to be taken into account. To the cost should be added the wasted time of the user in chasing see-references. This could easily be done by a simple extension of the present analysis, by simply adding the cost of the user's time. But disutility does not necessarily increase linearly with cost. If the user's wasted time exceeds a certain threshold, he will not bother at all to pursue a see-reference. This is especially true if search time varies linearly rather logarithmically with file size.

   (c) While the assumption that the directory entries can be maintained in order so that search time varies as the logarithm of file size, there are advantages to hashing (1), where search time is essentially constant for all file sizes up to a limit. It is simple to repeat our analysis under this condition, but this requires knowledge of bounds on M+N, the very quantity we wish to optimize. Perhaps a fresh approach to the entire formulation of the optimization problem - possibly as an application of dynamic programming - would prove to be fruitful.

   (d) The numerous assumptions in our model can be viewed as hypotheses for empirical test. Does the probability of a randomly chosen query matching one of the N direct index terms really increase as kN; and how can we estimate k so that the probability of no match at all is 1-k(M+N) ?

2. There are numerous easy extensions of the analyses presented here which give us experience and practice with such analytical work. Two examples follow:

   (a) Extension to the case where a query involves Boolean or more complex logical-syntactic combinations of index terms in the directory.

   (b) Extension to the case of partial matches between query terms and index terms, both for a complex query expression as a whole and when the query term is a single term.

B. Lessons for the Professional Information Officer.

With those responsible for the practical operation and improvement of the services of a library of information center, we can only share concern that the problems they face are intellectually challenging and difficult. It would be misleading to suggest that they can already apply our results to any practical problems of directory design. It will take considerably more basic research before this is the case. In the meantime they are well advised to continue using the rules of thumb that indicate that they are not likely to make mistakes of grave consequence. We strongly urge, however, that they begin to develop an appreciation for the kind of analyses made here and those likely to follow in this path. Most necessary is their support of the need for basic analytical research and possibly their collaboration to keep the models in better touch with the most essential features of reality.

References

1. Peterson, W.W. "Addressing for Random-Access Storage" IBM Journal of Research and Development 2, No. 1, April 1957.

2. Kochen, M. "A Study of Cross-Referencing" with R. Tagliacozzo. Journal of Documentation 24, (3), September 1968, pp. 173-191.

3. Morse, P. Evaluating Library Effectiveness, Cambridge, Mass., M.I.T. Press, 1969.

4. Mandersloot, G.B., Douglas, M.B., and Spicer, N. "Thesaurus Control - The Selection, Grouping, and Cross-Referencing of Terms for Inclusion in a Coordinate Index Word List" Journal of the ASIS , Vol.21, No.1, January-February 1970, pp. 49-57.

5. Jahoda, G. "Controlled Index Vocabularies" Information Storage & Retrieval Systems for Individual Researchers , Wiley-Interscience/ John Wiley & Sons, Inc., 1970, pp. 27-35.

SOME FACTORS RELATING TO THE ECONOMIC BASIS OF INTERNATIONAL SYSTEMS FOR INFORMATION STORAGE AND RETRIEVAL

Abraham I. Lebowitz
Acting Director, Elyachar Library,
Technion-Israel Institute of Technology.

SYNOPSIS

The scientific and technical information mechanisms that are established to serve a given country or region must be scaled to enable them to acquire, process and disseminate an increasingly high percentage of the world's output of scientific knowledge. Local systems will have to be based on international systems for information transfer whose costs would be allocated in terms of a country's GNP and/or the amount of research actually performed.

Scientific and technological information is a necessary input to the process of research and development in any country. The information required can be generated by means of local research and development programs (themselves, of course, partially dependent on external information inputs), or can be transferred from other countries. This latter process, that of technology transfer, has been defined by Bar-Zakay (3) as occurring "when scientific and technological information generated and/or used in one context is re-evaluated and/or implemented in another context."

Insofar as it is dependent on locally performed research and development the technological gap between the developed and underdeveloped countries has been increasing. This point is well illustrated by data in an OECD study (1) published in 1967 and covering the year 1963/64. The OECD data is useful bebause the OECD member countries provide a fairly good sample of the kinds of countries in the world today. The sample includes a highly industrialized superpower, the United States, large industrialized countries such as West Germany and Japan, small industrialized countries such as the Netherlands, and large and small developing countries such as Spain and Greece.

IV Economics of Information Systems

Table I shows the gross national expenditure on research and development, exclusive of space and defense, for the OECD member countries, as well as on a per capita basis, as a percentage of the GNP, and as a percentage of the total OECD spending for R & D. Those countries most seriously underdeveloped, Greece, Ireland, Portugal, Spain and Turkey with more than 30 per cent of all employment in agriculture also had the lowest spending for R & D. To the extent, then, that R & D influences economic growth we must conclude that the rich grow richer and the poor grow poorer.

Country	Gross Expenditure (in Millions of US Dollars)	Per Capita (in US Dollars)	As Percentage of GNP	As Percentage of OECD Total
United States	9500	49.5	1.5	60.8
United Kingdom	1460	26.6	1.5	9.3
West Germany	1305	22.4	1.3	8.4
France	1000	19.5	1.2	6.4
Japan	892	9.3	1.4	5.7
Canada	357	18.9	0.9	2.3
Netherlands	314	25.8	1.8	2.0
Italy	274	5.4	0.6	1.7
Sweden	239	31.2	1.4	1.5
Belgium	132	14.1	1.0	0.8
Norway	39	10.7	0.7	0.3
Spain	29	1.0	0.2	0.2
Turkey	27	0.9	0.4	0.2
Austria	23	3.2	0.3	0.1
Ireland	10	3.5	0.5	0.1
Portugal	9	1.0	0.2	0.1
Greece	8	0.9	0.2	0.1
TOTAL	15616			100.0

TABLE I

National Expenditures on R & D (excluding space and defense) for 1963/64 by OECD Member Countries.

Inspection of Table I also indicates that with the exception of the United States, no more than 10 per cent of all OECD research is performed in any one country. The percentages would, of course, be much lower if computed on a world-wide basis.

Technology transfer provides an alternative to local research and development. This process is usually thought of as a bilateral transfer of information based on a specific contractual or quasi-contractual (e.g. foreign aid) relationship between the parties. The process can be initiated by either the donor or donee and is usually limited in subject scope and time.

Another method of effecting the technology transfer process is the importation of scientific and technological information in bulk and utilizing it as required. By information in bulk form is meant that information which exists as primary and secondary scientific and technological publications including, but not limited to, journal literature, technical reports, patents, conference proceedings, monographs, and indexing and abstracting services. Both local R & D and specific technology transfer agreements are important and have been extensively studied and applied. However, the utilization of scientific and technological literature as an important factor in national development has, despite its potential, been relatively neglected. Among the reasons for this neglect we can cite the magnitude, complexity and cost of the required systems. It is, for example, relatively simple for an international organization to secure the services of an expert and send him to assist in the solution of a specific problem in a developing country. It is quite another thing for the developing country to set up and operate a large scale information system capable of handling a fairly high percentage of the world's scientific publications. Such a system would be out of proportion to the actual utilization that could be made of the information even though it would provide assistance in the solution of a broad range of problems.

We may define our objectives as ensuring the free flow of the information reported in the scientific and technological literature across national boundaries and providing for its effective utilization within individual countries. Realization of these objectives implies the existence of both national and international information systems whether or not they are formally recognized.

A national information system in science and technology must fulfill three functions:

a) it must record the results of the research and development performed within the country. This is performed by the 'primary publication' subsystem.

b) It must be capable of identifying, within acceptable time limits, those documents, no matter where published, which are required to support the country's research, development, and educational programs in science and technology. It is desirable that this 'secondary information' subsystem be capable of providing abstracts of the documents identified.

c) It must be able to provide any required document, no matter where published, in a form in which it can be conveniently used and within acceptable time limits. This 'document delivery' subsystem should include reproduction and translation services.

It is obvious that these functions overlap and interact in various ways and that given national information systems will implement them differently.

In order to fulfill the first function each country needs a system of primary publication which is approximately proportional to the research performed in that country. That such systems do, in fact, exist needs no demonstration and that they are actually proportional to the amount of research is partly corroborated by data from a recent study by Bourne (2) of the characteristics of coverage of the Bibliography of Agriculture (B of A). A random sample of the citations in the 1967 B of A was analyzed to determine, among other things, the country of publication of the bibliographic unit cited. A total of eighty different countries, including all members of the OECD, were identified in the sample. Table II compares the expenditure for R & D as a percentage of total OECD spending for R & D with agricultural publication by country as a percentage of OECD publications in agriculture as reflected by the B of A. Despite an interval of several years between the two sets of data, a rank order correlation, R = .947 (P = .001) is obtained. This data is only indicative, however, because no evidence has been adduced to show that a) the total amount of agricultural research varies linearly with the total amount of research and b) that the B of A is unbiased in its selection of material. In fact, the proportion of research devoted to agriculture decreases with the increasing development of a country. This is borne out by the data in Table II which indicate that, except for the United States, Japan and Turkey the percentage of agricultural publications exceeds the general R & D percentage by as much as 11:1. The anomalous cases of Japan and Turkey may be the result of language problems at the B of A.

Country	R & D	Publications
United States	60.8	42.7
United Kingdom	9.3	11.6
West Germany	8.4	15.2
France	6.4	6.8
Japan	5.7	3.0
Canada	2.3	3.3
Netherlands	2.0	3.5
Italy	1.7	4.5
Sweden	1.5	2.6
Belgium	0.8	1.5
Norway	0.3	1.9
Spain	0.2	1.3
Turkey	0.2	0.1
Austria	0.1	1.1
Ireland	0.1	0.4
Portugal	0.1	0.3
Greece	0.1	0.2

TABLE II

Expenditure for R&D by country as percentage of total OECD R&D and Agricultural Publications by country as percentage of total OECD Agricultural Publications.

There is little doubt that the B of A did not provide an unbiased representation of the world's agricultural publication. OECD member countries were represented in Bourne's study by 63 per cent of the sample. This probably implies a bias in favor of OECD at the expense of COMECON and other countries. The bias does not seem to extend to the United States which contributed only 26.9 per cent of the sample.

While the cost of the primary publication system is a relatively simple function of the quantity of research performed within the country the cost of the secondary information and document delivery sybsystems are influenced by many external as well as internal factors. The most important of these factors is the total amount of research being reported within the fields in which the given country maintains active R & D programs. The value of this factor is rapidly approaching a constant for all countries, principally as the result of the increasingly inter-disciplinary nature of most scientific fields. Thus, at the same time that each country, developed or undeveloped, is performing a diminishing percentage of all research performed, the secondary information systems designed to serve it must be capable of handling the ever larger quantities of information being generated. This problem began to affect the larger, highly-developed countries, several years ago. Their response has been two-fold: On the one hand they are making increased use of automated methods, sometimes tending to phase out conventional secondary publications in the process. The implementation of this solution, whatever its effects may be within the country adopting it, generally has adverse effects on the smaller countries which have tended to rely on the bibliographic services provided by the larger, more developed countries. Secondly, the larger countries are trying to take advantage of the economies of scale by developing document storage and delivery networks which will optimize document utilization by maximizing coverage while minimizing redundancy.

Another factor which militates against adequate information service in the underdeveloped countries is the unfavorable ratio of the prices of books, periodicals, and packaged information services to local salaries. Because the cost of these services includes the high labor costs prevalent in the more highly developed countries, the information user in a less developed country is at a serious disadvantage. This situation can be further aggravated where currency and exchange problems are involved.

For the past five years or so it has been recognized that the only long term solution available, even to the developed countries, lies in the emerging international systems. The technical aspects of such systems have been explored in great detail. The economic aspects also require intensive, in-depth analysis. If the intent is to maximize technology transfer and to encourage the broadest possible participation in international systems, the complete spectrum of benefits must be made available to each member country whatever its actual contribution to the systems' operating costs. One possible arrangement would be tying the 'membership fee' to the GNP. However, countries with a low GNP per capita not only spend less for science in absolute terms but in relative terms as well (see Table I). They would, therefore, probably consider participation prohibitively expensive if costs were assessed in terms of a fixed percentage of the GNP. This would also have the disadvantage that costs would not be directly related to benefits.

An alternative would be to allocate costs in proportion to the amount of research performed within a given country. This could be further refined by dividing the services provided in basic and optional groups. Basic services would be totally covered by the member's contribution. Optional services would be separately priced, perhaps on some kind of sliding scale.

One method of assessing changes might be the imposition of a page charge for inclusion in the system such as is now assessed by certain journals. This might have the additional advantage of cutting down excessive and redundant publication.

IV Economics of Information Systems

There is no doubt that large, international, and very expensive information systems are coming, primarily as a result of big power initiatives. If these are to serve as a force for peace by reducing rather than enlarging the development gap, their services must be provided to the less developed nations in a form which they can use and a price that they can afford to pay.

REFERENCES

1. International Statistical Year for Research and Development, Organization for Economic Cooperation and Development, Paris, 1967.
2. Bourne, Charles P., Characteristics of Coverage by the Bibliography of Agriculture of the Literature Relating to Agricultural Research and Development. IGC-PA-69-38, Information General Corporation, Palo Alto, 1969.
3. Bar-Zakay, Samuel N., "Technology Transfer Model," Technological Forecasting and Social Change, Vol. 2, No. 3/4, 1971.

A BREAKDOWN OF MANPOWER COSTS IN RELATION TO TASKS IN AN INDUSTRIAL INFORMATION SERVICE

R.B. Zaaiman
Head, Information Service, South African Iron and Steel Industrial Corp., Pretoria

SYNOPSIS

Direct labour costs for tasks in an industrial information service are determined. Manpower requirements are expressed in terms of time per task unit and of annual time needed for executing the prescribed number of task units. Activities analysed are Information Searches, Current Awareness Services and Co-ordinate Indexing.

The lack of published figures on task costs in information handling is frequently mentioned when the establishment of standard costs is discussed. The usefulness of published costs will depend on the accuracy with which they can be interpreted. A cost statement should accordingly be accompanied by a detailed description of the task costed and by an indication of the broader function of the unit which contains the task. Universally understood units should be used to express costs. Costs indicated in monetary terms, for instance, cannot readily be understood by anyone unfamiliar with the salary structure and cost of living in the country concerned.

This paper provides a break-down of direct labour costs for the Information Service of the South African Iron and Steel Industrial Corporation (Iscor). As an integrated steel company, Iscor mines its own ore and produces primary iron and steel products. It makes 3,5m tons of steel annually, to be increased to 8m tons by 1980 Iscor is the largest steel company in South Africa and needs to be self-reliant in the retrieval of information from publications. The Information Service serves all employees in Iscor's production, management and research teams. Just over 2 200 Iscorians actively use some or other aspect of the services provided.

The Information Service consists of four sections, of which the Library is one. Library tasks are excluded from this study. The Library contains 40 000 books, 10 000 published standards and specifications, and 13 000 bound volumes of periodicals. It receives 1 450 periodical titles, of which 64 are indexing journals. Publications acquisitioned number 6 800 items per year. The Library lends out 23 400 publications and supplies photocopies of 7 200 articles annually.

The other three sections of the Information Service are Information Searches, Current Awareness Services and Co-ordinate Indexing. All tasks in these sections are analysed, except those performed by two typists. Each section is led by a Senior Information

Officer who reports to the Head, Information Service. All information officers are graduate scientists. Posts are rated at 1 500 working hours (45 weeks) per annum when full-time holiday relief is not required, and at 1 700 hours (52 weeks) when full-time relief is required. The allocation of information officer time is shown in Table 1. Where a task uses significantly less than the full time available, the incumbent of that post assists in overloaded posts or acts as holiday relief.

Task and time analyses were made by Iscor's Systems and Procedures Division over a six-month period. Investigation methods included analysis of past performance, discussion with incumbents, personal observation and stopwatch measurement. The resulting standard times have been successfully applied for 30 months. Unit times are shown in decimal hours. Figures after the comma can be converted to minutes by multiplying with 60. The annual volume of work and number of hours required in regard to each task are also shown in the analyses. The analysis for only the first post is shown in full. For economy of space, tasks not directly related to the main objectives of subsequent posts are not shown.

It is realised that the manpower costs in the case study are peculiar to Iscor. A study of costs in many comparable information services will be needed to determine whether cost patterns of general validity underlie the cost figures of individual services. Information services vary so greatly in their approach and processes that the development of acceptable cost ranges may be more attainable than the establishment of specific standard costs. If the upper and lower limits of cost ranges, and points in between, are related to defined performance requirements, an information service can determine its own optimum position within the range.

Post No	Function	Available Time	Allocated Time	Allocated Time as Percentage of Total Allocated Time
	Information Searches			
1.	Senior Information Officer	1500	1595	9,41
2.	Information Officers (4)	6000	6332	37,36
		7500	7927	46,77
	Current Awareness Services			
3.	Senior Information Officer	1500	1337	7,89
4.	Patents Information	1700	1271	7,50
5.	Technical Press Review	1700	1869	11,02
6.	Business Information	1700	1624	9,58
		6600	6101	35,99
	Co-ordinate Indexing			
7.	Senior Information Officer	1500	1450	8,55
8.	Information Officer	1500	1470	8,69
		3000	2920	17,24
	Total	17100	16948	100,00

Table 1 : Allocation of time

TASK, VOLUME AND TIME ANALYSES	Volume p.a.	Unit Time	Annual Hours
POST 1: SENIOR INFORMATION OFFICER (SEARCHES)			
A. Scope of Post: Leads four other information officers in performing literature searches.			
B. Tasks			
1. Allocate work; control quality of information search results			
1.1 Extended searches (receive and discuss request with enquirer, 15 mins; decide on allocation and priority of search, enter in register, 7 mins; transfer request to information officer, 10 mins; discuss progress during search, 45 mins; evaluate and discuss search results with information officer, 20 mins; read customer's response to follow-up questionnaire, arrange for appropriate reaction, 20 mins.)	175	1,95	341
1.2 Medium searches (receive and discuss request with enquirer, 15 mins; decide on allocation and priority of search, enter in register, 7 mins; transfer request to information officer, 5 mins; discuss progress during search, 5 mins; evaluate and discuss search results with information officer, 5 mins; read customer's response to follow-up questionnaire, arrange for appropriate reaction, 5 mins.)	88	0,7	62
2. Perform information searches (receive request; orientate towards problem; choose search terms; look up search terms in published indexes; read, evaluate and select abstracts; write bibliographic details of required articles on work sheet; hand work sheet to clerk to obtain articles; perform similar search at library catalogue; ask co-ordinate indexing section to conduct search; mark selected items for photocopying; forward items to enquirer - photocopies directly and publications through library; conduct when needed continuation searches for enquiries previously regarded as completed)			
2.1 Extended searches	20	18,35	367
2.2 Medium searches	9	4,50	41
2.3 Short searches	30	0,25	8
(Calculated average time for all searches)	(59) Searches	(7,05)	(416)
3. Co-ordinate activities of Searches with following sections; Indexing, 20 mins; Patents, 15 mins; and Economic Information, 15 mins	48 Weeks	0,83	40
4. Control quality of bibliographies compiled	20 Bibliographies	2	40

	Volume p.a.	Unit Time	Annual Hours
5. Control quality of literature reviews written	4 Reviews	5	20
6. Liaise with 2 clerks on administrative and personal matters	48 Weeks	0,4	19
7. Discuss section's activities with supervisor	48 Weeks	2	96
8. Compile monthly report on activities	12 Reports	1	12
9. Counsel library on classification policy and on difficult classes or cases; recommend purchase, retention and circulation of publications; discuss convenient physical arrangement of library materials	48 Weeks	0,5	24
10. Inform Information Officer (Technical Press Review), Post 3, of Iscor information needs; discuss items for Technical Press Review	48 Weeks	0,5	24
11. Liaise with other sections of Information Service (e.g. library service counter; periodical registration; book orders; typists)	48 Weeks	0,33	16
12. Perform administrative functions (counsel information officers on work organization, discipline, Iscor approach, salaries, etc., 24 mins; write letters and internal memoranda on administrative matters, 9 mins; job evaluation and annual merit rating of staff and self, 7,5 mins; arrange for visits inside and outside Iscor for self and staff, 9 mins; sign leave forms, make relate leave arrangements, sign printing and stationery requisitions and revise all forms before reprinting, 22 mins; resignations and staff appointments, 22 mins; meet visitors, 24 mins; annual review and evaluation of activities, 9 mins; participate in Iscor productivity improvement programme, 4,5 mins; plan new services or revise existing services, 48 mins; evaluate own staff's recommendations of books for purchase or periodicals for subscription, 4 mins.)	48 Weeks	3,05	146
13. Go over outgoing and related incoming letters written in regard to searches; discuss action if necessary	60 Letters (Outgoing)	0,5	30
14. Arrange for and participate in training of personnel	48 Weeks	0,5	24
15. Do required reading			
15.1 Look through new books purchased	1950 Books	0,033	64
15.2 Look through first issue of new periodical titles	57 Titles	0,07	4

	Volume p.a.	Unit Time	Annual Hours
15.3 Read 10 periodicals on documentation	120 Issues	0,2	24
15.4 Read Iscor items related to work (e.g. Technical Press Review, Patents Review, research reports, minutes of committees, Market Research Bulletin)	48 Weeks	0,77	37
16. Attend Information Service staff meetings	96 Meetings	0,75	72
17. Visit Iscor's works and mines, visit other organizations, attend simposia, etc	12 Visits	7	84

POST 2: INFORMATION OFFICERS (SEARCHES)

The work of 4 information officers is considered. To obtain individual volumes and hours p.a., divide by 4

A. Scope of Post: Performs information searches, compiles bibliographies, writes literature reviews. Responsible to Post 1

B. Tasks

1. Information searches

1.1 Perform information searches (see Post 1, Task 2)

	Volume p.a.	Unit Time	Annual Hours
1.1.1 Extended searches	175	18,8	3290
1.1.2 Medium searches	88	4,75	418
1.1.3 Short searches	270	0,25	68
(Calculated average time for all searches)	(533)	(7,084)	(3776)
(Individual performance required)	(133,25) Searches	(7,084)	(944)
1.2 Conduct correspondence related to searches; discuss with supervisor when necessary	60 Letters (Outgoing)	1	60
2. Write literature reviews (each review stems from a search listed under Task 1.1.1)	4 Reviews	56	224
3. Compile bibliographies, with abstracts (each bibliography stems from a search listed under Task 1.1.1)	20 Bibliographies	14	280
4. Compile Meetings List (read announcements of meetings to be held by outside bodies; select, mark and edit those of interest to Iscor; give to typist; proofread typed list; supervise up-dating of distribution list)	60 Lists	0,33	20

POST 3: SENIOR INFORMATION OFFICER (CURRENT AWARENESS)

A. Scope of Post: Leads three other information officers in providing information on a current awareness

	Volume p.a.	Unit Time	Annual Hours
basis, including (a) personal service to senior personnel, (b) Patents Information (Post 4), (c) Technical Press Review (Post 5) and (d) Business Information (Post 6). Sub-sections (b) and (d) also perform information searches in the material handled there			
B. <u>Tasks</u>			
1. Technical Press Review (TPR)			
1.1 Edit TPR for quality and relevance of abstracts as well as quality of classification and cross-references	52 TPR's	2	104
1.2 Plan annual survey of readers' requirements; supervise execution; analyse results; arrange for necessary action	Year	30	30
2. Patents Information			
2.1 Edit Patents Review for relevance of patents chosen and quality of classification	26 Reviews	2	52
2.2 Go over outgoing and incoming correspondence; discuss action if necessary	30 Letters (Outgoing)	0,5	15
2.3 Control quality of patent information searches	120 Searches	0,4	48
2.4 Read responses by customers to follow-up questionnaires; arrange for appropriate reaction	120 Responses	0,033	4
2.5 Plan yearly survey of readers' requirements; supervise execution; analyse results; arrange for necessary action	Year	20	20
3. Business Information			
3.1 Discuss trends in information needs with Information Officer (Business Information)	26 Meetings	0,5	13
3.2 Control quality of work done during business information searches	230 Searches	0,083	19
3.3 Read responses by customers to follow-up questionnaires; arrange for appropriate reaction	230 Responses	0,033	8
4. Disseminate information selectively to members of top management, without computer aid. (Note: The service did not yet exist when the task sheet was written)	Year	504	504
4.1 Plan and develop service			
4.2 Maintain regular personal contact with members of top management to ascertain their needs			
4.3 Read abstract journals; instruct clerk to send copies of selected abstracts to customers			
4.4 Check response to abstracts; instruct clerk to obtain and send full articles to customers when asked for			

POST 4: INFORMATION OFFICER (PATENTS INFORMATION)

A. Scope of Post: Compiles and distributes weekly Patents Review; performs searches for information in patents; assists Information Officer (Technical Press Review), Post 5, and Information Officers (Searches), Post 2. Responsible to Post 3

Note: A new post, volumes/times are subject to revision as techniques are developed

B. Tasks

	Volume p.a.	Unit Time	Annual Hours
1. Patents Review (PR)			
1.1 Scan patent specification journals, select and mark items for cutting and pasting by clerks, Post 10			
1.1.1 United States	52	5,04	262
1.1.2 United Kingdom	52	4	208
1.1.3 South Africa	12	4	48
1.1.4 Germany (Derwent abstracts)	52 Journals	0,92	48
1.2 Develop and rationalize systems used for above tasks (time required is included in 566 hours listed under Tasks 1.1.1 - 1.1.4)			
1.3 Check PR for mistakes after it is pasted up (arrangement, numbering, lay-out, etc.)	26 Reviews	0,7	18
1.4 Supervise current updating and annual revision of PR circulation list	52 Weeks	0,38	20
2. Perform information searches in patents; instruct clerk to obtain and lend out relevant patent specifications to customers			
2.1 Extended searches	60	5	300
2.2 Short searches	60 Searches	0,25	15

POST 5: INFORMATION OFFICER (TECHNICAL PRESS REVIEW)

A. Scope of Post: Compiles and distributes a weekly Technical Press Review containing abstracts of articles from periodicals received by ISCOR; provides limited current awareness service on personal basis to top management. Responsible to Post 3

B. Tasks

	Volume p.a.	Unit Time	Annual Hours
1. Compile Technical Press Review (TPR) with three sections: Management; Technical, i.e. iron and steel production; Mining			
1.1 Scan 768 periodical titles with 10236 issues p.a. select articles; mark suitable abstracts (published ab-	8720 Abstracts	0,133	1163

	Volume p.a.	Unit Time	Annual Hours
stract or sentences in article); write page numbers, classification and cross-reference symbols on paper slip; affix slip to first page of article. (Hours p.a. include scanning time for all periodical issues)			
(Required daily performance for abstracts, 250 days p.a.)	(35) Abstracts	(0,133)	(4,65)
(Time per periodical issue)	(10236) Issues	(0,113)	(1163)
(Required daily performance for periodical issues, 250 days p.a.)	(41) Issues	(0,113)	(4,65) (hrs. p.d
1.2 Translate titles and abstracts from German and French into English or Afrikaans	780 Translations	0,183	143
1.3 Proofread typed abstracts and cross-references as typed on paper strips for cutting and pasting-up by clerks, Post 10. Each strip contains spaces for five items, including section headings, run-over of abstracts, etc	3120 Strips	0,058	182
1.4 Proofread corrected strips	52 TPR's	0,23	12
1.5 Check pasted up, final sheets before duplication for correctness; check final corrections	52 TPR's	0,67	35
1.6 Read customer's request for full articles; analyse interest trends; discuss with supervisor	1400 Forms	0,019	26
1.7 Make annual survey of reader requirements; analyse results; plan new approaches; review classification system	Year	32	32
2. Provide limited current awareness service on personal basis to executives	52	0,81	42

POST 6: INFORMATION OFFICER (BUSINESS INFORMATION)

A. Scope of Post: Selects newspaper and periodical articles; builds up cuttings file; disseminates news items selectively; performs information searches. Responsible to Post 3

B. Tasks

1. Scan newspapers, periodicals, annual reports, etc; select and mark items for (a) cuttings file; (b) dissemination to individual executives; (c) cuttings book for circulation to selected senior officials

1.1 Scan S. African newspapers (2190 issues @ 20,34 mins); overseas newspapers (1712 issues @ 8,1 mins); periodicals (656 issues @ 11,94 mins); annual reports (120 issues @ 5 mins)	4687 Issues	0,238	1117
(Time per item selected)	(10000)	(0,112)	(1117)

	Volume p.a.	Unit Time	Annual Hours
1.2 Mark items for photocopying or cutting; write item's page number on front page of newspaper; write page number on separate slip for periodical and affix to front cover (40 items daily @ 18 secs)	250 Days	0,2	50
2. Scan S. African Hansard (House of Assembly) during session of parliament; mark items of interest; write and attach circulation list of officials concerned with items; arrange for photocopies of items when requested (200 pages of 2 columns per week)	19 Weeks	1,32	25
3. Scan Debates of Senate as under Task 2 (120 pages of 2 columns per week)	20 Weeks	0,8	16
4. Perform searches for business information, using own files and other published sources	230 Searches	0,283	65
5. Review files for removal of obsolete material	741 Files	0,2	148

POST 7: SENIOR INFORMATION OFFICER (INDEXING)

A. Scope of Post: Leads one other information officer and 1,65 clerks in the co-ordinate indexing of ready-made abstracts, in maintaining a thesaurus, and in performing information searches on the indexed material

B. Tasks

1. Indexing

	Volume p.a.	Unit Time	Annual Hours
1.1 Read, analyse and index abstracts of articles			
1.1.1 Technical Press Review (index abstracts in technical section of TPR for manual card system; average of 3 indexing terms per abstract; write terms on abstract for entry by clerk)	4338 Abstracts	0,024	103
1.1.2 Iron and Steel Institute (London) abstracts (select abstracts of importance for Iscor; index for Termatrex; average of 6 indexing terms per abstract; write terms and abstract number on work sheet for hole-drilling by clerk)	3222 Abstracts	0,096	309
1.1.3 British Iron and Steel Research Association reports (as for Task 1.1.2)	95 Reports	0,096	9
1.1.4 Bureau of Mines (USA) reports (as for Task 1.1.2; average of 7 indexing terms per abstract)	22 Reports	0,23	5
1.1.5 American Society for Metals and/or Engineering Index abstracts (as for Task 1.1.2)	3333 Abstracts (Estimated)	0,096 (Est.)	320 (Est.)
1.2 Index bibliographies (as for Task 1.1.2)	46 Bibliographies	0,23	11

	Volume p.a.	Unit Time	Annual Hours
(Average indexing time per item)	(11056) Abstracts	(0,068)	(757)
(Required daily performance, 230 days p.a.)	(48,2) Abstracts	(0,068)	(3,27) (hrs p.d.)
2. Reading, evaluating and eliminating ISI abstracts not of interest to ISCOR (not included in volume under 1,2)	639 Abstracts	0,033	21
3. Perform information searches on indexed material (receive request from information officer, select search terms from Thesaurus, enter terms on work sheet for search by clerk, indicate combinations or terms required, instruct clerk on requirements of search, 20 mins; read and evaluate abstracts (30 average) found by clerk, discuss relevance of abstracts with information officer, continue search until satisfactory information is found, 76 mins.)	82 Searches	1,6	132
4. Review and improve thesaurus	Year	94	94

POST 8: INFORMATION OFFICER (INDEXING)

A. Scope of Post: Index ready-made abstracts; assist in compiling thesaurus; perform information searches in indexed material. Responsible to Post 7

B. Tasks

1. Indexing (Details similar to Post 7, Task 1)

1.1 Abstracts

	Volume p.a.	Unit Time	Annual Hours
1.1.1 Technical Press Review	5093 Abstracts	0,024	122
1.1.2 ISI abstracts	3823 Abstracts	0,096	367
1.1.3 BISRA reports	113 Reports	0,096	11
1.1.4 Bureau of Mines reports	26 Reports	0,23	6
1.1.5 ASM and/or Engineering Index abstracts	3960 Abstracts	0,096	380
1.2 Bibliographies	54 Bibliographies	0,23	12
2. Read and eliminate ISI cards not of interest to Iscor	790 Abstracts	0,033	26
3. Perform searches on indexed materials (details	98	1,6	156

	Volume p.a.	Unit Time	Annual Hours
similar to Post 7, Task 3)			
4. Review and improve thesaurus	Year	94	94

POST 9: SENIOR CLERK (INDEXING)

A. Scope of Post: Drills document number holes in Termatrex random numeric access cards; performs physical part of information searches. Responsible to Post 7

B. Tasks

	Volume p.a.	Unit Time	Annual Hours
1. Drill holes (receive coded indexing work sheet from clerk, Post 10; withdraw required Termatrex cards from file; check whether broader terms appear on Termatrex cards; if so, withdraw cards for broader terms; prepare Termatrex cards for new indexing terms not yet in file; combine all required cards and drill hole; replace cards in file. Write new indexing terms on 5 x 3 card, file in nursery list for possible addition to thesaurus. Write new company names on 5 x 3 card; file for later addition to company name thesaurus)			
1.1 Abstracts, with 6 indexing terms on average	14694 Abstracts	0,053	779
2. Use reference books to establish relationships of new company names; make entry for company name thesaurus	120 Company names	0,24	29
3. Perform searches (collect Termatrex cards from file according to search work sheet received from information officers; combine cards according to instructions on work sheet; view on light box; read and write down numbers of holes shining through; remove relevant abstract cards from abstract card file (average 30 per search); hand to information officer; replace Termatrex cards)	180 Searches	1,425	257
4. Prepare new set of Termatrex card (write each indexing term and its broader term, if any, on a Termatrex card)	Year	20	20

POST 10: CLERKS (2 POSTS)

A. Scope of Post: Renders clerical services to Searches, Indexing and Patents sub-sections. Responsible to Post 1

B. Tasks

1. Information Searches

	Volume p.a.	Unit Time	Annual Hours
1.1 Receive completed work sheet for searches from information officers; compare periodical titles on work sheet with list of periodicals received by Iscor; write symbol on work sheet to indicate titles available; look up location of other titles in "Periodicals in South African Libraries"; write holding library's symbol on work sheet. (60 entries per work sheet at 0,15 mins per entry)	292 Worksheets	0,52	152

	Volume p.a.	Unit Time	Annual Hours
1.2 From work sheet, write out alphabetical list of periodicals available at Iscor, with relevant date and page numbers; return work sheets to information officers. (30 entries per work sheet at 0,57 mins per entry)	292 Worksheets	0,285	83
1.3 Use written alphabetical list to collect periodicals from shelf; look up required page number; check article title against work sheet; insert flag; hand to information officer. Find out location of Iscor periodicals not on shelf; arrange with library for recall or reservation. (30 entries per work sheet at 1,4 mins per entry)	292 Worksheets	0,7	204
1.4 Arrange with library to obtain on interlibrary loan periodicals not in Iscor; receive; flag required article; hand to information officer	345 Periodicals	0,033	11
1.5 Prepare bibliographies (receive abstract journals from information officers; photocopy pages according to worksheet; mark required abstract on photocopy; cut out and sort abstracts in date order; paste abstracts in date order to fill A4 sheets; photocopy pasted-up sheets; hand to information officer; return abstract journals to shelf)	17230 Abstracts	0,031	537
2. Indexing			
2.1 File one 5 x 3 card alphabetically for each new indexing terms (subject terms and company names)	600 Cards	0,017	10
2.2 Stamp sequential number on items to be indexed	7411 Items	0,001	9
2.3 Receive weekly batch of already numbered Technical Press Review (TPR) abstract cards; check for correctness; place at back of numerical file. (183 cards per week)	52 Press Reviews	0,173	9
2.4 Enter TPR abstract numbers on manual co-ordinate index cards (read terms written on abstracts; withdraw cards from alphabetical file; write four-figure abstract number on cards; file index cards; file abstract cards) (3 terms average per abstract at 0,88 mins per term)	9431 Abstracts	0,044	418
2.5 Perform searches in manual system (receive search work sheets from information officer; withdraw manual cards from file; compare numbers on cards; write down coinciding numbers; remove relevant abstract cards from file index cards. (For average search terms are compared on 2 annual sets of cards and 5,4 abstracts are retrieved)	180 Searches	0,5	90
2.6 Cut out abstracts of BISRA reports; paste on 5 x 3 cards; file numerically	208 Abstracts	0,058	12

	Volume p.a.	Unit Time	Annual Hours
2.7 Numerically sort abstract cards retrieved during Termatrex searches and refile	5400 Cards	0,0054	29
2.8 Prepare new set of manual index cards for TPR-write indexing term on card; add broader term if on old card	2620 Index Cards	0,008	21
2.9 Receive completed indexing work sheets for Termatrex from information officers; find random numeric access codes for indexing terms on work sheet; write codes on work sheet; hand sheets to Post 9 for drilling. (Average sheet contains 2 abstracts or 12 indexing terms; code for each term average one letter and two figures)	3730 Worksheets	0,1166	435
2.10 Photocopy selected abstract cards after co-ordinate index searches; use Savin 220 photocopier	3000 Cards	0,0123	37
3. Patents Information (New activity: times measured but subject to change as system develops)			
3.1 Publication of Patents Review (PR) - photocopy pages from patent journals; cut out abstracts; paste on A4 page, arrange for printing; address 75 copies of completed Press Reviews. (Average of 96 abstracts per week)	26 Weeks	5,46	142
3.2 Order patent specifications; record receipt of specifications; certify accounts; write out record cards under name of patentee and patent number; sort and file cards alphabetically for patentee, numerically under country for patent number	1082 Specifications	0,1	108
3.3 Lend out specifications; keep loan records; refile specifications on return	218 Specifications	0,096	21
3.4 Circulate S. African patent journal; file all patent journals and patent specifications on receipt; maintain files in good order	52 Weeks	1,25	65
4. Photocopying			
4.1 Photocopy items not yet specified	2920 Exposures	0,0123	36

MANAGEMENT PROBLEMS AND COSTS IN SETTING UP A
MECHANIZED INFORMATION SYSTEM
(International Food Information Service - IFIS)

Dr. U. Schützsack
International Food Information Service, Frankfurt/Main,
Federal Republic of Germany

SYNOPSIS

In information science there is a general trend for international cooperation on a bilateral, a multi-lateral or on an international basis. This will reduce the costs on the national level by sharing the responsibilities and also ensure complete coverage of the relevant literature. This was the background for setting up the organization "International Food Information Service" (IFIS). Its first service is the abstract journal "Food Science and Technology Abstracts" (FSTA).

1. Subject Scope

The journal covers the whole field of food science and technology by abstracting about 1,000 journals (the main food journals cover-to-cover and the fringe journals on a selective basis) and by abstracting patents and books. Reports are only included on a limited scale. Food science is mainly applied research undertaken by the food industry. The results often are not published, but only distributed in the form of internal reports in the companies.

The first number of the journal was published in January, 1969. The technical details are:

a. it is a monthly journal containing about 1,000 abstracts a month, the number of abstracts per month however is increasing constantly.

b. Every issue has a subject and author index produced by computer. The subject index is a kind of a KWOC - index with descriptors, key words and phrases.

c. The monthly subject indexes and monthly author indexes are cumulated at the end of the year to prepare printed cumulated yearly indexes.

d. The subscription price for one year of the journal is DM 675.- including the cumulated yearly index.

e. For the time being, the main output of the system is the abstract journal. Gradually we are moving into a second phase, which could be called the phase of tailor-made information by tapes. The tapes available include all data of the abstract journal (bibliographic part, abstract part, subject index part). The cost for a yearly subscription is DM 6,000.- (12 monthly tapes). The tape of the cumulated yearly index is charged separately. Programs for SDI and retrieval are prepared by the organization and the respective mechanized services will be offered in the near future. Operated on a manual basis these services are already obtainable now.

2. Management (see figure 1)
The decision to create IFIS was taken by the sponsoring organizations in 1968. The decision was preceeded by about three years of discussion and planning.

The pressure for creating some kind of information service in food science did not come from the information scientists, but from the food industry after evaluating existing abstracting journals in this field, the main abstracting journals being "Chemical Abstracts" and "Nutrition Abstracts and Reviews", and by comparing the amount of literature scanned by these journals with the amount relevant to the food scientist.[1]

It was found that the existing journals only were covering a part of the literature. A broadening of the scope of these journals could have been foreseen. It was found, however, that it was desirable to have one service covering the whole field, food science having become a discipline of its own.

It was left to the planning institutions to find out what kind of information service was wanted, e.g.: bibliography, conventional abstracting service, computerized abstracting service, and what literature should be included in such a service.

This investigation was made by the Commonwealth Agricultural Bureaux (CAB) in the United Kingdom and the Institut für Dokumentationswesen (IDW) in the Federal Republic of Germany and resulted in two publications evaluating primary and secondary journals in food science.[2,3]

The two institutes mentioned above became the nucleus of the organization "International Food Information Service" together with the American partner, the Institute of Food Technologists (IFT). An agreement, which in reality is based on an exchange of letters, was signed in February 1968. At a later stage the Centrum voor Landbouwpublikaties (PUDOC - Holland) came in as a fourth partner.

The main features of this agreement are:

a. The organization is governed by a Management Committee, which meets at least once a year. In the Management Committee the sponsors are represented by two members, PUDOC by one.

b. The agreement is limited to three years. In January 1969 the first journal number was published, so the period of the agreement ends December 1971. After this period the organization is expected to be self-supporting.

c. The main contributions of the partners are in kind and not in cash (see division of work).

d. The work is divided as follows:

The English partner is responsible for preparing the manuscript including the following steps:

selection
scanning

cataloging
abstracting
indexing
preparation of the input in a standard form
overall editing

Abstracts are produced partly on a decentralized basis (see figure 2) through agreements with institutions organizations and individuals in all parts of the world. The abstracts are put in standard form by the collaborating bodies, but are not indexed.

Abstracts are indexed at the IFIS office in Shinfield and in most cases a considerable amount of reediting is necessary.

The German partner is responsible for:
data input on paper tape (done by a commercial service)
data processing (done by the Zentralstelle für maschinelle Dokumentation = ZMD)
printing and distribution (done by a commercial printer)
provision of magnetic tapes (for partners and for sale).

It was felt by the partners, that the above division of work would best meet the requirements to establish a mechanized effective information service in a relatively short time for the following reasons:

The English partner has had experience in abstracting by publishing about 15 abstracting journals in agriculture and related fields for nearly 30 years.

The German partner, the ZMD, responsible for data processing, has had experience in handling non-numerical data.

It would otherwise have been completely impossible to become operational in less than one year.

3. Costs
Costs arising can be divided into development costs and running costs.

A. Development (figures 3 and 3a)
For the development of the system the following projects had to be undertaken:

a. To define the scope (Food science) and on the basis of that to give a list of periodicals, which had to be abstracted (in FSTA mainly conventional literature is included). In this operation three institutes worked together (Institut für Ernährung, Potsdam-Rehbrücke; Institut für Dokumentationswesen, Frankfurt/Main, Commonwealth Agricultural Bureaux for Dairy Science and Technology, Shinfield).

b. Organizational changes in the Commonwealth Bureaux of Dairy Science and Technology, especially:

Organization of the input (input from all parts of the world), preparation of a categorized work sheet for data input,

interconnection between Dairy Science Abstracts and Food Science Abstracts
(about 2,000 abstracts published in Dairy Science Abstracts),

development of a concept for a subject index (KWOC).

c. A semi-modular programming system had to be built up by the Zentralstelle für maschinelle Dokumentation (30 programs).

It should be mentioned that the figures of ZMD (figure 3a) are calculated on a non-commercial basis. If the development of data processing had been handed over to a commercial firm, there would be an increase in costs.

Not included in these figures are the salaries of a large part of the staff involved in the operation on both sides.

d. It has been said above that tapes are available. In order to be able to use these tapes for SDI or retrospective search some programming is neces-necessary.

The kind of p.og. am may be different from country to country, from company to company an' therefore it is left to the customer to do this work.

The partners of the organization themselves are also doing some development work in this direction for those customers who have no computer facilities and therefore rely on the services produced by IFIS.

Running costs can be divided into:

a. Preparation of the manuscript (responsibility of the English partner)
b. Input)
c. Data processing) (responsibility of the German partner)
d. Printing)
e. Mailing and distribution)

a. <u>Preparation of the manuscript</u>
The preparation of the manuscript is entirely in the hands of the IFIS office in the U.K.

The office has the following scientific staff:
Editor and assistant editor
8 scientific information officers
1 indexer

It should be kept in mind, that above 40% of the abstracts are prepared by outside abstractors. Outside abstracts, however, need a considerable amount of reediting, which is done by the scientific information officers in addition to abstracting.

b. <u>Input</u>
Input is on paper tape, for which 3 paper tape typewriters are necessary.

Investment in 3 paper tape typewriters:	DM 42,000.-
The price for a simple tape typewriter, which is sufficient for this purpose, is	DM 14,000.-
The price for the punching of paper tape is calculated on the basis of characters. 1,000 characters including tax	DM 5.-
A calculation shows that 1,100 characters are equivalent to one abstract	DM 5,50

As the monthly output is about 1,000 abstracts, the monthly costs for input on paper tape are about DM 6,000.-

Annual costs	DM 72,000.-

The punching of paper tape is done by a commercial service bureau. The tape typewriters and material (paper tape and paper) are made available by the IDW. Moreover the Institute has to pay the costs for maintenance.

To be able to work on shifts two tape typewriters have been installed at the service bureau.

A third typewriter is in the IFIS office for text and structure corrections.

c. Data processing

Data processing is done by ZMD on a nonprofit basis on an IBM 1460 with additional equipment for data conversion and preparing of TTS tapes for hot metal setting.

d. Printing

The input material for printing Food Science and Technology Abstracts is a TTS tape for Linoquick. (Hot metal setting). The costs for one volume are:

Size	17.7 x 24.4 cm	
Contents	monthly edition about 1,100 abstracts = 160 pages of abstracts, 8 pages author index, 32 pages subject index	
Circulation	1,000 copies per month	DM 11,000.-
Annual expenses		DM 136,000.-

e. Mailing and distribution DM 34,000.-

It is intended to shift from hot metal setting to phototype setting in the near future. In this way the printing time is reduced considerably. The costs for phototype setting are nearly the same as for hot metal setting.

Experiences

(1) Subject Scope

The subject scope is based on the definition that the journal should cover the field of food science including fundamental aspects (chemistry, physics etc.). Food science again deals with the product from when it leaves the farm till it is on the table of the consumer.

This, however, only is a working hypothesis. In practice it means that the journal covers a fixed number of periodicals, which is the orientation mark for the user.

There is a tendency to increase the number of journals. The pressure is coming from scientists working in the fundamental disciplines (radiochemistry, biochemistry, biophysics), from scientists in the applied fields (complaining that their field - e.g. packaging - is underrepresented) and finally from inside the organization, from members of the staff becoming aware of new journals and of journals, which were not known when starting the service.

The journal started with about 10,000 abstracts the first year, and we now have about 16,000 in 1971. The number could be increased easily to 18 - 20,000 this year, but for financial reasons this is difficult, keeping in mind that the organization has to be self-supporting from 1972. So there is a conflict here between scientific requirements and financial possibilities.

The indexing problem, even strictly outside the scope of this paper, should be mentioned here. Indexing is based on an open word list with some structural elements (see and see also). New words are included as they appear in the scanned periodicals. The testing of words is done in the operational phase, which may create difficulties especially when cumulating the monthly indexes to a yearly index.

(2) Management

It has been said above that the whole system has been set up under the assumption that it would be self-supporting after three years.

Self-supporting in this connection means that all expenditures have to be met by income from subscriptions and sale of tapes.

The organization is more or less at this point now.

However, it should be kept in mind that the conditions to reach this goal were very favourable for several reasons.

(a) The number of food scientists and scientific organizations is increasing, food science being a young discipline;

(b) There is a change from food production on the farm and in smaller firms to food production on a large scale, which requires a scientific approach to solve the problems.

So food science may be regarded as a special case, but not as a singular one. The conditions may be similar in other areas (see Special Information Services).

In the above mentioned case the financial support of the system is limited to the development of the service and to the first operational phase. It was relatively easy to come to an agreement, because the sponsors only had to commit themselves for a limited period.

In other cases it might be necessary to support a system on a continuous basis, which in reality means support by the governments. To get such support is much more difficult, but in the long run such a system is more stable though perhaps less effective.

It is essential to know these factors when planning a new information service. An international system depending on continuous support from the start should have the support of the governments.

(3) Information services in special fields

FSTA is an information service in a special field. The reaction to this service on the part of the user is very positive, which shows that there is a demand for specialized information services. This demand obviously is independent of the subject.

The first number of another abstract journal in a special field, "Aquatic Sciences and Fisheries Abstracts", has just been published, a cooperation between;

FAO Rome
Institut für Dokumentationswesen, Frankfurt/Main
Bundesforschungsanstalt für Fischerei, Hamburg
INRA Biarritz
Information Retrieval Limited, London.

This journal has also already been accepted by the scientific community. So it might be worthwhile trying to come to some general conclusions.

It is obvious that information services in special fields can presnet the information in a more adequate form to the user than broad information systems by: deeper indexing, abstracts and subject knowledge. The scientific staff of these centers is primarily discipline oriented and is in constant contact with the scientists in its field.

The pressure to build up a specialized information service comes in nearly all cases from the scientific community, the main customer, who afterwards also is represented in the organization and has an influence on scope and coverage.

So the service is tailored to the requirements of the customer already from the beginning.

The number of specialists in a given field in one country is always small. Therefore it would be completely uneconomic and unrealistic to build up an information service for one country only. International cooperation is here needed even more than in broad fields.

In the two cases mentioned here, the carrier language is English, and one may assume that this should be regarded as a general axiom.

In some cases, the choice of langüage should depend on the scientific community which has to be served.

An example where English is a secondary language is Wine Research.

Some of the main wine producing countries are Germany, France, Hungary, Romania and some South American countries. An abstract journal for wine therefore should not be in the English language only, but also take into consideration the languages of the main wine producing countries. That is just what is done in the new abstract journal which is going to be set up: in this abstracts will be in English, German and French.

IV Economics of Information Systems

The concept given here is a network of information centers in special fields working in close cooperation with large information centers in: biology, chemistry, agriculture. Specialized information services do not compete with large information centers, but rather are complementary.

The elements of this network must be bound together to ensure compatibility in this way enabling an exchange of information, also in a technical sense between the services.

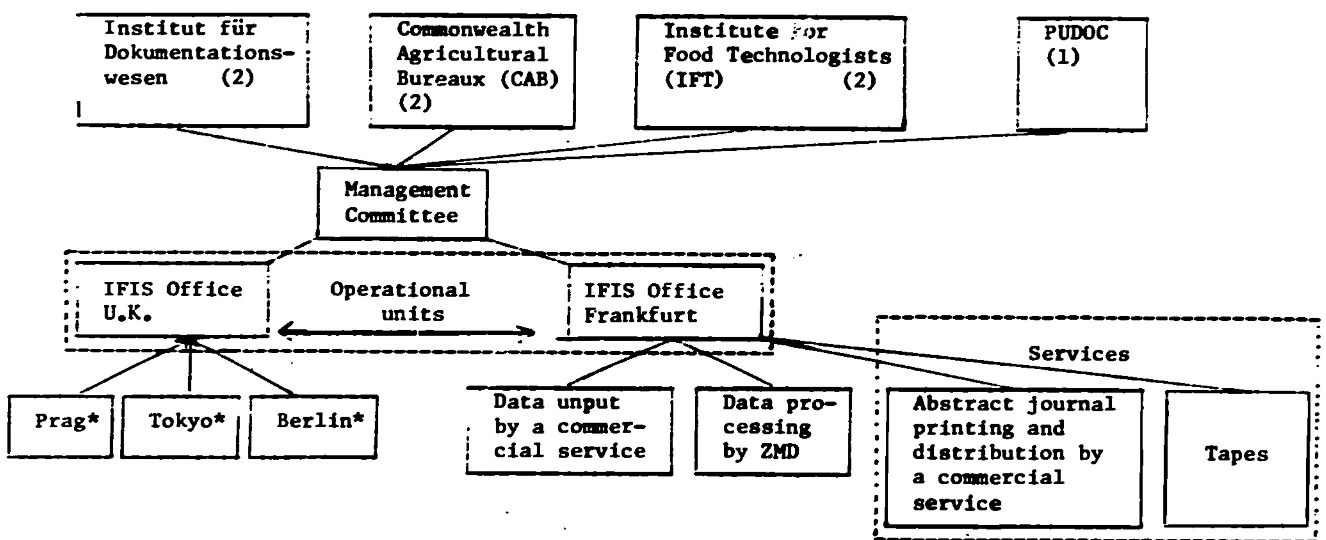
This could be done by agreeing, for example, on a common work sheet for data input, on the same way of cataloging and indexing, and on a macro thesaurus, in which microthesauri are included.

References:

1. Moyer, J.C. An information service for the American food industry. Food Technology (Champaign) 17 (1963) Nr. 1, S. 32, 35-36
2. Mann, E.J. Evaluation of the World Food Literature - Results of an International Survey. Farnham Royal, Bucks.: Commonwealth Agricultural Bureaux, 1967, 182 S.
3. Schützsack, U. Trends of documentation and information in food science, S.457-71. In:Proceedings of the 2nd International Congress of Food Science and Technology, Warszawa, Aug. 22-27, 1966. Warszawa: Wydawnictwo Przemyslu Lekkiego i Spozywczego, 1967, 512 S.
4. Bourne, P.C., North, B.J., Kasson, S.M. Abstracting and Indexing Rates and Costs. Minneapolis, Minnesota: Eric Clearinghouse for Library and Information Sciences, May 1970.
5. Marron, A. Costs of secondary service operations. In: Digest of 1970 Annual Meeting, National Federation of Science Abstracting and Indexing Services. Boston, Mass. 1970, S. 25-29.
6. Weil, B.H. Cost Effects. In: Digest of 1970 Annual Meeting, National Federation of Science Abstracting and Indexing Services, Boston, Mass. 1970, S.30-33.

Figure 1

Organization Chart of IFIS



Decentralized input
* for details see figure 2

Figure 2

Scanning of periodicals and abstracting for FSTA

Institutions	Number of periodicals evaluated	abstracts per year
I. IFIS offices		
Central IFIS office, Shinfield, Berks.	281	4,385
IFIS office in the National Lending Library (NLL), Boston Spa, Yorks.	235	1,886
IFIS office in the Tropical Product's Institute, London	127 = 643	1,184 = 7,455
II. Other institutions		
Institut für wissenschaftliche und technische Information in der Landwirtschaft, Prag	94	2.400
Institut für Ernährung, Potsdam	36	815
Japanese Section of the Institute of Food Technologists (IFT), Tokyo	37	400
British Food Manufacturers Industrial Research Association (BFMIRA), Leatherhead, Surrey	45	588
Metal Box Co., London	22	100
Biocentralen, Kopenhagen	14	88
Others	42 = 290	467 = 4,858
	933	12,313

Figure 3

Developments costs for the abstract journal FSTA

Analysis of literature	DM 60,000.-
Organizational changes in the U.K.	DM 60,000.-
Programming by ZMD	DM 179,260.- *
Programs for SDI and retrospective search	Estimated DM 100,000.-

* detailed analysis see figure 3a

Figure 3a

Costs for development of the system by the Zentralstelle für Dokumentation 1968

For the planning and development of the information system Food Science and Technology Abstracts 1968 the following figures are presented:

a.	Documentation work 91 working days, 8 hours a DM 20.-	DM 14,560.-
b.	Program-work 660 working days, 8 hours a DM 20.-	DM 105,600.-
c.	Machine-time IBM 1460 146 hours a DM 355.-	DM 51,830.-
	Machine-time IBM 7765 10 hours a DM 100.-	DM 1,000.-
d.	Material (Punched cards, Paper-tape, paper etc.)	DM 270.-
e.	Administration Costs (Telephone, Postage, Travel Expenses)	DM 2,000.-
f.	Magnetic-Tapes	DM 4,000.-
	Total	DM 179,260

Figure 4

Annual Total Running Costs (1970) for FSTA +

Preparation of the manuscript	DM 440,000.-
Input on paper tape	DM 72,000.-
Data processing	DM 130,000.-++
Printing	DM 136,000.-
Postage and distribution	DM 34,000.-
Annual Index (printing)	DM 35,000.-
	DM 847,000.-

+ Output 13102 abstracts
  monthly output 1100 abstracts

++ detailed analysis see enclosed figure

Figure 4a

Data processing	Nonprofit basis
Data conversion 190 paper tapes on magnetic tape IBM 7765 52,5 hours a DM 100.- a DM 300.-	5,250.-
Data conversion magnetic tape on TTS tapes GIER 10,25 hours a DM 100.- a DM 300.-	1,025.-
Data processing IBM 1460 285 hours a DM 355.- a DM 1,100.-	101,175.-
Costs for data processing	107,450.-
Costs for material, paper tapes, punched cards, paper	2,550.-
Costs for man-power	20,000.-
Total	130,000.-

FUNCTIONS AND ECONOMICS OF AN
OPERATING AUTOMATED LIBRARY SYSTEM

Richard S. Hirsch
International Business Machines Corporation
San Jose, California, U.S.A.

SYNOPSIS

A library management system was developed to evaluate its functional adequacy. The functions are designed to be interactive, integrated, and in real-time. The system is operating experimentally to assist in determining costs of the various functions as a first step in establishing economic feasibility of the total system.

Introduction

Over the past dozen years the writer has at various times, and in diverse capacities, been associated with information retrieval efforts of several kinds. A recurring observation stemming from these efforts is that an information retrieval system is difficult to achieve for at least three reasons: First, mechanization of a data base--in short a library--is a necessary first step. It seems that most information retrieval efforts falter for lack of a large data base whose content could be manipulated. Second, considering the costs involved in most sophisticated information retrieval proposals, it is difficult to obtain the funding necessary to implement even an experimental facility large enough to yield meaningful results. Third, the system procedures have sometimes tended to retrieve information in ways that are slower than those procedures currently in use. The problem in the last instance is sometimes aggravated by hardware and software interfaces between the human and the system that are less than optimally designed. On the one hand, the programming language may present users with diverse and inconsistent sets of system resources. The hardware, on the other hand, may be inappropriate for the application, as, for example, when a typewriter is used to print out long messages. These questions have been effectively addressed in a recent article by McAllister and Bell (1).

Accordingly, about five years ago, it was decided in IBM to investigate a somewhat different approach to answers to information retrieval questions. The approach was to look at mechanization of libraries, the data bases for information retrieval. Two assumptions prompted this approach: It was assumed that, since a library's procedures were better structured (and perhaps better understood) than information retrieval processes, a library would therefore be technically easier to implement. A second assumption was that the costs of a library could be determined and the economic feasibility of a system established more easily.

While it can be shown that implementation of a library system has indeed been easier than implementation of an information retrieval system, cost-justification has nevertheless continued to remain elusive. Some attempts, however, have been made (2, 3) and this paper will examine a recent effort made at identifying the costs of library operations in terms of their potential for automation as related to a particular set of mechanization specifications.

Overview of the System

In order to make meaningful judgments about savings that mechanization could realize, it was necessary to have in mind, if not in operation, a well-structured automated library system. In deciding on a system, the various functions of a library were at first considered as separate applications for mechanization. The work already done in this direction was examined: Catalog input and library bulletin programs (4) and circulation control systems (5), among others, were evaluated and tried. The result was to move as others (6, 7, 8) had toward a "total" integrated, real-time and interactive system (9).

Such a system is now operating experimentally in the Los Gatos (California) IBM Laboratory. The system is under control of a computer to perform the record-keeping functions of the library: namely, collecting, controlling and displaying information about the individual holdings in the library collection.

The work of a library is stated in six major functions: ordering, receiving, cataloging, circulation control, maintaining files, and searching the bibliographic data (in terms of a user's interest).

On initiation of an order by a librarian, the system accepts both bibliographic and purchase information about the item to be acquired. A file is established for the ordered document, and the subsequent controls include such operations as retaining the identity of the requester of the document, printing the necessary purchase order, and issuing claims and renewals (if and when necessary).

The receiving function changes the appropriate records on the basis of information available to the librarian from the document. For example, all the accounting records, such as, outstanding orders and fund allocations, are modified. A "Notice to the Requester" is generated, and bibliographic information available from the physical document is added to the file.

The system's cataloging function not only maintains information about the library's individual holdings, it permits the librarian (and only an authorized librarian) to add to, delete from, or correct the bibliographic file. Thus, for example, new subject headings can be added, misspelled words and names can be corrected, or dates and report numbers can be added, changed or removed. These single revisions will adjust all uses of the item throughout the system. This function also controls the display of the bibliographic information either on a display screen (IBM 2260) or as a printed document (by typewriter or high-speed printer) to produce various hard-copy catalogs.

Circulation control simply keeps track of the current status of a document and creates appropriate notices, reminders, and lists concerning borrowed items.

For odering, receiving, cataloging and circulating to be under the control and command of the librarian, the system must be able to display any information in the file at any time. Provided the librarian is authorized to make changes to the file, and within procedures designed to prevent accidental changes, the function of file review and maintenance permits the librarian to revise file detail.

Ordering, receiving, cataloging, circulation and file review and maintenance are collectively the record-keeping functions of the system. These functions are, however, of small value if they do not contribute to the needs of a user. Accordingly, the sixth function is called searching. By means of searching, the patron (whether a librarian or other user of the library) is able to explore the holdings according to his interests. The IBM 2260 Display Terminal is used to provide on-line searching. The user simply enters (through the terminal keyboard) titles, author's names, or descriptor terms related to the subject in which he is interested. If he wants a long list, or a hard copy of even a short list, it is provided from a printer, off-line.

There is obviously a great deal of detail that has not been mentioned about each of the six functions. For instance, there is not room to describe how the display interface is designed to guide the user through the system procedures. There are, moreover, other operations of the system that cannot be discussed at all in the brief space of this paper. For example, there are such subjects as how the files are structured, how the system administers date controls for the library, the way printing queues handle the many forms needed to run a library efficiently, and how the system collects statistics on the use of the library's facilities.

Specifications for the system were derived from prolonged and cooperative studies by librarians and system analysts. A group of ten working librarians within IBM provided an exhaustive list of desirable functions. The list was reviewed by several librarians outside IBM, and the specifications are the result of a consensus of librarians of what should be automated.

Economic Considerations

Discussions with library administrators led to the conclusion that such a system is needed and could be financially supported---if only one knew how to cost-justify it. Accordingly, an important objective for making the system experimentally operational was to provide opportunities to evaluate its various facilities both functionally and financially.

Warheit (10) recently reviewed and discussed the economic problems involved in mechanizing libraries: real-time versus batch-mode functions; time-sharing versus dedicated computer systems; and implementation of a total system versus selective applications. All these questions must be considered in any cost-justifying effort. Moreover, McCann (11) even more recently summarized what, after a decision is made to implement an integrated library system, the librarian must do organizationally to prepare for an orderly transition from today's manual operation to the computer system. Although some cost-savings are alluded to, neither article deals in any detailed way with the economic factors involved. Accordingly, what follows adds, in a sense, a few more building-blocks to the foundation laid by earlier workers.

In 1969 a group of system analysts in IBM undertook a preliminary 4-week study of library operations. What was attempted was to get a detailed statement of library costs, department by department, function by function, activity by activity within departments and functions. With cooperation from the librarians, they surveyed performance rates and costs, and they attempted to quantify those data processing and filing procedures in which the duplication of activity could be eliminated through automation by an integrated system such as the one described. Work sheets were developed on which were specified such items as job activities, work loads, queue size, number of files, etc.

Table 1 presents in summary form the kind of activities investigated. In each case, the system analysts and librarians agreed on the extent of duplication involved in an activity and on what improvements would result from eliminating the duplication. Thus a task was thought to have mechanization potential if it was judged that duplication existed and that mechanization would improve overall performance of the task. The percentages in the table are the fractions of the task, measured in labor, that satisfy the criterion of mechanization potential.

Table 1 does not, however, show the depth of detail analyzed in order to obtain the mechanization potential. It is presented in order to suggest that an extremely detailed survey of library activities and costs is necessary for a meaningful analysis to be made. Of particular importance, moreover, is the notion that the survey and analysis was made with a particular system of mechanization in mind.

In conclusion, it should be pointed out that it is not only an information retrieval system that is hard to cost-justify; the dollar-value of a library as such is also impossible to assess. While it may be agreed that the history of science, government, industry and the arts might have been quite different without libraries available for scholars in these fields, it is doubtful that a financial value could be placed on such sources of reference. Certainly no dollar figure, however reasonable, could be objectively justified in terms of corresponding savings. Consequently, the argument for mechanization has sometimes been made that the library will as a consequence of automation provide more information and better service. But since the initial value of the information and the service are unknown, the value of improvement is equally unknown. It is even difficult to carry the point that information will be more up-to-date because documents can be processed faster.

The cost evaluation described here was, therefore, designed to help identify library activities in terms of a given mechanized system in order to determine the activities that might, or could, be positively impacted by mechanization. That positive impact, what we have called the percent mechanization potential, may then provide some measure of offsetting costs to help justify the system economically. After some further refinements in costing procedures, what will be needed then is that one or more libraries install such a system experimentally on a scale broad enough to verify that services can be at least as good as those currently available, and to establish how much the overall costs are more or less than they are now.

Table 1
Percent Mechanization Potential By Department Activity

Processing Order Requests: In Print	54	
Out of Print	33	
Accessioning	76	
Invoice Processing	94	
Processing Gifts, Exchanges and Bulk Backlog	64	
All Other	11	
Acquisitions Subtotal		46*
Cataloging	47	
LC Card Ordering and File Maintenance	31	
Keypunching - catalog	100	
Cataloging Subtotal		53*
Serials Ordering	73	
Serials Title Preparation: Precataloging	100	
Postcataloging	69	
Follow-on Processing	72	
All Other	14	
Serials Subtotal		54*
Binding/Labeling: Supervision	0	
Materials Handling	0	
PAM Binding, Permabinding	0	
Date Due/Stamp/Typing/Sorting/Filing	100	
All Other	0	
Binding/Labeling Subtotal		23*
Circulation Desk	50	
Record Keeping	100	
Sorting/Shelving	0	
Verification	100	
Supervision	0	
All Other	20	
Circulation Subtotal		46*
Reference Desk	10	
Book Review	0	
Professional Reading	0	
Special Projects	0	
Typing/Filing	100	
Supervision	0	
Circulation	33	
Interlibrary Loan Search - Verify	33	
Interlibrary Receipt - Return	33	
All Other	0	
Reference Subtotal		10*
GRAND TOTAL		42*

*A department subtotal (and the grand total) is expressed as a percentage of the total departmental cost and represents a weighted average of the individual component activities.

IV Economics of Information Systems

References

1. McAllister, Caryl and Bell, J.M. "Human Factors in the Design of an Interactive System." Journal of the American Society for Information Science, 2 (March-April 1971).

2. "The Costs of Library and Informational Services." Knight, D. M. and Nourse, E. S., eds. In Libraries at Large, (New York: R. R. Bowker, 1969), 168-227.

3. Culbertson, D.S. "The Costs of Data Processing in University Libraries: In Book Acquisition and Cataloging." College and Research Libraries, 24 (November 1963), 487-89.

4. Bateman, Betty B. and Farris, E.H. "Operating A Multilibrary System Using Long-Distance Communications to an On-Line Computer." Proceedings of American Society for Information Science, 5 (1968), 155-62.

5. Kennedy, R.A. "Bell Laboratories' Library Real-Time Loan System (BELLREL)." Journal of Library Automation, 1 (June 1968), 128-46.

6. Black, D.V. "Library Information System Time Sharing: System Development Corporation's LISTS Project." California School Libraries, (March 1969), 121-6.

7. Bruette, V. R., Cohen, J., and Kovacs, Helen. An On-Line Computer System for the Storage and Retrieval of Books and Monographs, (Brooklyn, New York: State University of New York Downstate Medical Center, 1967).

8. Fussler, H. H., and Payne, C. T. Development of an Integrated Computer-Based Bibliographical Data System for a Large University Library, (Chicago: Chicago University, 1968.)

9. Alexander, R.W. "Toward the Future Integrated Library System." Proceedings of 33rd Conference of FID and International Congress on Documentation, (Tokyo, 1967).

10. Warheit, I. A. "Design of Library Systems for Implementation with Interactive Computers." Journal of Library Automation, 3 (March 1970), 65-78.

11. McCann, D.B. "Getting Ready." Datamation, 16 (August 1, 1970), 22-26.

THE RISE AND FALL OF AN INFORMATION CENTER

Bernard Bayer
Mechanized Information Center
The Ohio State University Libraries

SYNOPSIS

Somewhere in the cloud of statistics, on copies found and items searched, lie the remains of a dead information center. In this paper, the center will be reconstructed, so you will know how it was that it came into being and why it was that it then died. (I was there in Buffalo, New York, U.S.A., at the Center during most of the happenings.)

Six years ago was an opportune time to get into the information business.

In Washington, D.C., Congress had passed the State Technical Services Act, which would provide federal funds for universities and nonprofit organizations to help industry find and use research information. From Albany, New York, the State Capital, funds were flowing freely into the State University system. In Buffalo, New York, the State University of New York at Buffalo had the funds and mandate to be "Instant Berkeley". Four years earlier, it had been a private, underfunded university. Now it was building up in many fields, one of which was information retrieval.

So it was that the Technical Information Dissemination Bureau (TIDB) was founded at the State University of New York at Buffalo, in October, 1966. So it was also that one of its missions was to develop a computer-based information system and to provide services to industry under the State Technical Services (STS) Act of 1965.

The Bureau rose with the act, and then was sunk by it.

Prologue[7,1]

"If we had had this legislation 25 or 30 years ago, we might have prevented the economic depression that today exists in Appalachia," said President Lyndon Johnson about the STS Act, after signing it.

Congress gave its reasons for passage of the bill: the act was "...to promote commerce and encourage economic growth by supporting state and interstate programs to place the findings of science usefully in the hands of American enterprise."

The level of funding originally was to be for a three year period, starting at $10 million the first year, $20 million the second, and $30 million the third year. However, Congress appropriated $3.5 million the first year, and approximately $5.3 million for both the second and third years.

This appropriation was divided among those states that wanted to participate in the program, and the federal funds were to be matched by state funds. In 1969, 47 states did participate in the program.

In 1968-9, New York State's share of the funds was $355,000. TIDB's share was approximately 10% of that amount.

All the states proposed methods of disseminating information: seminars, conferences, field work, newsletters, and special courses.

Various institutions throughout New York State, in turn, proposed to the State's Commerce Department how they would spend the money, which was to be apportioned mainly geographically.

The State University of New York at Buffalo proposed that it provide services to five counties in Western New York through the Technical Information Dissemination Bureau by "maintaining a computer-based data bank of up-dated technological information with the capability to retrieve, filter, and match this information with special needs and problems of users in the area of Western New York." This was a unique approach in the State. The proposal was accepted.

The basic means for accomplishing these goals was to be a computer-based current awareness service.

Tooling Up: October 1966 - August 1967 [3,4]

An economic study of the region had already been performed for the New York State Office of Planning Coordination.

The report stated: "The conventional standard for assessing the economic condition of a region is its rate of growth--growth in employment, growth in income, growth in sales, growth in population. By this standard, which leaves out the qualitative aspects of an area's situation, the Western New York Region's economic history since the mid-fifties has been one of relative decline and/or stagnation."

The report went on to say that "there are signs that these absolute or relative downtrends are bottoming-out."

Items manufactured in the region include abrasives, ferroalloys, chemicals, cereals, business forms, graphite, tires, steel, electronic equipment, paint, and parts for automobiles.

The region covers 4,600 square miles and contains 1.567 million people, 9.6% and 9.3% of the state total, respectively. There are 1,478 companies, with most of them employing less than 50 people.

The area also has major universities and colleges, a nationally known research and development laboratory, and a major producer of defense items and space hardware, such as rocket engines.

How do you handle the diversity of industry? How do you provide services to them?

We thought the answer would be through computer-based systems, which could handle a diversity of information, but we wanted to find out for sure.

In the summer of 1967, we hired five men to find out what industry was doing for information services, and whether they would make use of our proposed service.

By the end of the summer, these men had called on 289 companies in the five counties of Western New York.

The response was in general favorable. Approximately 60% of the companies considered handling information a problem. Most of these companies said they were interested in our proposed solution to handling information.

But there were some disquieting notes. For example, one of our men evaluated his exchange of information with a president of a company in Niagara Falls:

(1) Is the company aware of the information explosion? Yes

(2) Does the company believe that handling information is a problem? Yes

(3) Was the company interested in the proposed computer-based system service? Yes, so much so that the president suggested we give packets of information about the Bureau and its services to the local Chamber of Commerce for them to use to attract industry.

(4) Last question: should we contact the company about our services when they become available later in the year? No, the man couldn't see how his company could use them.

We plunged ahead.

The system was programmed and became operational. In brief, this is how the system works:

> An interest profile is developed for each participant. The terms in this profile are weighted and matched weekly against terms appearing in the titles of articles and reports. In this manner, pertinent items are selected and printed out as a series of two-part notification cards.

In one year the system scans approximately 400,000 articles and government reports.

High Points: September 1967 - June 1968[5]

We wanted to find out if we could run a large-scale system, and meet deadlines every week.

Two major pilot projects were begun for: (1) 57 engineers and scientists from 29 companies in Western New York, and (2) approximately 200 faculty members, whose services were subsidized by the University.

The first pilot project ended in January 1968, the second in June 1968. During both projects, we provided free current awareness service and free copies of articles.

Needless to say we were a big success. It's amazing what you can do when the price is right.

No sooner did our cards go out, than the requests (one part of a two-part form) came back.

I remember receiving a call from a dentist on the Health Sciences Faculty on the last day of the faculty pilot project: "If I get these in today, can I still get free copies?"

"Yes," I said. "Now that the free pilot project is ending, how about subscribing?" (He had been notified about this before.)

"No," he said, "I wouldn't make enough use of it."

After the pilot project ended, it was necessary to charge for our services. Based on our level of funding, the fees were:

> $225 for one year of current awareness service, plus $150 in one-time fees, and $2-$3 for each copy of an article or government report furnished.

There was enough information on each notification for each man to obtain his own copy. We offered the copy service as a convenience. We bought tear sheets of the articles from a commercial service.

By the end of the year, we had sent out 30,000 notifications and furnished 1,700 copies of articles and reports, to industry alone.

Decline: July 1968 - May 1969[6]

We began this phase with approximately 60 paying customers. Faculty members were incensed at having to pay. They wanted the university to continue to furnish it without charge. Companies thought the price was high, and after thinking it over they said they really didn't need any service. Handling information was not much of a problem after all.

However, those companies and faculty who subscribed to the service seemed to like it.

During this time we provided another 127,000 notifications and filled requests for 5,000 articles and reports.

For some companies, our service was keeping whole engineering departments up to date.

And that was one of the problems. Large companies that should have been subscribing to 20 or 30 profiles were subscribing to one. Instead of obtaining $4,500 to $6,750 in fees, we obtained the basic $225.

TIDB was caught in a squeeze: if we raised prices, we would lose customers; if we lowered them, we would lose money on each customer, with STS funding at its current level.

The choices seemed to be:

(1) more STS funding, which as far as we were concerned was too low to provide service. (For example, four years later, the amount of a grant from the National Science Foundation to The Ohio State University for providing similar services to faculty and students was $204,000 for the first year. This is more than TIDB received in STS funds in four years.)

(2) higher fees

(3) go to management of local industry and ask for a subsidy for us to provide services. We felt that we were doing an important job for Western New York, and industry needed us.

(4) stop offering the service

(5) expand our marketing area.

The choices were evaluated as follows:

(1) more STS funding was impossible

(2) higher fees were not very wise. Our customers were complaining already.

(3) going to industry for a subsidy was settled for us. A provost at the University turned down the idea.

(4) before stopping the service, we wanted to exhaust all possibilities

(5) expanding the market into other regions seemed like the only solution left.

Therefore, we began to make presentations, through the other STS offices in New York State, to industry in their region.

However, we were too late.

On May 5, 1969, we were notified that the U.S. Department of Commerce had no funds to allocate in the coming fiscal year, and that we could not count on supplemental funding in the current year, for the STS program.

The End: May 1969 - March 1970[8]

It was to be a lingering death for the program.

Resuscitation efforts and conflicting stories were coming from all sides: Washington, Albany, and from the State University. These were some of the stories and happenings:

(1) On June 26, 1969, we heard that Maurice H. Stans, Secretary of Commerce in President Nixon's cabinet, had not requested funds for the program. He wanted to wait until the program was evaluated by an outside firm; he had assumed that there were funds available. "The decision was not political....The request by the President for a cut in the budget required reductions in all agency budgets and so this request and the Secretary's lack of knowledge of the program led to this decision."

(2) On June 30, 1969, we received an evaluation form to be filled out and returned to Arthur D. Little, Inc., the outside firm hired to evaluate the program. All STS offices were either visited or sent forms to be filled out.

(3) On October 17, 1969, we were told that Secretary Stans was now completely behind the program, and had written a "very strong letter" to the Bureau of the Budget requesting $5 million for the new fiscal year. This would be a supplemental appropriation.

(4) On October 20, 1969, the results of the Arthur D. Little evaluation were made public. The press release stated that "the STS Program provides a useful...service in transferring technological know-how which is of substantial benefit to the nation."

(5) In November, 1969, the Bureau of the Budget approved the request for supplemental STS funding.

(6) In December, 1969, Congressional hearings were held on all supplemental appropriations, including the one for the STS Program. The Senate approved the $5 million appropriation; the House did not.

(7) On December 20, 1969, a Senate-House conference committee on supplemental appropriations compromised: $0 for the STS Program.

(8) On January 15, 1970, we were officially notified of the termination of the program.

Also, the STS appropriation approved by the New York State Legislature for its support of the program was tied to federal funding. No federal funds, no state funds.

In addition, the State University of New York at Buffalo did not have extra funds to continue its subsidy of these services.

And, warring factions within the University made some decisions to help sink the Bureau. Instead of trying to lighten the financial load on the Bureau, these factions made it heavier: computer time for services to faculty was no longer subsidized, in part, by the Computing Center, and the $8,000 data bank that had been paid for by the Library was now charged directly to the Bureau.

Along with the STS Program, TIDB had gone out and obtained a number of contracts to do systems work for libraries and other information handling organizations. But these contracts were not enough to keep TIDB afloat against the combined efforts of the U.S. Government, the New York State Government, and the State University of New York at Buffalo.

The Bureau services were stopped as of February 2, 1970. The books were closed, contracts were finished, and the Bureau disappeared. All that remains are reels of magnetic tape, containing a data bank of approximately one million items and the programs for the current awareness system.

Epilogue

After going through my notes, the letters received from the various STS offices, and thinking back to those days, I am reminded of a quote from the book, In Cold Blood:

> **"I didn't want to harm the man. I thought he was a very nice gentleman. Soft-spoken. I thought so right up to the moment I cut his throat."[2]**

References

1. *Smith, W. P.; Proposal Submitted in Accordance with State Technical Services Act of 1965 for a Technical Information Dissemination Bureau for Western New York; Buffalo, New York; State University of New York at Buffalo; December, 1965.*

2. *Capote, T.; In Cold Blood, New York; Random House; 1965; page 244.*

3. *Myers, L. B.; An Economic Evaluation of the Western New York Region Basic Inventory and Analysis Phase Volume 1-Summary; Buffalo, New York; Cornell Aeronautical Laboratory, Inc.; August, 1966.*

4. *Bayer, B.; Task 5 Project: Establishing Liaison with Industry in Western New York; Buffalo, New York; Technical Information Dissemination Bureau; July, 1968.*

5. *Lazorick, G. J.; Annual Report New York State Technical Services Program 1967-1968; Buffalo, New York; Technical Information Dissemination Bureau; August, 1968; (Dr. Lazorick was the first director of TIDB and was responsible for choosing the direction TIDB was to follow.)*

6. *Bayer, B.; Annual Report of New York State Technical Services Program for Western New York 1968-1969; Buffalo, New York; Technical Information Dissemination Bureau; July, 1969.*

7. *Hamilton, A.; "State Technical Services: Congress Swings the Axe"; Science; volume 166; 26 December 1969; pages 1606-1608.*

8. *Many, many letters and memoranda from the coordinating STS office in Albany to the Technical Information Dissemination Bureau in Buffalo, New York, 1967-1970.*

Session Four - Discussions

ECONOMICS OF INFORMATION SYSTEMS
Chairman: Mr. W. Uhlmann (Sweden)

DR. W.E. BATTEN (UK): Over the last two or three years there has been a considerable change of emphasis in the general quest of information science. We have moved our area of inquiry away from the question of how shall we do it, to the rather more rewarding question of how shall we do it more cheaply, or better still, how shall we do it with a greater degree of cost effectiveness?

This I believe to be a very encouraging movement because first, it is intrinsically good to go after value; and secondly, you cannot go for value until you employ cost factor reasoning, which evidently records some progress.

Our ability to select a required sub-set from a large set is still extremely limited, even at substantial cost, but at least we now possess an expertise and a methodology for doing it and the world has already enough experience, if only communication and cooperation could be improved. We do not yet know how to obtain 100% retrieval and 10% precision through an outside agency, whether that agency be mechanical or human, but we are beginning to know why we cannot, and before long I think we may even be able to discern the likely limits of search performance and even the approximate shape of the cost performance curve.

This could lead to a situation where the user might be able to choose to what point of the performance curve he wants to go in a particular situation; in other words, he does not need to buy a Rolls Royce when a Ford Escort will do and vice versa.

It is not too early to cast an economic eye on the nature, size, and infrastructure of the data bases being produced today. I am not concerned with their various incompatibilities. I am far more concerned that we should recognize how they came into existence and how they get themselves paid for.

But let's not tackle the wrong problems. I doubt whether the oil technologists come together once a year to deplore the prevalence of sulphur in crude oil. The data bases that we now enjoy and at the same time decry are largely the products of the exertions and energies and investments of producers of secondary literature and let us never forget it. Their composition is, therefore, the distillation of many years in the publishing market. It represents a package for which there is a sufficient market to ensure its continuance. We might see a changing succession of products which will drive us even madder than we are. So it may be well to recognize that crude oil does contain sulphur and probably always will and that alcohol products do need to be carefully refined before they are potent.

While being appreciative of the existence of reasonably viable and stable data bases, let us not assume that they are necessarily in the best form for the information market. It is a plausible hypothesis that there is room for a whole new social activity in reprocessing these stable data bases into more opportunistic selections which might be both long-lived or short-lived. But this re-arrangement will never be done for a mere six people who want it for a mere six months - not unless there is a mere six million pounds of interest lying behind the quest.

IV Economics of Information Systems

I hope I don't strip us of dignity if I use a homely analogy: the good housewife does not require her vegetables in hundred weights. She needs one garlic, one pound of onions, one pound of carrots and so on. So a retailing organization was a very early component of commerce. The technological scientist, if not the research scientist, is a little like the housewife in that he needs a pinch of this and a cupful of that. We shall need information supermarkets where the stocks are both selected and organized and arranged to suit a certain clientele.

What will this mean in techniques and in policy? Here is an area in which we know little. If the search centers will merely acquire several data bases and offer all or some for the customer's choice, the overhead will be high. If we create mixed sub-collections out of the variety of today's collections, the overhead attached to re-arrangement will also be high. I will not speculate about how often the right answer will be "a" or how often it will be "b"; what is almost certain is that merging will be high and that sometimes it will be the right choice.

This points to the need for a unique and compact item identification and for an item record format which is at least uniform in its hospitality if not in its structure. We need these things not because we already know them to be essential for operations, but because we can discern that they will make developments so much more rapid, so much cheaper and so much more within the compass of even small organizations.

The re-arrangement and repackaging of data bases will be an area of trial and error that we cannot by-pass whatever our prejudices and whatever our hopes.

We now know a little more about how to make use of acceptable files. We know much less about creating acceptable files out of record files. This is an area that ASLIB hopes to be actively studying during the next two years. We shall be extremely happy to collaborate with anyone who is thinking of working along the same lines. In particular, we will welcome any statistical observations, any numerical observations on the degrees of overlap between different data bases and upon cases where the existence of overlap is not a nuisance, but is beneficial.

PROF. M. KOCHEN (US): Mr. Disch mentioned yesterday the "quality of life" with regard to an objective to be maximized in Norway in connection with information systems planning. "Quality of life" is measured mostly by the GNP per capita. I propose the following measure of the "quality of life": the variety of the goods and services available to a person, together - and this is very important - with the wisdom to choose among this variety. So what I call "wisdom", or the ability to choose among these alternatives, is just as important as variety.

A few examples: in the US there is a large number of services and a tremendous variety of services for the poor which the poor have no idea about. In Israel, for example, there is - I have just discovered - a tremendous amount of know-how about urban planning as a potential world-wide service that I am willing to bet very few of the people who ought to know do know about. The information referral center is now becoming one of the services of the library. In Philadelphia, for example, the libraries are stepping into this role, matching the poor of Philadelphia - there's 25% unemployment there - with the many agencies that provide services for the poor, but which the poor don't know about.

Within the library itself, the number of services is increasing. As the number of services increases and the number of needs too, what is lacking? Lacking is a way of coupling the two. The referral specialist needs directories to help him beyond his memory. He needs directories that minimize the possibilities of large resources and would enable a variety of problem solvers to meet needs for which resources exist.

This is a key problem because if it is solved, it opens for analysis a whole host of other problems which are of theoretical and practical significance.

We need to bring some analytical thinking into the area. There is some progress, although very much in its infancy, in the analytical methodology of directory structures. By directory structure I mean, for example, poison control. Somebody calls in: my baby has just swallowed a white liquid and has the following symptoms. The person at the other end doesn't necessarily have to be an expert on toxicology, but there are four kinds of directories that could help him: the white pages of the telephone directory lists all the poisons and he could look through all 100 pages of this, but this is ineffective. Something a little more sophisticated would be the index of symptoms and the properties of the substance, but that isn't very useful either for a poison control situation. A classified directory such as the yellow pages is a little bit better, but still really not adequate for poison control, and finally, there are new things that we could think of such as directories that have constant zoomability which permit the person answering this emergency request on the telephone to get a bird's eye view of what the poison might be and then zoom down on it like an eagle does and crawl along the terrain like a worm until he pinpoints the precise or close to the precise substance that might have been the poison.

The purpose of these various kinds of directory structures is to match resources to the real needs of problem solvers in a cost-effective way. There is progress in this area. There are today a number of programmes, computer programmes and mathematical methods of analysis that can help design directories. For example, there is a program to estimate the cost of using a directory. It is a computer program which takes its input from four types of tables. Basically these tables tell you to whom a referential consultant could refer a question that comes in. For example, if the question "where can I get food stamps?" or "who can help me with this urban problem?" comes in, there are a number of sources the consultant can turn to in this directory and there are a number of other direct sources where he knows he can get the answer and there is a cost attached to each such look-up.

There are a number of people to whom he can "pass the buck" with a probability of success and he knows that there is a cost attached to that. There is the possibility of referring the entire request to somebody who he thinks can handle the query better than he himself can. There is a cost attached to this as well, and the possibility that the second person will not be able to answer, but will in turn pass it on. So there are three possibilities: either to answer it by look-up or directly from one's expertise, or refer to somebody else. There is the danger of being trapped in the bureaucratic cycle run-around and never getting out of it, while the customer waits.

There is a computer program for analyzing what the average cost per query is for a whole variety of queries and one can run this program and determine what is the best structure of a directory which will serve the purpose. There is another program which permits us to see under what conditions a network of referential consultants, each of whom passes the query on to another and uses the other's resources, does better than the best of these resources alone. It is a rather interesting situation where five heads are better than one, even though one is the best of the five. There are conditions under which this is the case. The conditions are connected with the utility of a response and the cost of transferring.

There is also a model which permits one to predict the success of information referral centers. For example, in the US in Philadelphia, the center which matches services available with the needs of the poor, measures the number of queries per month and their growth, and they are predictable from mathematical models. Most of the questions that come in this way to the center in Philadelphia, in time get answered successfully. A three-way telephone hook-up is used in which the referential consultant makes sure that the connection which he suggests between an agency and the client is in fact satisfactory.

Analytical approaches are possible and, in my opinion, necessary, because when data bases get very large, we cannot afford trial and error; and to a large extent we cannot afford experimentation. It is important before gathering a lot of data to have some idea about what to do with this data. There are no practical applications for this kind of analysis as yet, but they lead to ideas such as the following. We could set up experimental systems which more than pay for themselves. The yellow pages which have been around for many years are just the beginning. They are a remarkable way of utilizing the technology of the telephone and it seems to me that there are tremendous opportunities for redesigning the structure of the yellow pages to make them more responsive to the needs of the users, and to make them available to a greater variety of users.

A system such as an information referral center which could be set up in many developing countries at rather low capital investment, would answer the needs of decision makers in problem solving at all levels, and could repay itself very amply. Poor people have all kinds of needs; for example, the need for water. There is a lot of know-how about water, but the poor don't always know how to tap this information; they don't know whom to ask. It is a question of getting names of people, telephone numbers, all very practical things, to lead them to the source which would help them solve their particular problem.

Such a thing could be easily implemented and a directory like the yellow pages is only one aspect of this. One could do a lot with census taking. A mobile computer which takes a census, can at the same time disseminate information like a mobile library which goes into less developed zones and collects demographic data which it then makes available to the people and their representatives to make them aware of the problems of which they perhaps weren't previously aware.

There is real opportunity to help decision making in the public sector. We are beginning to get some empirical data on the needs and practices of decision makers with a view towards knowing what kind of directory structure would be useful to them.

Let me terminate on this note. The kind of directories we really should strive for should allow maximum cost effectiveness. Those which have some capability of synthesis, some capability of showing a person what is the over-logical structure of the topic he is concerned with, have educational value - it is only such factors which maximize benefit.

<u>*MR. A. LEBOWITZ (Israel):*</u> The origins of my paper are in my work in the US during the early stages of the INIS and AGRIS information systems, and my opportunity to view international systems both from the point of view of one of the large, highly developed, super-powers, as well as to view those same systems from the viewpoint of a developing country such as Israel. I have been trying over the course of the past year to develop a conceptual model of the information system which exists here and its relationship to the world-wide information scene, just to help me understand some of the forces that are at play.

I firmly believe that scientific and technological information is really a necessary input for the industrial development of any country. Information can be obtained either through R&D within the country, or through the exploitation of R&D taken from someplace else, the usual method being technology transfer.

Insofar as information depends on locally produced R&D, what is happening now is that the rich countries are becoming richer, the poor ones becoming poorer. In my paper I cited OECD data from which you see that the more developed the country, the higher the percentage of its GNP is used for R&D, so it gets even further ahead. In 1963-4 the percentage of the GNP used for R&D was 1.5% in the US as compared with 2/10 of 1% in such countries as Portugal and Greece.

One method of handling the technology transfer process is to import information in bulk. I am not talking about the non-formal, non-library systems, but about the documentary systems where it is possible to buy data bases, books, journals, etc. and to utilize them in another country.

What do you do when you have a system that has to be able to provide for the research within a given country, as well as for the world's production of information, while it itself is based on a research program which is only a fraction of one percent of the world's production of knowledge? The only real solution is to base national systems on much larger international systems.

This brings up the next question. How does one allocate the costs of those international systems? Can you do it by dividing up the total cost in terms of how much is made of the information? Probably not, because that puts a very large amount of the cost on small developing countries who are using the information developed by the super-powers to a much greater extent, proportionately, than the super-powers themselves.

There are, however, some alternatives. The first is to relate the contribution that each country makes to the cost of operating the overall systems, to the GNP or some function of the GNP within each country. At first sight that looks like a relatively simple thing to do. However, it has the disadvantage that it does not take into account the actual use made of the research within the using country.

For example, a country whose emphasis is on agricultural research ought to pay a higher percentage of the cost to the agricultural information system than a country whose principal emphasis is on some other branch of knowledge. If you relate those two factors - the GNP and the quantity of research - formulas for fairly allocating the amount can be developed.

Another possible method might be to allocate a page charge for each page of publications issued. A proposal was made in the US some time back that when an automobile is sold the price ought to include the cost of the disposal of that automobile when its utility ends. I think the same kind of thing could be applied to publications. A country which wants to participate in an international system and wants its information to be indexed, abstracted, and distributed through the international system, should pay a share of the cost in proportion to the amount of publications that it actually generates. Perhaps this would cut down on some of the redundant publications. However, I haven't noticed that page charges on journals have had that effect.

Technological progress is extremely important for the under-developed countries. It was Pope Paul who recently said that another name for peace could be technology.

MR. R.B. ZAAIMAN (South Africa): When one undertakes the expensive task of costing of information services, one does so to improve the efficiency and effectiveness of the service. It may help one to reach this objective if one can compare the cost of one's own information center with those of other countries. However, when making inter-center comparisons, one must be careful to compare comparable activities. Important differences in approach may exist between information centers which are attached to industrial organizations on the one hand, and research institutes on the other.

IV Economics of Information Systems

The requests which we recieve at Iscor - South African Iron and Steel Corporation - for information on technical problems form 79% of all requests received. Searches on management problems form the other 21%. The research department initiates only 8% of all information searches. Within the information service one finds four types of activity: direct services to users which make up 63% of all staff time; administration and communication within the center take up 14%; tasks which are performed purely to create input take another 14%; liaison with Iscor personnel and with outside information sources make another 9%.

In research institutes, these relations may be different, especially as regards input. Iscor's information center does not attempt to pioneer revolutionary developments in information science. The non-experimental nature of our work is reflected in the negligible amount of time allocated to planning and development of new services.

As we are attached to a non-subsidized, profit-oriented organization, we must keep costs as low as possible. Salaries form 76% of our budget and accordingly, we attach great value to the measurement of the amount of human effort which is required to perform each task.

One hopes that the practice of publishing the costs of information in terms of money, which is now prevalent, will yield to the use of time as an international measuring factor for cost comparison.

The times given in my paper indicate what it costs to run a service regardless of the actual success rate achieved. In this sense the time costs are merely descriptive of activity and thus are of limited value. It is difficult to interpret the efficiency and effectiveness of an information center by looking at its costs alone. A standard of measurement must be found by which the time costs of activities can be interpreted by means of a common denominator.

It has been suggested that the cost per user should be used as a common denominator. However, we found that two thirds of our users use one third of the time that we spend on searches. The other one third of users use the remaining two thirds.

A more acceptable standard may be the time cost per "successful unit" of output. The concept of a "successful unit" must in each instance be defined within the context of the service rendered. If the cost per "successful unit" is determined, time costs are useful in the evaluation of a service and also in comparisons between information centers.

We can analyze the cost of success in the case of the Iscor Technical Press Review. The Review contains short abstracts of periodical articles. Readers of the Review can ask for copies of the full articles.

When 8,720 abstracts were published annually, the time taken for the preparations per abstract was 0.13 hours. The number of requests for copies of full articles then came to 36% of the number of abstracts published. We consider a request for an article a sign of a successful abstract. The time per successful abstract was 0.37 hours. Since getting these results the choice of the items for the Technical Press Review has become more selective and we now publish 40% fewer abstracts. The number of requests for articles has actually increased by 5% and is now for 63% of the abstracts published.

Being more selective, the time for the preparation of each published abstract increased from 0.13 hours to 0.22 hours. However, the cost per request or per success decreased from 0.37 hours to 0.35 hours. This is an instance where the publication of a large number of abstracts lowers the unit cost per abstract published. This may create an impression of greater efficiency, while the effectiveness of the service may in fact be lower. The measurement of the cost of successful units of output seems to be necessary for effective inter-center comparisons.

MR. R. HIRSCH (US): I think it might be of interest in view of the things that have been discussed earlier to present a brief history of the life and near death of our automated library system. It started officially about five years ago. Unofficially it began well over 15 years ago. The work was carried out by a large number of people. But even a large organization like IBM is not given to developing expensive programs without worrying about their economics.

It was approximately two years ago that the company asked us to look at the economics and to prove that indeed there were libraries that could use this system if it were developed. We tried to obtain figures from libraries. Dr. Batten suggested that the reason that some of these figures are not available is because companies are jealous of their successes. I would also suggest that companies and universities are reluctant to give this kind of information because they are also jealous of their failures.

We were able finally to obtain help of a small university near the IBM laboratory where I work in California to help us identify some of the operations which could be improved through a mechanized system of the kind that is described in my paper.

About this time our company decided that there was no future for this program and it had to be stopped. It is a truism that in governments and in large corporations, it is very difficult for the management to stop things that people want very much to do. We stopped what was called the program, but continued it under a different name. The corporation has tried to stop many programs which ultimately proved to be successful so that even if management didn't encourage our work, they were still sufficiently uncertain about their own judgement not to interfere.

I am happy to report that the program is now indeed operational and a large governmental library is negotiating to install it for experimental purposes. I would like to give you some indication of the amount of work that was required to obtain even this amount of information. IBM supplied six people; the university supplied an equal number and these 12 people worked for a total of three months to collect the data.

Now in order to persuade IBM or the libraries that this program was desirable and worthwhile we looked for ways in which money could be saved. Money in this case means salaries. We looked for all those operations in the library that were duplications of effort. For example, if a card had to be inserted into the main catalogue and into several other catalogues, we considered that a duplication of effort. However, there was one other consideration; namely, that if operation was mechanized, it would indeed improve the overall performance of the system.

The reason for having six IBMers and six librarians working on the survey was to make sure that we would come to some common agreement about what was duplication and what was an improvement of the system. We came to the conclusion that in this particular library, approximately 40% of the salaries had a potential for cost saving, which could then be applied to the cost of the mechanized system.

For the last 15 years we have been working on information retrieval. We are continuing to do that, but at some point it seemed desirable to do something about the repository of information rather than the content of information. If we could successfully mechanize a library and find out where the information is and a little on what the information content is, then we might have less difficulty in mechanizing the contents of books for retrieval. I think that was a fairly mundane kind of approach and not one that was very popular in the company. However, a few of us felt that until such a data base was available for information scientists to manipulate, it would be very difficult for information science to progress. I agree that this position has yet to be proved, but at least now we have a system which will allow this to be investigated.

The system is operational and hopefully as a consequence of the negotiations being undertaken with one or two libraries in the US, we will be able to report at a later date that this system has not only been successful in terms of its function, but that a large percentage of the costs have been identified and that it is available for manipulation by information scientists.

MR. UHLMANN: To Dr. Batten. I have some experience which borders on what you said, but unfortunately most of the commercially available data bases are made for purposes other than searching. They are mainly made for producing printed journals and this fact makes their use for international retrieval sometimes very awkward. One is occasionally faced with the problem which can be compared to that of the reader of a book without an index; to find what one wants one has to read through the whole book. Besides, data bases often contain much information which is quite useless for information retrieval. I know of one fairly big data base which consists of only 50% of information bearing characters. In other words, you have to search 20 million characters per year of which 10 million characters you don't need. Also, many, or most of the commercial services object strongly to transfer information from their tapes to others for redistribution.

DR. BATTEN: To Mr. Uhlmann. You referred to the fact that commercial data bases are often full of needless characters which I suppose are there for no better purpose than telling the computer that it has to underline at some point or to leave two spaces or something like that. Of course, it is wrong that when a base is put on the market it should be cluttered up. This is just the sort of task which the "information supermarkets" should do. I'm sorry to press the analogy still further, but it is their task to take the sulphur out of the crude oil. Many of the commercial data bases were not designed for services, but neither was the crude oil provided by providence in a form suitable for driving jet engines.

Many of the commercial bases indeed have many of the imperfections of raw materials because often they have been created for other purposes. But it's still good that they exist. Let us do what we can with them rather than start entirely anew.

MR. UHLMANN: To Prof. Kochen. The data applying mathematical thinking (mathematical systems and operations) in the design of information systems has been largely neglected. Some attempts have been made in Sweden, but I think it has been done too late.

MR. C. KEREN (Israel): I would like to introduce another aspect of the economics of information systems which has been rather neglected, e.g., the economics of the information system as a part of the input which goes into making a product, be it research or a physical product. We have to decide what kind of an approach to take, how deeply to index, how much to abstract - in short, how much input do we put into what we want to produce.

It has been said that minimizing information means maximizing risk, and I know that we all more or less intuitively decide on which point of this scale we stand. But we need much more information about the economics of information systems in order to decide where to stop. The results of such research would probably weigh heavily in deciding which techniques we should employ in order to achieve our purposes.

DR. BATTEN: *To Mr. Keren.* The total logistics and economics of our large information problems are frightening in their complexity unless one is prepared to be satisfied with a fairly pragmatic solution. You are absolutely right. We do know the optimum form of item representation. We do not know the optimum nature of indexing; we do not know the proper admixture of classification techniques and of indexing techniques and of correlation techniques. What an item means to one person is not what it means to another and I am not talking about polemic assessment where it is a case of saying I think the man is wrong or I think the man is right. I am talking about the kind of creative thought that the reading of that contribution is going to produce in different people.

MR. S. ISAACSON *(Israel): To Dr. Batten.* Your analogy to the grocery store and supermarket is a good one. In the US, the Department of Defence adopted this concept many years ago in their specialized information centers.

I think what is going to happen in the information field of the future is that we are going to have many different levels of the grocery chain. For example, there is going to be the wholesaler who sorts the information into very large batches that overlap considerably. We are also going to have some specialized retailers who provide very highly specialized information. It is what is happening with the highly specialized doctors. However, we don't want to become so specialized that when you call an eye doctor, he asks "is it the right eye or the left eye?"

DR. E. HOFFMANN *(Israel):* We have been acquainted with two approaches - one presented by Dr. Batten which is to set up national information systems for specific purposes, because the material is too large to be coped with otherwise in a cost effective way. The other approach presented by Prof. Kochen, is the building of mathematical models of information systems and applying mathematical theory to evaluate cost effectiveness. However, there is a third approach which has been neglected. There do exist cost effective information systems in the commercial community. Large chemical companies, for example, in the US, have information systems of their own which are definitely considered cost effective. It might be a good idea to analyze these systems.

DR. BATTEN: *To Dr. Hoffmann.* I entirely agree with you that there are cost effective systems in industry which would well repay study. Let me however add a word of caution. I was associated with a fairly large industrial organization for a number of years and I know of one example there of information system which was cost effective for no other reason than the fact that owing to the tenets of accountancy and financing in that particular company, that particular service never had to pay for any of its duplicating charges and consequently of course the product looked very cheap, but it jolly well wasn't.

PROF. KOCHEN: *To Dr. Hoffmann.* Regarding the cost benefit figures of business enterprises. What comprises this knowledge industry? You will find the components of that in a book by Machlup "Production and Distribution of Knowledge in the United States" (Princeton University Press, 1962) in which he shows that the "knowledge industry" constitutes about one third of the GNP. Machlup gives direct cost prices rather than costs and there is an important economic difference; you get profits rather than utility. For the most part you get information about the private sector which does things the way it was done yesterday, so you will get a somewhat distorted picture.

If you really go into the basic problem of measuring benefits you will have to estimate and project and create demands and try to understand what is the central nature of information and be specific about what kind of information one wants to get. Production of magazines such as Time, Inc. produces, or, on the other hand, those providing agricultural information or information about water, are in the public sector where the figures on cost and benefit are quite different because they are government subsidized.

MR. W.K. LOWRY (US): I would like to mention some of the problems which we who have large information organizations are concerned with, as far as the economics go. The variety of data bases that are now being developed for computer processing are making choices essential. This choice is not always desirable, particularly if the user is interested in many disciplines. The alternative to making choices is to develop new data bases. Since this is an international conference, it seems quite appropriate to enter a plea here that the developers of large data bases not increase the size of their data bases indefinitely, each doing the same thing, thus making the choice more difficult and the cost higher. There should be some sort of standard worked out so that not only one, but several data bases could be used.

MR. UHLMANN: There is probably an optimum size for a data base because if I have one enormous interdisciplinary base which covers everything, I am obviously forced to search every time for a needle in a haystack. On the other hand, if a data base is too small there is the danger of missing something in the border fields and one is forced to search through several bases. There are some attempts at international cooperation which hopefully will make it possible to find a suitable compromise. I know about one because I am concerned with it. There is an international research documentation service which was established about eight years ago to distribute documentation sheets. A year ago OECD called a meeting to decide what is to be done with this service in the future. A format on magnetic tape was proposed. It was further proposed to design documentation sheets so that they may be used directly as input for the system. The communications format is roughly identical with the proposed ISO format. Distribution of tapes is to start on January 1, 1972. This is a small data base with only about 12,000 documents a year, but I think it is a successful international venture.

DR. BATTEN: To Mr. Uhlmann. You thought that there might be an optimum size for a data base and I think you shuddered a little at the overlarge one. I entirely agree with you. When I spoke of supermarkets, I certainly didn't visualize one gigantic fish pond containing all the sharks and all the minows. I thought that having formed the pond, one would then form from it a number of puddles which might well change their shape and form from time to time, but little puddles to which small chaps like you and me could go with little fishing lines.

PROF. H. BORKO (US): To Mr. Lebowitz. Regarding the charging algorithm you suggest, I understand that the more highly developed and richer the country, the higher the percentage of the cost it should pay. It's sort of a graduated income tax. I don't see it in terms of the input of the amount of research. I would rather see it in terms of the total budget proportional to the budget spent on R&D in general. What you are saying is that the more research he puts into the system, the more he has got to pay, rather that the more money he funds for research the more he has to pay. There is an important difference.

MR. LEBOWITZ: To Prof. Borko. I tend to agree with you. I think that one of the rules that would have to be established would be that a country that participates would be expected to do so fully, with the exception perhaps of classified information.

MR. B. VICKERY (UK): To Mr. Zaaiman. I would like to support your plea that if people are doing costing of effort, they should report this in terms of time as well as money. From the point of view of comparability, the types of tasks that are reported should aim to be standardized. Often one gets very widely diverse figures for the time spent on cataloging a book because people's interpretations of what the task of cataloging means, vary a great deal. Have you found difficulty in comparing your figures with the figures of others because of this?

MR. ZAAIMAN: To Mr. Vickery. It is extremely difficult to make any comparisons at the moment. The only way that one can compare is to look at the figures published by other centers and find one which looks like your own and then try to make out what they actually did to achieve that figure. That is why detailed descriptions of tasks of the sort I tried to put into my own paper must be published. I think that in some cases they are not detailed enough. The problem with time, as with money, is that one gets widely divergent figures reported in the literature. In an article by Penner some time ago, costs were expressed in terms of money, and he indicated that these costs were calculated at ranges of $2.50 to $8.10 an hour.

MR. M. KELLER (US): To Mr. Zaaiman. I was very interested in your statement that a request for a document indicates a successful abstract. This is partly true in my own experience. A good abstract persuades me that I ought to read the original. On the other hand, this may be untrue since sometimes I have to go to the original text because the abstract is not good. For example, it may fail to give the number or sex or age or ethnicity of subjects or something doesn't add up or something seems ambiguous. So I don't know whether the request for original documents is automatically a good test of the quality of abstracting.

We should aim at developing standards for excellence in abstracting so that the abstract serves two purposes: to persuade the user of the abstract that this is a paper worth reading; or, on the other hand, to show that it would be a waste of time. For example, if the abstractor knows that in this subject you really cannot evaluate the experiment without knowing the sex of the subject and the author has not reported it and the editor publishes such a silly paper, I think it is very nice to put into brackets in the abstract "sex of subject not stated" so I wouldn't bother to read the paper.

MR. ZAAIMAN: To Mr. Keller. The question of successful abstracts is one wide open to criticism. We consider it a success if we have chosen an article which is of interest to Iscor and have summarized it in such a way that somebody wanted to read it. This is our only practical test at the moment. We have no further information as to how successful these abstracts really are. That would take a lot of research which we cannot go into. We have to apply some standard which works in a practical situation.

IV Economics of Information Systems

How much information do you put into the product? We should accept that information work is a service function and in many companies service functions are not costed. How can you cost a personnel department? You will find that 10-15 years ago personnel departments were record departments. Then management woke up to the fact that personnel is one of the most valuable raw materials so they started to hire psychologists and sociologists and others, and management does not expect these people to show profit. They want them to retain the staff; if staff turnover is too high, they will get critized. I think that information work is a service function that can hardly ever be really costed as far as the profit it shows. One has to deliver the service which is required at the lowest possible cost.

PROF. V. SESSIONS (US): *To Mr. Zaaiman.* You are to be congratulated on having presented probably one of the most uniquely valuable papers at the meeting. I would have liked to have heard some summaries by function; what were the managements' costs in total, not just the breakdown. I was glad to hear you devote time to such things as writing reports. These are things that people usually don't put down. However, I would have liked to see some summaries by general function and I would also have liked to see some R&D costs.

A great deal of discussion here has been on the assumption that you have a system that has somehow been miraculously given to you or developed under some funding and it is static from then on. This simply is not true. Even if you are not converting to a different type of computer, you are constantly finding new methods of editing and validating your data and you are inevitably going to have new data and new devices. You are going to have to train people, you are going to have certain feedback costs, etc., and I think that if at some time you redo your paper, some of this information should be included because these are costs that are usually neglected but they are extremely important.

MR. ZAAIMAN: *To Mrs. Sessions.* We have quite a lot of additional figures available and we can supply them to anyone interested. The amount of work that goes into a survey like this is something which we would not be able to undertake ourselves. Fortunately we have a large Systems and Procedures Division which undertook the study. I warn anybody who wants to undertake a study like this to be very careful as it has an extremely demoralizing effect on the staff. People think that certain things can't be measured. As a matter of fact, I know of a case where work performed by doctors is done according to time standards and done successfully, but librarians and information officers do not accept this and you have to prepare them very carefully and then still watch out for a lot of trouble.

MR. ISAACSON: *To Dr. Schutsaack.* In your paper, you cite a cost per abstract of $15-16. This seems rather high. In 1962 I headed an information center for a marketing company in Ohio, Florida. An abstract cost us 49¢. We were handling about 50,000 abstracts a year, servicing 15,000 engineers and scientists. But this difference in price may be partially due to the inflation throughout the world today.

I don't think it is unrealistic to have very highly structured international information centers dealing with special subjects. Last month we finished a study in plastics, looking at new products. Scanning the information in Israel alone occupied five people for about three months. In addition we bought a lot of information from highly specialized centers overseas.

I think it is well worth the money to get the information that you need in time and I think what has to be done in something like a group hospitalization plan for the world. You pay a certain amount to join and when you get sick you get certain services. The same can be done with information; you pay some minimum fee to become a member of the information society and when you need the information there will be a charge which you will consider worthwhile.

DR. U. SCHUTZSACK (Germany): To Mr. Isaacson. The figure of 49¢ is rather low, but I should like to say that in our figure of $15 are included not only the preparation of the abstract, but the data, handling by the computer, the printing, the distribution and so on. Besides, it is very difficult to compare figures from different centers. Some do in-depth indexing, some short abstracts, others informative abstracts and so on.

DR. J. ROTHSCHILD (Israel): To Mr Hirsch. Could you please compare your findings with the experience of some other large American university libraries. No. 8 of your bibliography is the publication of Fussler and Payne of the Chicago University Library. It should be interesting to hear a discussion about their experience.

MR. HIRSCH: To Dr. Rothschild. With regard to comparing our figures with figures of other reports, we have had the same difficulty that everybody else has in identifying all of these detailed activities, but in general I have observed that the Chicago report and ours would agree in general on the overall potential for mechanization savings.

MRS. A.M. DE BUSTAMANTE (Mexico): To Mr. Hirsch. Would saving 42% of the salaries of the personnel of a library by installing your system be enough to pay for the computer system, and what would be the advantages of putting the system in?

MR. HIRSCH: To Mrs. De Bustamante. Wherever possible, the activities were looked at in terms of how data was entered into the system. If data was entered more than once, that was considered duplication. As to whether the 42% mechanization savings potential is adequate to pay for the system, the answer is a qualified yes. Let me explain why it is qualified. In the particular university where this study was carried out, they did not have a computer that could be shared with another department. The library system that I have described in the paper is one which shares a computer. There are many ways in which a computer can be shared. It can be shared in terms of information being stored so that it can be retrieved quickly, or less expensively by it being retrieved a little slower.

In the case of fast retrieval it could be as if you and I were talking back and forth to each other all the time. In the case of slower retrieval the time delay may be as much as 20-30 seconds. This makes a considerable difference in the cost.

However, in terms of the salaries of that particular library of half a million volumes with additions at the rate of approximately 100,000 per annum, 42% of mechanization savings potential would have justified a computer on a share basis.