DOCUMENT RESUME

ED 061 946 LI 003 532

AUTHOR	Smith, William A., Ed.
TITLE	The Management of Information Analysis Centers: [Proceedings of a Forum] Held at National Bureau of Standards, Gaithersburg, Md., May 17-19, 1971.
INSTITUTION	ERIC Clearinghouse on Library and Information Sciences, Washington, D.C.
SPONS AGENCY	Federal Council for Science and Technology, Washington, D.C. Committee on Scientific and Technical Information.
REPORT NO	COSATI-72-1
PUB DATE	Jan 72
NOTE	203p.;(62 References)
EDRS PRICE	MF-$0.65 HC-$9.87
DESCRIPTORS	Abstracting; *Electronic Data Processing; Indexing; *Information Centers; *Information Processing; Information Retrieval; Information Services; Information Systems; Input Output Devices; *Management; *Marketing
IDENTIFIERS	*Scientific and Technical Information

ABSTRACT

The papers given at the May 17-19 forum on Management of Information Analysis Centers held at the National Bureau of Standards in Gaithersburg, Md. are presented in four sessions separated by topic. Besides the welcoming remarks and the keynote address session 1 contains three general papers on information analysis centers and automatic data processing, abstracting and indexing services, and markstem, inputting techniques, and computer usage in a large data center. The third session, on abstracting and indexing services, includes case studies of ERIC/CLIS and the Air Force Machinability Center, the role of secondary services, a profile of scientific-technical tape information services, NSIC computerized information techniques, and the trends in services. Session 4, marketing, includes case studies of DoD, Cooper Data Center, Plastics Technical Evaluation Center, and information analysis centers' liaison with professional organizations, commercial firms, and the private sector. (SM)

THE MANAGEMENT OF INFORMATION ANALYSIS CENTERS

Sponsored by
Panel on Information Analysis Centers of the Committee on Scientific and Technical Information (COSATI), Federal Council for Science and Technology

held at
National Bureau of Standards, Gaithersburg, Md.
May 17—19, 1971

William A. Smith, editor
Office of the Secretary, U.S. Dept. of the Interior

prepared by
ERIC Clearinghouse on Library and Information Sciences,
American Society for Information Science
1140 Connecticut Ave., N.W., Suite 804
Washington, D.C. 20036

with
partial support from the National Science Foundation,
Office of Science Information Services, Washington, D.C.

printed by
U. S. Atomic Energy Commission, Technical
Information Center, Oak Ridge, Tennessee 37830

January 1972

This publication was prepared pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official Office of Education position or policy.

Available from:
ERIC Document Reproduction Service,
LEASCO Information Products, Inc.,
P.O. Drawer O, Bethesda, MD 20014;
and
National Technical Information Service,
U.S. Dept. of Commerce,
5285 Port Royal Rd.,
Springfield, VA 22151.

CONTENTS

Contents—*Continued*

Contents—Continued

SESSION 4. 1:30-5:15, Tuesday, May 18, 1971

MARKETING

Chairman: Dr. Joseph F. Caponio, National Agricultural Library

MEMBERS OF PANEL ON INFORMATION ANALYSIS CENTERS

Dr. Edward L. Brady
Chairman, Panel 6
National Bureau of Standards
Washington, DC 20234

Mr. Harold B. Atkinson, Jr.
Executive Secretary, Panel 6
Office of Information Services
U.S. Atomic Energy Commission
Rm. P-602
Washington, DC 20545

Dr. Joseph Caponio
Associate Director
National Agricultural Library
U.S. Department of Agriculture
Beltsville, MD 20705

Mrs. Elsa S. Freeman
U.S. Department of HUD
451 7th St., S.W.
Washington, DC 20410

Dr. David Garvin
Physical Chemistry Division
National Bureau of Standards
Washington, DC 20234

Mr. Raymond A. Jensen
Water Resources Scientific Information Center
U.S. Department of the Interior
Washington, DC 20240

Mr. Harvey Marron, Chief
Educational Research Information Center
U.S. Office of Education (HEW)
400 Md. Ave., S.W. - Rm. 3013
Washington, DC 20202

Mr. Harry E. Pebly, Jr.
Plastics Technical Evaluation Center
Picatinny Arsenal
Dover, NJ 07801

Dr. Stephen A. Rossmassler
Office of Standard Reference Data
National Bureau of Standards
Washington, DC 20234

Mr. William A. Smith
Office of the Secretary
U.S. Department of the Interior
Washington, DC 20240

Mr. James R. Trew
Science & Technology Division
Library of Congress
Washington, DC 20540

Dr. Herman M. Weisman
Office of Standard Reference Data
National Bureau of Standards
Washington, DC 20234

Mr. James F. Williams, Chief
Library Branch, NA-41
National Aviation Facilities Experimental Center
Department of Transportation
Atlantic City, NJ 08405

CONSULTANTS

Dr. Francois Kertesz
Coordinator Information Centers
Oak Ridge National Laboratory
P.O. Box X
Oak Ridge, TE 37830

Mr. Gustavus S. Simpson, Jr.
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201

Mr. Charles De Vore
Liaison Member for Panel 1
Headquarters
Naval Material Command
836 Crystal Palace No. 6
Washington, DC 20360

LIST OF SPEAKERS AT THE FORUM ON THE MANAGEMENT OF INFORMATION ANALYSIS CENTERS
MAY 17, 18, and 19, 1971

Dr. Burton W. Adkinson
Director, American Geographic Society
Broadway at 156
New York, New York 10032

Col. Andrew A. Aines
Federal Council for Science and Technology-COSATI
Executive Office Bldg.
Washington, D. C. 20506

Mr. William T. Black
Copper Data Center
Battelle Memorial Institute
505 King Avenue
Columbus, Ohio 43201

Dr. Edward L. Brady
Assoc. Dir. - Info. Prog.
Nat'l. Bur. of Standards
Washington, D. C. 20234

Dr. L. M. Branscomb
Director
Nat. Bureau of Standards
Washington, D. C. 20234

Mr. William E. Burgess
CCM Information Corp.
909 Third Avenue
N. Y., N. Y. 10022

Mr. Walter C. Christensen
Director of Tech. Info.
ODDR&E
The Pentagon, Room 3D1021
Washington, D. C. 20301

Dr. Donald H. Coombs, Director
ERIC/EMT
Institute for Commun. Research
Cypress Hall
Stanford University
Standford, California 94305

Mr. William B. Cottrell
Oak Ridge Nat'l. Lab.
P. O. Box X
Oak Ridge, Tenn. 37830

Dr. Ruth M. Davis
National Bureau of Standards
Washington, D. C. 20234

Dr. David Garvin
Physical Chemistry Division
National Bureau of Standards
Washington, D. C. 20234

Mr. H. W. Koch
American Inst. of Physics
335 East 45th Street
N. Y., N. Y. 10017

Mr. Charles T. Meadow
COSATI
Executive Office Bldg.
Washington, D. C. 20506

Mr. Jeffrey Norton, Publisher
Information Services
Holt, Rinehart & Winston
383 Madison Avenue
N. Y., N. Y. 10017

Mr. Young Park
ERIC/JC
Room 96, Powell Library
Univ. of California
405 Hilgard Avenue
Los Angeles, California 90024

Mr. Harry E. Pebly, Jr.
Plastics Tech. Eval. Center
Picatinny Arsenal
Dover, New Jersey 07801

Mr. Byron Riegel
G. D. Searle & Co.
P. O. Box 5110
Chicago, Illinois 60680

Dr. Vladimir Slamecka, Director
School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Georgia 30332

Mr. Joshua I. Smith, Associate Director
ERIC/CLIS
1140 Connecticut Avenue, N. W.
Suite 804
Washington, D. C. 20036

Mr. Fred Tate
Chemical Abstracts Service
University Post Office
Columbus, Ohio 43210

Mr. Walter H. Veazie, Head
Electronic Properties Information Center
Hughes Aircraft Corporation
Aerospace Group
Culver City, California 90230

Dr. James Vette, Director
National Space Science Data
Goddard Space Flight Center
Code 601
Greenbelt, Maryland 20771

FOREWORD

Scientific and technological progress are today the major underlying forces of economic and social growth. They afford great stimulus to viable economies, improved standards of living, adequate health care, effective transportation, and an international communication network. Continuing progress requires a continuing increase of our understanding of the physical world that surrounds us, and of the laws of nature which govern it. Such understanding is the goal of scientific research, valid for both the physical world and the social world. The companion effort, the application of understanding in the solution of practical problems, constitutes technology.

One key to the effectiveness of science and technology is their ability to apply yesterday's discoveries to today's problems. The warehouse of knowledge already gained - the scientific and technical literature - is a major national and global asset. But using that asset requires locating the relevant information packages in the warehouse and matching them to the problem at hand. Since the warehouse acquires more than two million pages of new information each year, and since the content is not always fully identified, the user frequently requires help. Moreover, information quality is not uniform, and the wise user seeks assurance of the accuracy of the information he uses.

Meeting these user needs-accessibility of relevant information, and assurance of its quality, is the reason for existence of the Information Analysis Center.

The Committee on Scientific and Technical Information (COSATI) and its parent organization, the Federal Council for Science and Technology, have been keenly aware of the vital role of Information Analysis Centers. In 1967 COSATI established a Panel on Information Analysis Centers to provide government-wide focus on their functions and problems.

In the first year of its existence, this Panel sponsored a Forum to help information analysis center representatives to exchange ideas with one another, and with their sponsors and other government program officials. The Proceedings of that Forum, held November 7-8, 1967, have been published and are available from the National Technical Information Service as document number PB 177051.

Continuing explorations during the following four years sharpened the Panel's awareness of the technical problems and opportunities faced by Information Analysis Centers. We also recognized that government-wide requirements for cost recovery in information programs were becoming a major concern for Information Analysis Centers. After receiving a strong favorable recommendation from a canvass of a number of Information Analysis Centers, Panel 6 decided to sponsor a second Forum in the spring of 1971. The theme selected was the management of information analysis centers, with particular attention to key problems identified by center managers.

The Forum was held on May 17, 18, and 19, 1971, at the National Bureau of Standards in Gaithersburg, Maryland. The first session, an overview, featured a keynote address by Dr. Lewis M. Branscomb, Director of NBS. Also included were introductory comments on the three major problem areas which composed the subsequent three sessions of the Forum. Dr. Ruth M. Davis spoke on Information Analysis Centers and automatic data processing. Dr. Byron Riegel commented on

Information Analysis Centers and abstracting and indexing services. Dr. H. W. Koch surveyed the problems of marketing in relation to Information Analysis Centers.

The afternoon session on May 17 provided a detailed examination of automatic data processing operations and applications. Following dinner that evening, the Honorable James H. Wakelin, Jr., Assistant Secretary of Commerce for Science and Technology, addressed the group on the national stake in better technical information. The next morning and afternoon sessions resumed close scrutiny of important problem areas - the use of abstracting and indexing services in information analysis centers, and marketing. The final session, on the morning of May 19th, was given over to tours of information analysis centers in the Washington area, demonstrations of computerized operations, and meetings of several common-interest groups under sponsorship of individual Federal agencies.

The members of the COSATI Panel on Information Analysis Centers consider that this Forum provided important understanding to center managers about future developments in acquiring, handling, and disseminating technical information relevant to their missions. The 1971 Forum may also be regarded as a fulfillment of promises, made at the 1967 Forum, for closer attention to problems which the earlier meeting defined, but did not study in detail.

The Proceedings which follow record the discussions that took place on May 17 and 18, 1971. They contain much practical advice beneficial to all managers of technical information activities, whether or not they attended the 1971 Forum.

E. L. Brady, Chairman
COSATI Panel on Information
Analysis Centers

WELCOMING REMARKS

Col. Andrew A. Aines
Office of Science & Technology*

Ladies and Gentlemen:

Let me open my remarks by stating that I am delighted to see you all here today. It is evidence that the conference is deemed to be important by all of you who operate or sponsor information analysis centers. It is also good to see so many survivors of a grim period for science and technology. Despite our troubles, it betokens, it seems to me, a growing interest and understanding on the part of management. If this were not the case, there would be fewer information analysis centers in existence.

Again, I would like to extend to Dr. Branscomb the thanks of COSATI for making available this beautiful facility for this meeting. The National Bureau of Standards by any way of reckoning is one of the most outstanding laboratories in the world. Its receptivity to, and its participation in, the crusade for better dissemination and use of scientific and technical knowledge is second to none. Its assistance to the Science Adviser to the President in both his Office of Science and Technology and Federal Council Programs has brought me great pleasure and the country even greater profit.

Since the first Conference on Information Analysis Centers a few short years back, there has been considerable progress for IACs.

This is evident in the growing number of centers, in the increasing interest in IACs in other countries, in the introduction of management principles that will, in the long run, strengthen the position and the contributions of IACs, and in a relative way, the growing understanding of the promise of IACs by scientists, engineers, and managers.

To a man, Presidential Science Advisers and their staff people have been staunch supporters of IACs, but, in recent months, I have been strengthened in my belief that they will play an even more vital role in the future. They are a logical intellectual extension or balance to growing mechanization of information and data systems. They will help us in the task of information and data utilization as well as screening and compacting the literature. I used to think of them earlier as being most useful at the cutting edge of science, but I have become convinced that they can make great contributions in the solution of the complex problems of society, certainly to aid decision-makers and problem-solvers, as well as scientists and engineers.

It is my hope that you have assembled - you leaders in the IAC community - with well-defined, hard-boiled objectives; that you will make it a serious meeting of serious people with serious problems. I hope that you will ask such questions as: Where are we now? Where do we want to go? What obstacles and problems do we have to overcome? What actions do we need to take as individuals? What progress can we make via group action? I am not suggesting that you wear hair

*Now located at the Office of Science Information Service, National Science Foundation, Washington, D. C.

shirts and engage in self-flagellation while you are here. I hope that this will be a meeting to enjoy because all of the vibrations will be favorable to new insights and progress.

So I will conclude my brief welcoming remarks with the friendly challenge to make this Conference a great success, a high watermark of accomplishment. I urge you to achieve new insights and understanding so that when you leave on Wednesday, you will carry away with you the feeling of certainty that you have made a seven-league leap forward, and that you return to your centers refreshed and ready to make successful inroads on your problems. Good luck. Have a wonderful conference, ladies and gentlemen.

KEYNOTE ADDRESS

Lewis M. Branscomb
Director,
National Bureau of Standards

INFORMATION ANALYSIS CENTERS: THE CHALLENGE OF BEING NEEDED

"The wise man does not act without attempting to know the consequences of his actions. Contemporary societies must be more prudent in their actions if technology is to be a boon rather than a curse for mankind. Information is the key to the wise management of our future."

"Perhaps the most important event of the next decade will be the recognition of the true value of information—the right information, reliable and relevant to our needs, available in useful form to all those who need it."

So begins a report to the Secretary General of the OECD entitled "Information for a Changing Society: Some Policy Considerations.[1]" No better case need be made for the importance and future potential of information analysis centers.

I will not attempt to define what I mean by an information analysis center. COSATI must have an official definition that satisfies its needs. Let me say only that such a center is a contemporary institutional mechanism for organizing, evaluating and making available the numerical and phenomenological information which results from research and observation and which is needed by people other than those who generated it. I will restrict myself to science and technology since COSATI is similarly oriented. But neither I, nor COSATI, believe that such a restriction implies that scientific and technical information is always fully useful to decision makers unless they also have associated economic and social data. Indeed the most useful contribution of the information analysis centers in science and technology may well be to demonstrate the importance and practicality of achieving objectivity and credibility in the effective utilization of organized information. The opportunity for the social sciences to contribute decisively to rational decision making in public policy depends critically on their developing similar capabilities.

The information analysis center serves as the cerebral cortex of the technical nervous system. In the human body, a signal from eye or ear calls up a search for stored information, which must be selected for its relevance and its reliability. The output signal activates the appropriate muscles and produces whatever action is required. Similarly, the information analysis center couples the unassimilated knowledge of basic science to our technological muscles. When the brain works well, we take it pretty much for granted. When it works poorly, walking down steps or riding a bicycle can be a terrifying experience. In technology it is not quite so obvious that the cortex is too small, is deprived of oxygen, and is coupled to only a small fraction of the brain's memory cells. Why?

First, our society is—from a technological point of view—at a very primitive stage of evolution. We are just beginning to crawl up out of the sea, so to speak. We accept that research is a high risk proposition and that scientific creativity is a delicate flower. But we fail to realize that this is no excuse for failing to organize our science and technology system into a functioning whole. Nor is it

[1]Unpublished at this writing. Publication by OECD anticipated during 1971.

an excuse for failing to introduce improved quality assessment into science. Most people expect the data they use to be uncertain, and so place minimum reliance on it. It would be very interesting to know how many small scale pilot plants might have been unnecessary if there had been an adequate basis for confidence in the theoretical prediction of the efficiency of a full scale system.

Unfortunately, this skepticism about the reliability of published data is well justified. If we need a reminder, there is the story of the rocket fuel production plant that was shut down - after the reported expenditure of 200 million dollars - when it was found that its process was based on an erroneous value of the heat of formation of a light metal oxide.

Second, the provision of stimulation for civilian industrial technology characteristic of the science policy of the past decade has depended too much on "trickle down" and "drip off." The Federal Government buys research and development (R&D) to get technology for such government operations as defense and space exploration. "Spin off" benefits to the private sector—or "drip off" as Ralph Lapp calls them—are often incidental to this investment. Agencies like the Commerce Department's National Technical Information Service and the Smithsonian's Science Information Exchange are admirable means for facilitating access to the publications and ongoing projects of this federal effort. But the mission-oriented Federal R&D system is not intended to focus primarily on provision of useful data to the public. There are, of course, some excellent exceptions - represented by the information analysis centers that do exist. In the absence of enough information analysis centers, those who need data produced in government projects must search the project literature for it.

"Trickle down" refers to the weakness of the mechanisms for ensuring that the results of our $2 billion national investment in basic research find their way into applicable technological choice. The Federal Government pays for most of the research but too often it stops short of accepting responsibility for the effective evaluation, secondary processing, and dissemination of the results. If this responsibility were accepted we would have to put the horse back in front of the cart by focusing applied research on information needs, buttressing this investment by appropriate new research to ensure availability of information not yet in existence.

The third reason that information analysis centers have been undervalued is reflected in the quotation I read from the OECD report. People are not accustomed to placing a realistic value on information. Part of the problem, of course, is that information is used in decisions, be they managerial or technological. How do you place a dollar value on a decision? We all know that a bad decision can be very expensive - witness the experimental rocket fuel plant mentioned earlier. But there is no established economic measure for decision quality. How then can we put a price on information? Perhaps this ability will come in time from decision theory, which provides a technique for quantifying the value of information. But the more traditional answer is that information can be priced as other commodities are priced - in the marketplace. But the ability of information analysis centers to command a good but appropriate price for their product is limited by a number of factors:

- Inadequate economies of scale resulting from reaching too small a fraction of the potential market;
- Traditionalist attitudes of the technical community toward information transfer mechanisms of new kinds, combined with;

- A tenacious and well justified desire of scientists to reserve their dependence on information sources to those sources whose continuous availability and quality are reasonably well assured;
- Less than full confidence in the reliability of the information products offered, so that the user does not risk defraying a substantial cost for information with an even greater cost avoidance achieved by relying on it;
- Less than fully effective marketing of information analysis center products, combined with the fact that information is more efficiently and effectively wholesaled than retailed. Thus economic return depends on intermediate institutions such as libraries, which are not in a position to recover or even measure economic benefits from good information.

All these handicaps are made greater by the fact that the universities do virtually nothing to prepare their students for contemporary innovations in the evaluation and handling of knowledge—a rather striking indictment of education since evaluation and transfer of knowledge and experience is what education is all about in the first place. In addition, we have not yet had enough experience at identifying potential clients for information analysis services—other than the data generating community with which most centers are in close contact. A NASA-sponsored study by Denver Research Institute[2] showed that design and production engineers in commercial enterprises of moderate technological sophistication get their technical information primarily from commercial product sales literature and sales representatives. Government publications ranked near the bottom of the list in significance as an information source. Government sponsored science and technology information transfer programs have yet to reach the majority of industrial uses outside the R and D community.

Finally, the intellectual challenge of information center activity has not yet been fully recognized by the peers of those who engage in it. Nevertheless, the information analysis center is an old idea whose time has come. The information analysis center will prove indispensable as a means to make scientific knowledge quickly available to policy makers in a useful form. It will thus be a major factor for ensuring that technology is wisely used for human benefit. The impact of technology on society now moves at so swift a pace that there is no longer enough time—or at least no longer enough patience—for research to be launched from scratch when hard choices have to be made. The dilemma facing those who must make policy on the use of washing detergents is an excellent example. We simply do not know enough today about alternative chemicals and their environmental effects, their efficiency as laundering agents, their economic impact on washing machine design and life. All this information needs to have been gathered, evaluated, and made available yesterday. The resultant cost of making the wrong decision with inadequate information might well run into the hundreds of millions.

It sometimes happens that there is no choice between time and economic cost, for the time to get new information is intrinsically limited. When Apollo 13 failed on its trip to the moon, the cause of the explosion had to be diagnosed and corrective action taken before the next mission could be launched. In this case, NBS cryogenic engineers were able to help NASA track down the cause. Some very accurate, evaluated data on the thermodynamic properties of liquid oxygen - available from the

---

[2]"Commercial Application of Missile Space Technology" by John G. Wells, L. G. Marts, et. al.; Denver Research Institute Report No. N-64-24335; 1963, 262 pages.

NSRDS center in Boulder - were decisive in choosing correctly between two alternative chains of events that might have characterized the failure.

Indeed, the course of science and technology is strongly influenced by the information base on hand at the time of a new conceptual break through. NBS established its Atomic Energy Levels program several decades ago, largely because of the needs of basic research in atomic structure and in astronomy. In time, thanks to the great work of Charlotte Moore Sitterly and her colleagues, this center achieved a position as authoritative focus for the analysis of atomic spectral data. In the '40's and '50's there were times when budget makers questioned the steady drain of investment in this large effort. But suddenly there was the MASER and the evident possibility of making an optical oscillator through coherently stimulated radiation emission: the LASER. In 1966, Professor W. R. Bennett of Yale said[3], "The three volumes of critically-compiled data on Atomic Energy Levels (NBS Circular No. 467) played an essential role in the development of the gas laser. Without the existence of these data the development of the gas laser field ... would have been delayed for many years." What economic benefit shall we ascribe to this application alone? The laser industry is growing exponentially, and is now well beyond the $100 million level. Even if the Atomic Energy Levels program brought the laser only one year sooner, the carrying charge on the fraction of the national debt equal to the annual taxes paid by the new laser industry would have paid for operation of the atomic energy levels program for a decade.

Timeliness of data availability is indeed one of the most important values of the information analysis center. But in my personal view, quality and reliability enhancement are by all odds the most important benefits that flow from well managed information analysis centers. I suspect they also pose the most subtle and difficult problems for information analysis center managers. Accordingly, I am a little surprised to note that this management problem does not seem to be on the agenda for this meeting.

I will illustrate the role of reliability by discussing numerical data evaluation, but this principle can be extended to nonqualitative forms of information. First let me call to your attention a small and informal symposium being sponsored in this auditorium on July 21, 1971, by the U.S. National Committee for CODATA. The occasion is the CODATA General Assembly in Washington, and the topic will be a discussion of criteria for data evaluation. We will have a panel discussion - perhaps a debate - involving information analysis center managers and primary journal editors. The meeting is open to the public, and I know Dr. J. Ross MacDonald, Chairman of the Numerical Data Advisory Board of the National Academy of Sciences would want me to invite each of you to attend if you can.

The manner in which data are evaluated determines the impact of that evaluation on the primary scientific work to follow. The traditional, or *ex cathedra* method consists of asking an expert (a species for which there is as yet no objective performance criterion) to select the "good" from the "bad." In the absence of an information analysis center, many scientists do not use the results of unknown younger scientists until those results have been used in the work of one of the "great men of the field." The frequency with which original data are referenced by citation of a theoretical paper by a famous man who used the datum himself, rather than by citation of the original research paper, is an indication of this tendency.

[3]1966 Letter to Dr. L. M. Branscomb

But certification by imprimatur is neither objective nor fair; nor does it lend itself to the identification of measures of reliability. Unless quantitative statements can be made about precision and systematic errors, you cannot expect users to place reliance on the information, for they cannot compute the risk they are taking by using it.

A second alternative to *ex cathedra* evaluation is the Delphi or consensus method. Those of you familiar with technological forecasting will know that the Delphi method is regarded as a great improvement on the educated guess as a means for predicting the future of technical events. It consists of the projection into the future of the average of a number of educated guesses by people whose education differs at least in some respect. But a committee gives no more assurance of an objective reliable result than does a high priest of science. Indeed, the pressures of a committee will work for inclusion of everyone's data whether these data are valid or not.

The preferred approach is to apply the principles of scientific objectivity to the evaluation process, and to ensure that the individuals applying it have demonstrated competence sufficient to the task. This process calls for criteria by which data are to be evaluated. If the data are derived from experiments, these criteria cannot be developed without a complete theory for the experiment, giving systematic error phenomena equal weight and attention with the phenomenon under investigation. (It is we scientists who chose to assign the phrase "experimental result" to one phenomenon and "systematic error" to other phenomena in an experiment. Nature is unaware of our prejudice in this regard, and holds all phenomena in equal esteem. She sometimes punishes us for our narrow vision by letting us delude ourselves about what the experiment is we have actually performed.)

Thus the criteria for evaluation of experimental results become an algorithm for doing a valid experiment in the first place. Information centers are obligated to publish the criteria they use to referee literature. Those that have won the respect of both users and data generators will find that prudent future investigators will take more care when new work is done, not only to do a valid experiment but to publish the evidence that their errors were under control.

The impact of information analysis center operations on fundamental science is well documented. For example, the landmark paper issued by Lee Kieffer and Gordon Dunn in *Reviews of Modern Physics*[4] discussing the state-of-the-art in reliable reference data on cross sections for ionization in electron collisions has been analyzed. The article in question appeared in 1966. Since then, many authors have cited this article in their own papers. Fifty-three of these articles were read and analyzed as to what the impact of the citation was. Twelve of the articles made reference to the Kieffer-Dunn article for general citation, background purposes, or newstype items. Twenty-two articles cited Kieffer and Dunn in making use of, or referring to, the data which Kieffer and Dunn presented, for purposes of computation or comparison between experiment and theory, for comparison between one experimental value and another, or for calibration purposes. Nineteen articles explicity recognized the main conclusion of Kieffer and Dunn, that a very large fraction of the previously reported results in the field of electron collision cross sections were deficient in their reporting of experimental data and in the analysis of systematic errors. These last nineteen articles made clear identification of their own attempts to avoid such inadequacies and to present their own data in a reliable and meaningful form. Thus, it is clear that over 1/3 of a significant group of

---

[4] "Electron Impact Ionization Cross-Section Data for Atoms, Atomic Ions and Diatomic Molecules; I Experimental Data" *Reviews of Modern Physics*, Volume 38, No. 1, pp 1-35, Jan. 66.

references to this article showed that research had been influenced by the output of an information analysis center, improving the future input.

Admittedly, the numerically specifiable properties of elementary substances are most suitable for this kind of objective evaluation and for the specification of accuracy values. Such data are the concern of many of the information analysis centers of the National Standard Reference Data System, coordinated at NBS. But it is a generally valid principle that managers of all kinds of centers should strive in this same direction. Not only will such procedures have more favorable impact on the quality of data sources, but they are also essential to insure a broad and continued acceptability of the products of the center.

What can we say about the role of automatic data processing in the information analysis center operation? I would not wish to duplicate any of the discussion already scheduled on this topic. Let me only note that if we view the information service business as a service industry and ask what we can expect in the way of productivity increases, we see an immediate role for the computer. And indeed, the computer is being widely applied to vastly increase the output of product - at constant product nature and quality - per unit of input labor. Most centers use computers for internal operations - storing, searching, and retrieving bibliographies and often data files as well. Some use computerized access to indexing and abstracting services. Some use computers in output or product preparation—as in computer type-setting. But only a few, such as MEDLARS, use teleprocessing as a means for dissemination and marketing. In my view, the reasons are rooted in simple economics and in the handicaps of traditionalism noted above. I am certain that the day is coming when real-time access to evaluated data files will be at the fingertips of the individual scientist. Research efforts to hasten the day when this will be practical and economic are well justified. But, I also believe that the surest way to kill the concept of information analysis centers is to oversell present market demand for their products and to force them into symbolic if not actual bankruptcy by tying their viability to excessive requirements for capital investment in communications and software at a premature stage.

But the computer can also increase the effective productivity of information analysis centers by even more important qualitative changes. Computers permit the application of quantitative criteria for data validity of much greater complexity than those normally applied by the subjective evaluator. X-Ray crystallographers today already have in operation a system for subjecting the data in proposed publications to a computer's evaluation. Thus, journal editors have added another referee to their staffs, one of great perseverance and undeniable objectivity - a computer. The extension of this principle may bring the day when a quantitative method for data evaluation can provide guidance for the synthetic creation of new knowledge.

Theorists are beginning to develop general purpose programs for computing arbitrary numbers of useful quantities in terms of input parameters that can be specified by the users. Rather than applying the program to a few published cases and destroying it, the information analysis center keeps these programs as powerful tools for producing new data on demand. As the center also comes into possession of thousands of experimental data values, it can update the accuracy of much of these precise, but less accurate, data when new benchmark measurements of great accuracy are made. By combining such normalized data with theoretical programs parametrically dependent on the data, thus synthesizing new knowledge, the information analysis center will become a focus for analytical research of a new kind. As the error control in the data banks of natural properties improves, the information noise level represented by today's data uncertainty will drop. And with improved information signal-to-noise level provided by improving selectivity, phenomenology contained but concealed in the data will begin to emerge into view.

The emerging role of the information analysis center in dealing with basic physical properties as a source of new knowledge is perhaps only dimly seen at this time. If we look to the information analysis center dealing with information on much more poorly characterized systems - such as data on the incidence and circumstances of building fires - we see that the information analysis center creates valuable new knowledge today. Nowhere is this better illustrated than in information centers designed to identify the needs for government regulation of technology.

It is relatively easy to prove that there are too many accidental fires, and too many people are killed. It is easy to show that thousands are hurt by needlessly dangerous household products. But it is often very difficult for the official with responsibility for setting mandatory standards to identify the chain of events that leads from exposure to risk to the initiation of a dangerous set of events and then to the final tragic injury. Only well organized systems for acquiring data on injury, on the frequency of exposure to risk, and on the nature of individual vulnerability, and systems for the critical evaluation of such data can provide the degree of confidence needed to be sure the mandatory rule will in fact lead to a reduction of injury. It is urgent that this nation insure its ability to regulate technology in a rational way that does indeed satisfy public expectations of benefit which are expressed in the new authorities being provided by Congress. These new authorities, when exercised, drive up costs and limit technological alternatives. If a better environment, safer cars, toys, and household products, uncontaminated food and less risk of fire do not result, the present unhappiness concerning science and technology could become a rebellion. A large array of increasingly sophisticated information evaluation centers are required for such purposes.

I hope so far I have convinced you of the value, potential, and present need for information analysis centers. Perhaps, as managers of information analysis centers, you weren't very hard to convince. Let me now address briefly the question:

How can you measure the success of an IAC?

I would ask three sets of questions:

Of users, I would ask: Tell me in what way you rely on the information from this center to a greater extent that you would on access to the original literature from which the information came? What would you have done instead if the center's product had not been available to you? What are you prepared to do—including paying money—to insure your continued access?

Of the center itself, I would ask: What evidence can you show me that the information you are now receiving is of better quality as a result of your prior work? To what extent have the demands of your customers reflected themselves in the priorities of the data generators who feed you their material? How much cooperation do you get from the data generators, whose scientific reputations you hold in trust when you evaluate their data?

And of the R&D policy makers in government and industry, I would ask: How confident are you that the research you are buying reaches those who need it in an optimum fashion for its effective utilization? Is your own access to the information you need for decisions being provided for? What fraction of the market for evaluated information is being satisfied?

What does it take for a Center to be successful? I believe six requirements are necessary and sufficient:

1) Competence: Reflected in the continued involvement of the Center's intellectual leadership in creative work and thus in the scholarship of the evaluation done.

2) Continuity: A long-term commitment to generating confidence by the user and the data generating communities in the competence of the team.

3) Completeness: The user must know within the scope of his inquiry that the Center's coverage is complete; otherwise he cannot assess his risk in relying totally on the Center's information product.

4) Conscience: Realizing the plight of the user from another discipline who cannot evaluate the information even if he could find it, and the fate of the data generator whose paper will never be read by the scientists of tomorrow, because they will instead rely on and refer to the evaluated output of the Center.

5) Cash: To finance the very demanding and expensive scholarship without which the primary values of information analysis centers are lost. This cost cannot be passed on to the retail customer.

How big is the job, and how much cash will it take? Today there are over 500,000 scientists and engineers employed in a $27 billion national R&D enterprise. Over 50,000 of these are in basic research (3.8 billion). I have no idea how many are engaged in generating quantitative, storable information in the public domain, but let's suppose the number is similar to those supported by the total basic research budget. A good review covers careful study of perhaps 50 to 100 papers on an average, of which about 10% might have been contributed in the preceding year. It takes at least as long to do the research for such an evaluation and review as it does to prepare one of the original research investigations (about a year on average). If one updates the criteria for evaluation every 2 to 5 years, we need from 2 to 5% of the data generating workforce engaged in data analysis evaluation and review. This might be somewhere in the ball park of $40 - to $100 million per annum.

Now that estimate is not a budget justification, because it is an input investment, not a measure of product value. But it does let me ask: Should we view information analysis center activity as a necessary adjunct to our national long-range research investment? Should we tax basic research 2 to 5% to provide for evaluation and preparation for use of their data? A good case can be made for this approach. But if the R&D user community were prepared to pay from 0.08% to 0.2% of their costs for organized, reliable information the information analysis centers would be well provided for. If we can then find ways to overcome the present handicaps faced by the Centers and finance their production and marketing operations in a reasonable way, we may so upgrade the regard with which good information is held that we generate a thirst for more basic research to feed the information system. When the IAC's thus become the main force for generating public support for increased financial support for basic research, the informational evaluators can stop worrying about being second class citizens in the scientific community.

I would like to close my remarks as I opened them, with a quotation from the OECD report referred to earlier. While the advice was originally intended for all OECD member governments, I believe it should be taken particularly to heart by our own. I quote from selected paragraphs of the report's conclusions and recommendations.

> "Current information systems generated by research workers primarily for their own requirements are well established but most are quite inadequate for users in other disciplines and in technology, and are increasingly inadequate in their own disciplines.

"*Recommendation 5* - We recommend that governments give greater support to mechanisms for insuring effective interchange of information among scientists, giving explicit recognition to the key importance of informal systems, of which international personal contact and oral communication are an increasingly vital part. We further recommend that governments devote more effort to experiments in improving information transfer between scientists, particularly between scientists of different disciplines, and between scientists and non-specialists. Various kinds of information analysis, consolidation, evaluation, and repackaging can be envisaged here, and the different kinds of specialized information centers and information analysis centers have a vital role to play in improving the value - to science and to technology - of the national investments in R&D. These activities will improve both the quality and the usefulness of information in the hands of those who need it.

"We recommend that governments at the highest level accord priority attention not only to the development of policies for the generation of scientific and technical information, but also to the development of policy for the efficient and prudent use of such information in policy formulation, in the conduct of the affairs of government, and in R&D management.

"Proper handling of scientific and technical information must not be regarded as an administrative or mechanical matter, to be considered apart from (and often after) the design of R&D strategy. Systems for dealing with scientific and technical information have quite different requirements for the four spheres of activity in which information is used: for the conduct of science itself (for which most current systems are designed), for the effective generation of technology and its application in industry, for decision making and policy formulation, and for the enlightenment of the general public through education and public information.

"We recommend that policies and strategies for scientific and technical information should be developed as an integral part of the design of policy as a whole and R&D policy in particular, in such manner that in each of the above areas of public concern provision is made in advance for the scientific and technical information system requirements. Thus, the focus for policy concern in scientific and technical information should be closely associated with the focus of responsibility for R&D policy itself."

INFORMATION ANALYSIS CENTERS AND AUTOMATIC DATA PROCESSING

by

Ruth M. Davis, Ph.D.

When one is discussing computers and Information Analysis Centers at the same time, one feels obliged to start out with the phrase, "In the beginning there was Vannevar Bush; then came Alvin Weinberg's report." In that report, in 1963, some pertinent recommendations were made. Looking at them today would make everyone believers of the assertion that there is nothing new under the sun.

Dr. Branscomb said that one has the responsibility of deciding how to determine one's own progress. I think, in many instances, we are not sure, ourselves, when we have actually achieved a result, or anything we could call an innovation. We have not ascertained a means for making those kinds of assessments. For example, in 1963 the Weinberg Report said

> ". . . colleges must educate in the art of handling information more professionals who can lighten the burden of the technical man and can invent new techniques of information retrieval".

In 1971, we have a little over a hundred departments in universities which grant either a Bachelor's, Master's, or a Ph.D. degree in computer sciences; certainly computers have been recognized as a principal means for manipulating information. We have, however, only about 25 departments of information science in universities; a surprising number of these are actually "interdisciplinary" programs. A degree in an interdisciplinary program has, at the moment, rather uncertain market value after a few years. While "interdisciplinary" has the connotation of encompassing more than can any one given department in a university and, therefore, in some way not being able to be subsumed under any of them, it also has the connotation of not being adequately enough defined to become a department in itself, with a developed curriculum that will be able to meet stated objectives in education or in research.

This same report (Weinberg) in 1963 said

> ". . . The technical community must explore and exploit new switching methods. . . . The information transfer network is held together by an array of switching devices that connect the user with the information (as contrasted with the documents) he needs. As the amount of information grows, more ingenuity will be needed to find effective switching mechanisms. . . . The technical community must courageously explore new modes for information processing and retrieval."

Now when we say switching we normally think of that type which involves some sort of automation. In this area, when we try to determine progress, we reach this impasse of not knowing how to judge progress. Certainly, if we look at connecting the user with information, as opposed to connecting him with the documents that he wants, we are faced with the complex problem of electronically connecting users, geographically dispersed, with information at locations different from their own. The implication in the Weinberg report was that one neither moves the users to the information or moves all the information to each individual user. We then realize that we haven't even been able, as yet, to define the reality of this goal. The goal of delivering information, as opposed to documents, is essentially the goal for information networks. But the means of meshing network components, and making them work, will keep us all busy for some time.

Let me now describe a goal of remote browsing. This simply means that I can sit in my office and with a telefactoring device, meaning any kind of mechanism which extends my capabilities, I can connect myself to remotely located information media, be they magnetic tape, paper, films, photography or books. I can connect myself in such a way as to remotely browse through these holdings, make my own selection of information through query, pull books from stacks, look at a set of photographs or query a computer from a console, xerox or duplicate at my own convenience the information I need, stamp it and put it into a mail chute at that information center and mail it to myself. Now, surprisingly enough, this is a capability that has been costed out to a limited extent and for which most of the technology exists in scattered segments. It is a capability towards which the Weinberg report pointed some eight years ago.

Another recommendation in the Weinberg report stated

> "Among the schemes that ought to be exploited more fully are:
> a. Specialized Information Centers. . .
> b. Central Depositories. . .
> c. Mechanized Information Processing. . .
>
> Commercially available equipment is not the remedy in every case; . . . There is a need for equipment specifically designed to retrieve documents from very large collections. . .
>
> d. Development of software. . . Software, including methods of analyzing, indexing and programming, is at least as necessary as hardware for successful information retrieval. . . .
>
> ". . . Uniformity and compatibility are desirable. . . Switching will be fully effective only if the different subsystems adopt uniform practices towards abstracting and indexing".

There were a number of people, both in and out of the Government, who took these recommendations seriously, both then and now. In the early 1960's, there were a tremendous number of projects aimed at this particular set of objectives, all of them oriented around the use of automatic data processing equipment or complementary type equipment. These projects included the development of associative memories which one could rapidly scan, in a non-structured way, large stores of textual information, and retrieve with one search all of the information that met certain criteria. They included the first steps toward automatic on-line indexing with consoles, where one could have displayed on a cathode ray tube the text that one was editing and have, perhaps on a second cathode ray tube, the rules for indexing and/or abstracting. Using a light pen, as an extension of one's pencil, one could index and/or abstract these articles on-line with the computer, have the information automatically inserted into the computer, and thus be available for retrieval. They also included the development of techniques for coupling document storage media with the indices or the means of retrieving documents.

We have since seen the design and development of several models of trillion-bit memories and of film storage devices for handling microtext. None of these have been operating long enough to permit cost-effectiveness or comparative analyses. We are still in what I call the stage of "Tweezers Technology" with respect to retrieving documents in reduced image forms; this means that the best techniques still involve the use of tweezers to pull a microfiche out from its storage place. We still worry about wear and tear on microfilm documents in their retrieval process.

We still worry about automatic indexing because we haven't solved the intellectual problems of indexing. The time-dependency of indexing is one of our most crucial problems. We don't generally have associated with index terms that very essential qualifier that lets us know when that index term was useful. We don't have a way, therefore, in most automated systems of being able to cross-reference to permit automated updating of indexes.

At the time Dr. Weinberg wrote his report, in 1963, there were 400 Information Analysis Centers that he identified. COSATI Panel 6 in the first edition of its Directory in 1968, identified 113 Federally-sponsored Centers. In that list, 24 of these 113 indicated that they were using computers; that's about 21%. In 1970, the updated version of the Directory of Federally-sponsored Centers shows 110 Centers of which 34 indicated that they were using computers; that is about 28%. So, there was about a 5% increase in the number of Centers in two years and about a 7% increase in those that were using computers.

The manner of using the computer by these Centers was fairly uniform; namely, for the compilation, manipulation, and retrieval of data. In some instances, the Centers offered copies of data compilations on tape; in other cases, the Centers offered computer programs or use of computer facilities for outsiders. In the majority of cases, the computers were the tools by which Centers stored and retrieved data in order to answer specific inquiries or compile particular lists of data. In a very few cases, the computers were being used as repositories for the data. But in no cases were the entire text, or the entire information bank that the Center worked with, stored in computer form.

In discussing the application of computer technology as related to Information Analysis Centers, one generally distinguishes between information handling and document handling systems. The products inputted, processed and retrieved in document handling systems are documents, document surrogates and/or document reference. The equivalent products of information handling systems are information and/or data elements. In designing composite systems or service networks the general guidance given is to separate the two types of systems. The reasons are that the supporting technologies, equipment and manpower resources are significantly different.

In a way, the Information Analysis Center as it has evolved represents the worst of both worlds. It inputs and processes both documents and information. It is generally supposed to output only information. The technologies of both document and information handling must be brought to bear in building an Information Analysis Center. That particular merge of technologies is one that remains a challenge, as opposed to an accomplishment, at this date. The motivation behind merging these two technologies has been stated very well by Dr. Branscomb: it is the urgency of information, and the urgency in which users place their requirements for information, that has somewhat strained not only the technologies of information analysis, but also the intellectual prowess necessary for this information analysis.

One of the reasons for this strain, and perhaps unendurable stress, is that the increased amount of information that has been available for the last 60 or 70 years has had some remarkable effects. First of all, the elapsed time between the initial discovery of an innovation and its recognition as a commercial product has decreased, from about 30 years in the early 1900's to about 16 years following World War II. That means a number of things: indexing schemes have to be changed twice as often; thesauri, and the number of terms used, are increasing twice as often. The time to translate a technical discovery into a technical product is now down to five years, from about 12 years at the end of World War II. The half-life of information, before it becomes not only out of date but—in the case of fields such as health and drugs—dangerous to use, is decreased from an average in scientific fields of about seven years to less than five years, and in some cases, in the computer field, for example, to about three and one-half years.

All of this means that it's difficult not only to read the information, but to use it: a state-of-the-art review which now takes two years may well be considered an historical survey. Unless we find ways of aiding the intellectual process of making state-of-the-art reviews, and unless we find ways of assimilating information faster than we can do it manually, we simply are not going to keep up with the rate of introduction of technology. Technology transfer then cannot be dependent on the essential hard, complete, and accurate kind of analysis that it should have in order to achieve its greatest utility.

Now there are some interesting counter-balancing effects introduced by technology. It's a sort of "check-and-balance" effect. More information is being required for decision-making before the decisions are allowed to become final; therefore, there is a slow-down in the decision-making processes when introducing new products and determining the uses for new products. As an example, the drug industry, one of the fastest growing in the country, is regulated by the Food and Drug Administration. Certain requirements have to be met before a new drug can be introduced to the commercial market. The amount of information needed by the government—information largely based on the results of experimentation—is such that the length of time before introduction of drugs is increasing. As a result, the time which elapses until the drug is available to you, or to your physician, is also increasing. The demands for additional information on an ever-increasing number of new drugs act as a drag effect on the introduction of the technology.

One of the major problems faced by any Information Analysis Center in attempting to keep up with technology and, at the same time, to make sure that its analyses are just as correct and comprehensive as before, is the storage requirement for the documents, and the information itself. It appears to computer technologists that the storage requirements are of several types: for storing documents themselves, for storing document surrogates, for storing document references, for storing the information and data that are generated, for storing the information and data elements that enter the system, for storing the management information necessary for Center operation. All of these are different. Attempts to use the same kind of storage mechanism for all of these requirements introduces a situation that makes effective storage difficult. As was stated earlier, in no presently operational system is the full text of all documents stored on computer media for search or retrieval. The normal reasons for this are, very obviously, the expense of converting from text to digital form, the difficulties of digitizing graphs and photographs, the cost of present computer storage and the unproven value of having full texts in digitized form. Unless techniques are developed for retrieving or manipulating texts based on content, there is not much reason for going to the expense of putting text in digital form.

The role of Information Analysis Centers now is such that there can indeed be an impact on their operation through computer use, even though only 21 to 28 percent of the Information Centers use computers, and these are generally simple uses. The situation is such that a little guidance from COSATI Panel 6 and a little guidance from computer scientists and information scientists can indeed affect these Centers critically in the costs of their operation and in their plans for the future. I think that one of the objectives of these Centers, as implicitly defined by the agenda for this meeting, is to determine how to keep up with the processing and analysis of information through the most effective use of computer technology. I heartily ascribe to that objective and I will be looking forward to future progress towards that objective.

RELATIONSHIP OF INFORMATION ANALYSIS CENTERS AND ABSTRACTING AND INDEXING SERVICES

Byron Riegel
President, ICSU Abstracting Board

I wish to express my appreciation to Edward L. Brady and Harvey Marron and their staffs for their hospitality and arranging for my participation in this forum on the management of Information Analysis Centers. The COSATI Panel No. 6 on Information Analysis and Data Centers has supplied a definition of an Information Analysis Center (IAC) that is out of this world. I had not read the definition until I started preparing for this talk. This definition is so broad and big, I believe I can say almost anything and it will be appropriate. The purpose of the IAC's greatly overlaps the responsibilities and purposes of most of the Abstracting and Indexing (A&I) Services, particularly in the acquiring, selecting, storing, and retrieving of information and also compiling, digesting, and repackaging information. This is not all bad. Practically all scientific and technical information services must overlap with each other in order to establish a recognized continuum. It seems to me that my problem is to outline the policies for management of A&I Services and IAC's so that there is the minimum of duplication of effort, chiefly the intellectual input. Furthermore, we should establish a great dependence upon each other so that we have an excellent feedback to help guide the development of each of these types of services.

Whenever I speak before a distinguished group of this caliber, I am afraid that I must be masquerading around under false pretenses. I am not an expert in information handling, either in relationship to A&I Services or to IAC's. However, I have been closely associated with the development of the Chemical Abstracts Services since 1959 and have served as president of the ICSU Abstracting Board for the last two and one-half years. Most of my time has been spent on raising money and working on methods that would save and conserve the money. These then are my credentials as an authority on scientific and technical communications.

Many times, we who are deeply involved in the information transfer process lose sight of our purpose. Our purpose is really simple. We are trying to serve the individuals who have need for the information. The attempts to regulate our operational procedures so as to serve the user have not been entirely unsuccessful. The user has to be educated to the new methods and doesn't appreciate all of the things that are available. Furthermore, we scientists and engineers who generate the information are the principal users. Everything that I want to say has been said, or written, and all that I can do in this talk is emphasize some of the high points of the past. It is almost impossible at the present time to make an original contribution to this complex problem.

There is a genuine desire in the United States, and worldwide, to standardize the methods for the transfer and dissemination of scientific and technical information. We have made good progress in this area of standardization during the last three years. May I ask you what do you think was the driving force to promote this international cooperation? If you haven't guessed, it was the lack of funds. We have been forced to cooperate in order to survive and carry out our missions.

Who provides the money for all of this work? Again, the answer is simple. It is the public. It makes no difference whether it is a government-supported operation, a scientific or technical society, industry, or so called not-for-profit foundations. The funds in every case come from the public. This is a responsibility to the public and is not solely a scientific and technical society, government, or industrial responsibility.

The greatest problem to me at the present time is how to price the services in a fair and equitable manner. We have all heard the problems of the primary journals. COSATI was very instrumental in arranging for page charges to help keep our primary literature viable. Most of the A&I Services of the world are in trouble because of the immense cost of mechanization. In practically all cases, the A&I Services are running parallel services using the old classical method of hard copy while they are trying to develop more sophisticated computer manipulation of information. It is not cheap to develop a computer data bank of information from which there can be instant recall or access. The question is how should we distribute the costs? Who should pay for what and how much? Also, very few people have tried to study in depth how to market scientific and technical information. I have a very wholesome respect for any commercial group that can survive in this field.

Combined with the cost of the storage and retrieval of scientific and technical information is the whole problem of copyright. Sometimes I envy the Russians and their VINITI operation. We have copyright of the primary publications with rights to the authors and editors. We have copyright for the A&I Services including their magnetic tapes, microfiche, software for the computers, and finally, the compiling, digesting, and repackaging which may all involve innovative contributions. Sometimes I think it is a shame we just cannot ignore the whole copyright business. If I were the author of a few successful textbooks, I would have an entirely different viewpoint. Here again, this was well discussed under the "Freedom of Information" act by COSATI Panel No. 6.

There are so many important areas in the handling of scientific and technical information that cannot be discussed in this paper, one of which is our research libraries. There is no doubt in my mind that, on a worldwide basis, they will be the real information dissemination centers. In fact, I am not too sure that some of these centers you have called analysis centers should really be called distribution centers. But you admit this.

At this point, I would like to say a few words about my own personal philosophy on the policy for handling scientific and technical information. This is highly colored by my own background which carries a chemical bias. The most effective way to handle information is to subdivide it into very small groups. Let each group handle their own information in the way that helps them the most. IAC's are a fine example. This may sound like heresy for me to talk this way when I have been so closely associated with one of the world's largest A&I Services. I was very much impressed by the IEG's (Information Exchange Groups), the so-called "invisible colleges" that were established by the NIH and then had to be discontinued because of their conflict with other established methods. They were so successful, I would like to see some modified form of them started again. The second thing which I believe is that well-operated A&I Services can, and should, supply the necessary bibliographic material for practically all mission-operated information distribution centers. This also includes data. Third, the A&I Services will be forced to develop methods of classification where the index terms for any one service will be defined and understood by the other services. Fourth, I am not too impressed by the classification systems that have been designed to date and thoroughly believe that UDC will not be acceptable on an international basis. Fifth, I am not too impressed by efforts to develop multilingual thesauri. Probably my lack of knowledge influences my strong convictions. Finally, I am overwhelmed at the progress that is being made in transliteration of languages that do not use the Roman alphabet and the universal agreement which is being made at a most rapid pace through UNISIST, ICSU AB, ISO, IFLA, FID, and a few dozen other organizations. This talk was designed to stimulate discussion, so now it is your turn.

MARKETING THE PRODUCTS AND SERVICES OF INFORMATION ANALYSIS CENTERS*

H. William Koch, American Institute of Physics
and
Walter Grattidge, General Electric Company

Abstract

Information analysis centers must perform the function of evaluation as well as compilation, in order to generate products and services with increased utility and user acceptance. Also, these centers must perform a dual role as wholesaler and retailer. These roles, as well as problems and examples of production and marketing experiences, are examined so as to elucidate the present and future potential of information analysis centers in improving communication among scientists. The full potential will be shown to require marketing by information analysis centers operated by a complete spectrum of institutions, including governmental agencies, scientific and technical societies, not-for-profit groups, and commercial firms.

Introduction

In 1963, the Weinberg Panel on Science Information [1], with great foresight, envisaged information centers that were technical institutes rather than technical libraries. Such centers would, with the aid of dedicated and knowledgeable interpreters, "collect relevant data, review a field, and distill information in a manner that goes to the heart of a technical situation," and thereby would be "more helpful to the overburdened specialist than is a mere pile of relevant documents." The panel projected that such information analysis centers would eventually become "the prime retailers of information to scientists." [2].

This ultimate potential is apparent from the present developments of several different types of information analysis centers based both on the nature of the particular information base being covered and on the requirements of the user group to which the output of a center is primarily directed. Garvin [3] has summarized the scopes of such centers. As he indicates, the important factor in the Information Analysis Center Concept is evaluation and those products that result from it.

Many centers, whose function is to process information already in the public domain, are now well established. Within the National Standard Reference Data System there are some 26 information analysis centers concerned with the review and evaluation of data in the physical sciences. In addition, there are almost a hundred other federally-supported analysis centers.

Simultaneously, there are developing within industry comparable operations devoted specifically to internal company users and with coverage of both public and proprietary data. In addition, there are commercial services, both traditional and new, available at both a "wholesaling" level and also at a direct user "retailing" level.

This report on the marketing of their products and services assumes as one of its basic premises the evaluation model of an information analysis center. Further, while the primary focus is on evaluated numerical data produced by the centers as distinct from documents, the latter type of product is important and will be referred to. In addition, since the user is the dominant factor, the production and marketing functions must be closely intertwined and directed to the ultimate user. Therefore, both production and marketing are considered in this report.

Starting from these premises, what are the production and marketing limitations and opportunities? How can we successfully market products and services that have predominant characteristics determined in the production phases of those products and services? How do we grapple with the vast producer-oriented stores of data being generated by scientists and technologists? How can we best user-orient the data at information analysis centers? For that, in effect, is the next important phase between production and marketing that must be accomplished if we are to market the data. How do we work towards information analysis centers of the future as "prime retailers" to scientists?

Problems of Production and Marketing

a. Problems of Production

The production problems of data compilations and evaluation (see Table I) are understood and have, unfortunately in some instances, become an accepted unsolved tradition among research workers. These problems must now be tackled and research workers must be involved in their solution if research and development work is to be kept efficient and effective. Today's new solutions involve non-traditional along with traditional methods.

It has long been recognized that it is much easier to do a piece of research and report on it than it is to review the literature and data in a critical manner and produce an authoritative review or data compilation. Many research professors have graduate students who are given individual research assignments and from whom research results can be monitored and evaluated. In the case of an authoritative review or data compilation, it is usually necessary for the professor or senior researcher to remove himself from the research environment, with little or no support from assistants, and to examine the information in a scholarly manner. The professional rewards have traditionally been larger for the research professor discovering new concepts than they have been for the same individual reviewing, evaluating, and compiling the data of others.

Because reviews and compilations require special encouragement and support, the National Bureau of Standards established in 1963 a National Standard Reference Data Program. By means of this program, it is possible to have manpower for data compilations fully supported with Federal funds in a manner that has become traditional in the support of original research work. To date, the program is still in its infancy. While the Federal Government is supplying funds of the order of $300 million for research in physics, the Standard Reference Data System (NSRDS) is only funding physics data compilations at an annual rate of less than $2.0 millions. It would appear that more support in the latter area would yield high leverage in increasing the productivity of the former investment.

TABLE I

Problems in Production and Marketing

Problem Area	Problem	Solution
Production	Inadequate Professional Encouragement and Reward	1. NSRDS 2. Review and Compilation Fellowships
Publication	1. High Cost 2. No R & D Funds 3. No Involvement of Science Community	1. NSRDS 2. Proposed Journal of Chem. & Phys. Ref. Data
Technology	Uneconomic Computer Storage and Accessibility	Combination of Computer Tape and Microfilm Service
Marketing	Diversity of Products, Applications, Users	Successful Response to Marketing Challenge by Government, Society, Not-for-profit and Commercial Sectors

While the Federal Government has recognized its responsibility to encourage reviews and compilations, various scientific societies are also recognizing their responsibilities in this direction. For this reason, the American Institute of Physics, in representing its seven member societies, plans to make a major "review and compilations" proposal to the National Science Foundation to obtain financial support. If this support is forthcoming, it is proposed that the U.S. physics community will be directly involved in providing fellowship funding to outstanding specialists so that these specialists can spend a sabbatical year at centers of their choosing to undertake specific review projects. Such centers will undoubtedly include many of the information analysis centers that are represented at this conference.

b. Problems of Publication

The second problem in numerical data compilations is their publication. By their very nature, reviews and compilations frequently result in articles that are longer in length and frequently more complicated and detailed than are the primary research articles. Extensive tables of data are frequently very difficult to have published because of the attitudes of publishers and because of the lack of funding of authors. The authors and their institutions frequently have funds to publish the results of their research work, but do not have funds to publish the results of reviewing and compiling the data of others.

A further aspect of the problem of publishing data compilations has to do with the lack of involvement of the scientific community in the publication process. In the case of primary research articles, the scientific community has established a system of referees who review, for acceptance, articles submitted for journal publication. In the case of data reviews, there is not as yet an accepted system for refereeing compilations and data values. This is because the compilations have not, in general, been published in the scientific literature operated by scientific societies. The result has been that the scientific community has not been involved in a formal way. Somehow peer acceptance and prestige needs to be developed by the scientific community to those who will analyze, review, and compile, and also those who referee the results prior to publication.

Most of the results of the NSRDS program to date have been published by the Government Printing Office, and their availability has been announced in media not generally available to members of the scientific community on an individual basis, as are the primary research journals. An attempt to develop a solution for these publishing problems has been the recent proposal for joint sponsorship by the American Chemical Society, the American Institute of Physics, and the National Bureau of Standards, of a new journal, **The Journal of Chemical and Physical Reference Data.** This journal will be able to publish the reviews and compilations originating in the centers supported by NSRDS as readily as any research article. Under the proposal, the principal elements of the scientific community involved in the work of NSRDS will be intimately involved in reviewing, refereeing, and preparing data compilations as they are in the same functions for primary research articles. The societies also would take care of publishing and marketing.

c. Problems of Technology

Another problem is the technological one of how to disseminate numerical data. The same determining factors involved in document handling and dissemination are involved in data handling and dissemination. In the case of documents, the full text of documents are not going to be disseminated in the form of a computer tape for a good many years to come. The case is similar for data, although there are now some examples of data being available in tape form for analysis and evaluation by the users. An example of such data are the neutron data tapes being produced by the Brookhaven National Laboratory for use by reactor design groups. An interim compromise to disseminating the full text of documents on computer tape is the announced plan of the American Institute of Physics to produce a combination package of techniques. One part of the package will be a computer-searchable magnetic tape describing the complete bibliographical information about all the articles contained in full text on the second part—a microfilm tape issued every two weeks or every month, simultaneously with its computer tape counterpart index. As soon as **The Journal of Chemical and Physical Reference Data** has been placed in production, it will be available in this dual format.

d. Problems of Marketing

Government agencies, scientific and technical societies and not-for-profit groups, and firms in the commercial sector are all becoming involved in marketing information services and, especially, in marketing the specialized products of information centers. An understanding of the relationships of these different organizations and their respective types of marketing requires an awareness of what is encompassed in the new terms of wholesaling and retailing now beginning to be applied to information services.

Wholesaling includes the production, evaluation, and marketing by, and for, the producing scientists as well as the serving of customers who in turn repackage or produce specialized information products for retailing to the ultimate user. Thus, if one is to have an economically stable industry, the criteria for determining economy, timeliness and quality at the retail level must reflect the equivalent criteria incurred in packaging and dissemination at the wholesale level. The extent to which wholesalers see their principal activity as information generation (through critical reviews and data evaluation) may determine the extent that they consider themselves participating in the primary research and development function. In the words of the Weinberg panel report, "Transfer of information is an inseparable part of research and development." However, transfer and dissemination without a contribution of evaluation does not appear to command a large value-added factor in the market place.

Marketing has to be done not only by wholesaling to the specialist groups who require specialized services by agencies and societies, but by retailing to non-specialist public audiences and to specialist audiences of other specialties. To date, customized public retailing has been done primarily by the commercial sector. This sector deserves to be encouraged and stimulated in continuing in these areas for which it has particular capabilities and expertise.

The major problem in marketing at both the wholesale and retail level results from the requirement to disseminate or deal with a wide diversity of data products, of access and application, and of secondary information generators and ultimate users. The dissemination, in turn, has to be done under conditions of economy, timeliness, and quality that are acceptable to the user.

The marketing challenge is therefore to identify and reach the group of potential users even when this group is of a narrow scientific or technical discipline. Specialized libraries and information centers are possible marketing contact points. However, as previously indicated, not all potential users are linked to identifiable specialized libraries. Particularly in relating to academic users, it may be necessary to use broad marketing channels, such as professional journal advertising and broad library mailings [4] to ensure coverage of that user segment. The efficiency of this notification process becomes a significant factor in the cost of marketing and must be seriously considered in establishing a pricing policy which reflects full cost recovery, at least of the secondary dissemination costs.

With this explanation of marketing and its major problem, let us consider other problems created by the acceptability criteria of economy, timeliness, and quality.

1. *Criteria of Economy (or Costs and Prices)*

There appears to be experience accumulating regarding what customers are willing, or perhaps better, are now conditioned to pay for information products and services.

Table II lists types of information services involved in the handling of data compilations by the commercial sector. There are different types of compilations available with varying degrees of evaluation involved in the production process. As shown in the final column, the attitude of the market at the present time is that the charges for providing such data services cannot be much above the distribution cost level. The experiences undoubtedly reflect the user's evaluation of the ease, ability and costs of reproducing the data himself compared to having the data supplied. If problems of acquiring the data by purchase are of a comparable order to reproducing the data, then the data will not be bought. If the research is Federally sponsored, then the threshold for the buy decision may be still lower. One may surmise that as the availability of sponsorship for research and development tightens, program managers will increasingly evaluate the full costs of data duplication and be prepared to buy when the data is available.

TABLE II

Data Compilation Products and Markets

Product Type	Principal Identifiable Market Segments	Contact Channels	Competitive Products	Market Attitude on Value
1. Unevaluated Data Compilations.	• Research Peer Group	• Profess. Soc. Memb. • Univ. Dept.	• Journals (Special Issues)	Users value the information only at the distribution cost level.
2. Evaluated Data Compilations.	• Research Peer Group • Industry Design Eng. • Education	• Profess. Soc. Memb. • Business (SIC Groups) • Univ. Depts. Libraries • Spec. Libraries	• Publishers Monographs • Material Suppliers Catalogues • Handbooks	Users appear to value the information at the distribution cost level plus a *small* return to the expert to partially cover his cost of commentary.
3. Combination Data with Expert Forecast (principally economic).	• Industry Business Planning Function	• Business (SIC Groups)	• Specialized Newsletters	As 2 above.
4. Engineering Design Data (usually proprietary).	• Industry Engineering Function	• In-House Distribution	• Usually Non-Marketable Due to Proprietary Content	Most buyers are suspicious of anyone offering this type of product.

Since run-off and distribution costs appear to fix the level at which users consider it economic to buy compilations, how can pre-run costs be met? The whole question of cost recovery in the dissemination of scientific data has been the subject of a study of an ad hoc Panel on Marketing of the Numerical Data Advisory Board of the National Research Council—a study related specifically to the products of the National Standard Reference Data System. In a memorandum by the Panel transmitted to the Director of the National Bureau of Standards, the following two recommendations were made:

"1. The Panel recommends that the scientists engaged in the important and necessary work of data evaluation should be supported by the Government, in similar manner to the Government's position in funding primary R & D scientists; this work should not be a target for cost recovery.

"2. The Panel recommends that the accepted page charge concept for R & D results be applied to the publication of NSRDS products as well. In practice, some (if not all) of the pre-run costs of publication of data compilations should be considered for Federal support."

These recommendations are consistent with the concept that publication is a necessary part of research as well as its compilation and evaluation; and that users can be expected to pay for data compilations at a rate that covers run-off, distribution, and very little, if any, of the pre-run cost for evaluation.

2. *Criteria of Timeliness and Quality*

Just as with costs, the criteria on what the specialist, as well as the non-specialist, will be willing to afford in time delays and in quality are determined, primarily, in subtle ways at the input by production standards and criteria. A prize example in the present context is the extensive growth during the 1960's in the use of preprints and governmental reports that competed with the more conventional information transfer mechanisms available in the primary research journals. These journals have been, and will continue to be, produced by both society and commercial publishers. As the journals have become bigger, more costly, and more delayed, other communication mechanisms, such as the preprints and reports, were invented to bypass the problems of the journals.

It has become recognized that preprints and reports are very effective ways of communicating quickly to a specialized audience. However, these mechanisms are extremely costly and are, in general, non-public communication mechanisms. There have been concerns expressed that these mechanisms have been getting out of control, and will result in the disappearance of journals in their present form. However, the multiple advantages of journals and the increased attention to costs and timeliness are resulting in renewed recognition that the conventional, proven techniques in the form of journals must be strengthened in order to accomplish wide, public, and economic dissemination of scientific information.

Examples of User-Oriented Data Products and Services and Their Marketing

a. Handbooks

The conventional method of bringing comprehensive data compilations to the market place has been by the publication and marketing of handbooks. Such handbooks have traditionally been published by commercial publishing houses or by specialized subsidiaries. (See Table III.)

In such publications, masses of data covering a broad scientific or technical discipline are compiled and arranged in an accessible form for the user. The compilation is then published in book form [5]. By including within the one publication many sets of data which cover a broad spectrum of users, the publication has broad market appeal.

The data in many cases represent standard values having a useful life-time (to the user) of several years. Thus a specific edition is not immediately outdated on publication, and by bringing out new editions every two or three years, the publisher sustains a continuous impact on the market.

TABLE III

Examples of User-Oriented Data Products and Services

Item	Characteristic	Example	Traditional Publisher
1. Handbooks	Compilation of Data in Broad Scientific Discipline Published in Book Form	Handbook of Chemistry & Physics	Commercial
2. Data Subscription Services	Initial Set of Data Followed by Updates	1. F & S Index of Corp. & Ind. Monthly 2. GE Data Books on Heat Transfer and Fluid Flow	Commercial
3. Individual Compilations	Determined by Data, Author, Institution, Publisher	NBS Report of Superconductive Materials	Government, Society, Commercial
4. Specialized Compilations	Proprietary or Otherwise Restricted	GE Eng. Mat. & Process Info. Service	Commercial
5. Data Bases of Literature	Secondary Services on Computer Tape	SPIN, CAS, and Commercial Services	Society and Commercial

b. Data Subscription Services

Over the last decade, specialized data services have been developed and marketed on a direct subscription basis. Included in this category are services for which the user receives an initial set of data followed by updating revisions or extensions on a pre-arranged periodic basis. From time to time a new up-to-date comprehensive data base is issued which supersedes all earlier editions. Such a service is attractive to the user whenever the data values change with time or in time and where the market places a premium on up-to-date validated data values.

One major segment of data services of this type cover economic or technical-economic fields where new data values become available at fixed calendar dates. For such services, quarterly, semiannual, or annual data values are important to users. Reference 6 is a typical example.

In most technical fields, data values do not become outdated or superseded quite so fast so that periodic updates, where they occur, are much less frequent. One example is the recently introduced series of data books on heat transfer and fluid flow each of which is marketed on a subscription basis [7]. With each service there is an annual up-dating of the data included in the subscription price. This annual up-dating includes the addition of new sections as well as the revision of existing sections.

A more recent development of such data services has been the provision of the data to the user in a computer accessible form. This may be either by the provision of data on a computer magnetic tape or by a computer accessible data service. For the magnetic tapes, the subscription covers periodic updates or supplements, and with many services of this type there are specific computer programs available or provisions for user education and training. In the case of the computer-accessible form of service, the cost may be made up of a fixed subscription plus a variable amount based on monthly access usage of the data base. An example of this type of data base is one on organic chemical compounds [8]. Data is supplied to the user either on magnetic tape for in-house manipulation or the opportunity is available to use the data base via a remote access computer terminal.

The principal problems involved in marketing data subscription services involve the identification both of potential users and also of the most effective channels to make the availability of the service known. In addition to the ultimate users of these data (scientists and engineers), there are related services (libraries, information centers, and computer centers) whose personnel also have interests as intermediate handlers of the information. Marketing techniques, therefore, involve brochures, mailings, and advertisements, and in the case of computerized services, may also include demonstrations and exhibits at scientific and professional meetings, one and two-day invitational demonstration and training institutes, and on-side demonstrations and trials.

c. Individual Compilations

In many instances, a specific compilation of data is published by itself. The form of product depends on the scope of the data, its author or editor, the sponsoring institution, and the characteristics of the user group for which it is intended.

Examples can be cited where the finished product is of size and of such broad market impact that a recognized publishing house will publish the compilation in book form [9]. In other instances, the compilation is more appropriately published through the sponsoring agency as a monograph [10]. Recently, specialized compilations of data have become available on magnetic tape [11].

In this type of individual compilation, the compiler of the data is very often aware, professionally, of the principal generators of the data and in many instances they, in turn, are aware of the compiler's assignment and responsibility [12]. In fact, the compiler or editor has a professional responsibility to evaluate and select the data prior to incorporating it in a publishable data base. The result is that the editor or compiler performs a gate-keeping or quality control function on data values which, to a large extent, become accepted in the profession [13].

The difficulties associated with the marketing of such data bases arise from delineating all potential users other than those who are data generators themselves. As we have indicated, the latter group are known professionally to the compiler, and communications arising during the compilation and interpretation process often occur directly. Identification of other potential users is less straightforward. While one can list general disciplines or sub-disciplines that should be concerned with the data, the specific identification of individuals in colleges, industrial laboratories, or government agencies, who would or should have a direct interest, is very difficult. Thus a major marketing effort is required to attract the attention of these potential users to the availability of the compilation.

Many times in the past, when the sponsor for the compilation of the data has been a Federal department or agency, then the publication and marketing activities have occurred through the U.S. Government Printing Office (GPO) and the Office of the Superintendent of Documents. It is now clear that potential users have not always been aware of the availability of such publications, since they personally may not be exposed to the GPO document listing, and they may not always have local librarians or other information center personnel aware of their specific data interests.

To overcome such gaps in coverage requires such things as advertising in professional journals, direct mailing to university departments or to companies in specific industry classifications and, whenever possible, secondary advertising through newsletter, etc. All these methods will be recognized as inherently inefficient since they employ broadcast techniques to communicate with a narrow interest group.

An alternative marketing approach is to seek to develop on an individual basis a list of names of the potential users for each data base. Hopefully, this list grows as the data base itself becomes more complete and comprehensive. Direct advertising to these users then becomes a more efficient marketing technique, though it may miss many potential users of the data.

A recent challenge, particularly for Federal agency sponsorship of such compilation and evaluation activities, is for the sponsor to demonstrate the broad social value of such data compilations by market place criteria. In particular, if the data compilation and evaluation functions are recognized as research and development activities to be Federally sponsored, as such, then the utility of their output should be evaluated by the extent to which they satisfy a significant segment of the recognized potential market at a price level which covers at least the marketing and distribution costs.

This latter is, of course, most easily recognized when a commercial publisher is willing to serve as the publishing channel. This has recently happened with the multi-volume compilation, "Thermo-physical properties of matter - The TPRC Data Series," in the process of being published [14].

d. Other Specialized Compilations

There are certain data compilations in existence for which the distinguishing characteristic is that they are considered highly proprietary, or otherwise restricted by security, to a particular company or organization. Almost by definition, such compilations are not for sale or release to the public or to others on an individual basis except by specific authority. Marketing problems are at a minimum. However, consideration is occasionally given to making such compilations accessible to a wider public. Dominant factors in the consideration are the identification of the market for the data and recognition of the marketing channels through which to contact such groups.

For example, compilations of preferred design data on materials are created in many engineering design organizations in industry. Once created, the question is occasionally asked as to whether such an information base would not be saleable especially to manufacturers in related industries. Usually the answer is that such information is too sensitive for proprietary reasons to release. Occasionally the decision is made to offer such a system for sale. In that case, the marketing challenge becomes one of identifying corresponding industrial users and establishing contact channels.

One such example of this type of data base is the Engineering Materials and Processes Information Service (EMPIS) [15]. This is an extensive information bank covering descriptive data and specifications for manufacturing materials. The service was test marketed for three years, but is not presently offered outside the company producing it, though it continues to be an internal system within that company for material specifications. One of the peculiar marketing problems encountered in the test marketing of EMPIS concerned the inability of the potential users of such information (design engineers) to convince appropriate top management that the subscription cost of the service was a necessary expense.

e. Data Bases Covering Scientific and Technical Literature

A recent report [16] has presented the results of a survey of commercially available computer magnetic tape services which can provide libraries and information analysis centers with data bases of scientific and technical literature. This directory lists the general characteristics of each data base, the most frequently used access points, the frequency of the tape issues, and the number of items reported on an average tape issue.

This particular report is the result of cooperation by a special interest group of a scientific society—the American Society for Information Science—and the American Institute of Physics. It is to be anticipated that similar compilations of available data bases in other areas will become available through journal articles and other media.

User Access to Data Compilations: The Test of Successful User-Orientation

In the traditional printed form, a data compilation is immediately accessible to the user once he has located the volume either on his own bookshelf or in the library. The existence of xerographic copying has further reduced any tedium that there may have been in transferring specific data values to his personal information files. As the volume of primary research information has grown, most scientists have been forced into a mode of selectivity of exposure to the literature resulting in a decrease in awareness of pertinent information. This, the Weinberg Panel foresaw and postulated the development of intermediate information centers for subgroups of users.

In the long-range plan [17] for an information system for physicists, this type of center was envisaged as an integrated information control center. Its major function would be to monitor the interests of user groups in subdisciplines and interdisciplinary combinations relating to physics and astronomy, and to devise and operate procedures for manipulating its files to provide references and back-up documents for dissemination to users. When one adds the function of information analysis, the generation and publication of topical status reports and annotated bibliographies to supplement conference proceedings, the center expands beyond the concept of a conventional library or information store to a technical information institute which would attract consultant scientists and visiting scholars to engage in the preparation of reviews and compilations.

If one can forecast the effects of further significant decreases in the costs of information transfer through present day land-line or microwave communication channels, augmented by communication satellites and cable T.V., one can speculate that there will develop a close, direct relationship between the user and his particular information analysis center, regardless of geographic distance.

It is the direct user interface which is most crucial to the effective working of information transfer systems and it is one where our efforts to date have made little headway. In most instances, the user now, in answer to this query for some factual data, is invariably given a series of detailed sign-post instructions to original papers. Copies of the papers are not attached and his library is invariably some distance away. Consequently he loses enthusiasm for schemes which tell him how well the primary generators are doing, while he must still hope that his problems will not be forgotten.

Information in the public domain will need to be made more accessible to user enquiry. There are many ways of key-word indexing, subject identifiers, machine methods of self-indexing all directed toward more rapid query access. This the user is coming to expect, though he may require considerable education regarding the price level at which such service can be offered.

Another area of direct concern to the user is the collection, organization, and dissemination of data within his own research environment, whether that be a research institute, commercial company, or government agency. Convenient methods of standardized data collection are required with corresponding convenient access methods for co-workers with related interests. Many fields of research now appear to have reached the point where organized data stores would enable researchers to expand the scope of their own research studies with little increase in cost, and, thereby, increase their research productivity.

These are problems that information analysis centers and others who seek to participate in this new industry must address themselves to if they are to retain the interest and support of the user. We are convinced that these centers will be able to solve these problems and to fulfill the need for evaluated data and knowledge compilations. The Weinberg Panel should be credited with being a major force in encouraging the appropriate development of centers. It pointed the way toward avoiding, in the future, the stifling effects of the avalanche of information on individual research workers.

References:

[1] *Science, Government, and Information,* A Report of The President's Science Advisory Committee, Panel on Science Information, A. Weinberg, Chairman, January 10, 1963.

[2] C. Herring, "Distill or Drown: The Need for Reviews," *Physics Today,* Vol. 21, No. 9, p. 27 (September, 1968).

[3] D. Garvin, "The Information Analysis Center and The Library," *Special Libraries,* Vol. 62, No. 1, pp. 17-23 (1971).

[4] Jerry A. Minnich, *Approaching The Librarian,* Scholarly Publishing, January 1971.

[5] *Handbook of Chemistry and Physics,* Chemical Rubber Publishing Co., and *Standard Handbook for Electrical Engineers,* McGraw-Hill.

[6] *F&S Index of Corporations and Industries Monthly,* Predicasts, Inc., Cleveland, Ohio 44106.

[7] *Data Books on Heat Transfer and Fluid Flow,* General Electric Company, Technology Marketing Operation, Corporate Research and Development, Schenectady, New York.

[8] *SDB,* Science Data Bank, Inc. 18901 Cranwood Parkway, Cleveland, Ohio 44128.

[9] S. C. Brown, *Basic Data of Plasma Physics,* M.I.T. Press, 2nd Edit., 1967.

[10] B. W. Roberts, *Superconductive Materials and Some of Their Properties,* NBS, Technical Note 482, issued May 1969, Supt. of Documents, U. S. Government Printing Office.

[11] *Evaluated Nuclear Data File - ENDF* (Nuclear Properties of the Elements), Brookhaven National Laboratory.

[12] For the basis for establishment of the Superconductive Materials Data Center, see: *Physics and Chemistry of Solids* **23**, 181 (1962).

[13] Amitai Etzioni, *The Need of Quality Filters in Information Systems,* Science, **171,** January 15, 1971.

[14] Being published by IFI/Plenum.

[15] Produced by General Electric Company.

[16] *Survey of Scientific-Technical Tape Services*, Compiled and Edited by Kenneth D. Carroll. AIP ID 70-3, September 1970. American Institute of Physics - American Society for Information Science.

[17] *A Program for a National Information System for Physics and Astronomy 1971-1975.* Information Division, American Institute of Physics, June 1970, AIP-ID-70P.

CONSIDERATIONS IN ESTABLISHING A COMPUTERIZED FILE

Charles T. Meadow
Executive Secretary, COSATI*

Introduction

I would like to discuss this topic in terms of a series of questions that the prospective establisher of a mechanized file system should ask himself and to which he should provide tough, considered answers.

1. "Do you *need* a computer system?"
2. "If so, what kind?"
3. "Can you afford a computer system?" (i. e. can you pay *all* the associated costs, not just the purchase price?)
4. "Is your data in condition for computer processing?"
5. Lest these seem all pointed in the negative direction, "Can you get along without a computer?"

To get into more detail, we should start with consideration of the data and work back toward consideration of the need for a computer.

Preparing Data for Computer Processing

More than one information activity has discovered that the cost of converting existing files to machine readable or processable form can be the dominant cost in the development of an information system. By readable, here, I mean physically sensible to a machine; by processable, I mean sufficiently comprehensible to permit the machine to act upon data. These are quite different concepts.

Here is an example. One study of computer applications at the Patent Office estimated the *minimum* cost to convert their entire existing file of 6 million patent documents to machine readable form at $180 million. Or, consider the problem of converting the Library of Congress catalog—a file that cannot be taken out of service and which contains some handwritten records, not able to be read by any automatic reader. These, and other information operations, would face the problem of low return on initial investment in a computer system, until a sizable portion of the files is converted. Not always a happy prospect.

Information items which can be read and interpreted by humans may not be able to be read by a machine. In fact, if we think of business files we often find that records contain cryptic notes which serve to recall the real information, which is stored in the mind of the clerk—perhaps the very

*Now at the Atomic Energy Commission, Gaithersburg, Maryland clerk, seemingly expensive and inefficient, whom you are trying to replace with a machine. We must face the question, then, "Can you actually do without human analysis as a part of the retrieval process?" In recognition of this problem, modern information systems are increasingly relying on interactive techniques, so the human observer can remain in the loop, even with a mechanized information retrieval system. Also, it is the failure to recognize this question that, in commercial systems, causes so many of the irritating problems of computer-generated invoices—dunning letters for a zero-balance account, for example, or failure to remove a disputed item from an account after a verbal agreement to do so, because "there is no way to change the machine." Looked at in this way, many information files are more complex than they may appear and are less susceptible to totally mechanized processing than the energetic computer salesman might realize.

Do you have all the data that you need for your file? I. e., is your collection or file complete? If not, how are you goint to get it? *Can* you get it? Can you operate with an incomplete data base? Whether because the data is not yet assembled or because of conversion delays, is it possible for the proposed system to operate with a partial data base? Can you afford the complete hardware system before you assemble your complete file? If not, is there compatible software that will enable you to start with a smaller computer? Or, can you rent computer time?

Are there dissemination restrictions on your data that might affect the performance of your proposed system? Decide *now* how you are going to handle matters of privacy or security. Design your system around these restrictions. Do not postpone consideration of these restrictions until it is too late to do anything about them.

There are, of course, many information systems in which these problems, or most of them, do not arise. These are mostly systems characterized by the use of volatile data—data which is not stockpiled for any great length of time. This eliminates problems of conversion and gives the user the chance to change his procedures for creating or collecting the data to suit the requirments of the information system. An example is an airline reservation system. There is no great wealth of historical data to contend with here—schedules change, and reservations, once used, need not be accounted for. Still we find, even here, occasional references to problems. One is the releasability of information. The traveller who is pleased that the computer system keeps track of his business associate travelling companion, assuring that the two can be seated together even if they board an aircraft at different times, may not be so pleased with this infallible memory if the companionship is not on a strictly business basis.

It is not unusual for the cost of input preparation, including continuing handling as well as initial processing or conversion, to account for half the total cost of operating an information system. Yet, the subject rarely gets half the attention. It is not glamorous, but it is extremely, and in data processing, supremely important.

Paying the Price

Can you afford a computer system? Purchase price or rent is the most obvious cost in acquiring a computer and is perhaps the easiest to anticipate. But there are other costs, including:

—A staff of systems analysts, programmers, operators; people with high salaries, poorly-defined job functions and high turnover rate;

—Space devoted to the computer and its staff; and

—Replacement cost. Just as airlines cannot continue to use aircraft thr ughout their mechanical life, because of the pressure of competition, computers relatively rarely remain in place throughout their mechanical lives. As new generations of machinery are developed, the cost of maintaining the old, provisioning them (spare parts), finding programming staff willing to stay, and exclusion from the benefits of new software developed for the newer machines, all militate toward replacement, even where performance is apparently satisfa o-ry. Traditionally, the cost per unit of computation has gone down rapidly with succeeding generations of computers, so that replacement has some attractions, but the cost of hardware, to a continuing operation, is not going to be limited to the initial investment.

—Possible non-availability of data. While a computer dramatically increases the accessability of data, as compared to manual search methods, when a computer is "down" there is no access to the data. On the other hand, while data in file cabinets is not as accessible, the file cabinets are rarely down. Computers are becoming more and more reliable, but the possibility of a total outage, even if only for a few minutes, always exists. What will you do in this eventuality? For a typical library or information analysis center, the answer probably is to just wait. But not everyone can do this.

Another aspect of the availability problem is in the use of a time-sharing facility belonging to a contractor. This complicates your security and accessability problem. Cases have been reported of theft of information through a remote console, where the owner of the file, the computer from which the data was stolen, and the thief, are all at different locations.

The communications network lying between user and computer introduces further reliability problems. Furthermore, storing your files remotely means you do not have direct control over physical access to file storage areas, fire protection, etc.

What Kind of Computer System?

At this point, we assume we have decided that the data were in good condition and that the costs were bearable. What kind of system should be obtained? This is obviously much too broad a subject to try to cover in any great detail. Therefore I will try to concentrate on software systems, on the assumption, which may not always be valid, that if we know what we want a computer system to do, hardware selection is relatively easily. But even in hardware, there is so little in the way of performance standards that selection is often more arbitrary than we would like.

Software is the more difficult to evaluate because all of the problems of performance measurement are magnified, as compared with hardware, and because it falls upon software to take up the burden of the user who does not know enough about his data or his usage patterns. Mis-selection of software remains a widespread problem. The typical buyer does not know what questions to ask. The typical seller is unable to answer them, if asked. We have the further difficulty that, to a large extent, computer hardware performance is now dependent on the performance of systems programs, which are supplied by the manufacturer but whose efficiency is independent of hardware features that may promise greater speed. If this software is inefficient or defective, the total system will have problems, and there is little the unsophisticated user can do about it.

There are few information system users who make computer selection decisions based on a thorough analysis of the details of system software—the characteristics of the data access method which, for example, may have more effect on retrieval speed than the physical access speed of a disk unit. The particular software component of the computer's operating system that manages data storage and retrieval on a disk has some important characteristics. These concern the sequence of records in the file, the method of indexing the records, the method of changing or deleting records, and sensitivity to changes in record sequence. The overhead involved in a multiprogramming or time-sharing monitor can wipe out the speed advantage of a new computer. How many users question this while evaluating raw computer speed?

Available methods for evaluating software include "bench-marking" and simulation. Bench-marking is the testing of an application, usually with an approximation to the eventual software, and probably with an approximation of the eventual data files. The success of the method depends upon the success at approximating both the software and the files. The problem is that one cannot really know how successful the approximation is beforehand. Also, this tends to favor the large, rich bidder who can afford to set up a bench-mark, over the smaller company that may have better software, but no funds for elaborate demonstrations of it. Simulation programs are available, but these tend to simulate at a too detailed level. This can have the effect of making the validity of the simulation model dependent on the user's ability to predict fine detail when he is unsure of even general parameters. Some examples of parameters which are hard to predict are: rate of query, rate of change, and the area within a file receiving the most change (if not uniformly distributed).

This introduces the subject of just how much the user knows of the operating characteristics of his system at the time he makes a computer selection decision. Let me list a few of the critical characteristics, repeating some of those just mentioned:

—Usage rate of files;

—Modification rates (not just additions, which often can be accurately predicted, but also changes);

—Reliability-induced requirements for multiple copies of files, audit trails, file access protection;

—Performance speeds (e.g. retrieval time) required;

—Need for time-sharing or interaction (For retrieval? For file changes? If the answer to the latter is yes, do you understand the effect on the performance of the system? Was it available with the last time-shared system you saw demonstrated?); and

—Are there standards to which you must conform within your organization? Your profession? (of hardware, software, data structure or content)

Need for a Computer

At last, we come to the crucial question. . .are you going to need a computer? Not really a separate question, but the answers to the previous questions will largely determine the final resolution of this one. My questions have been mainly intended to steer away the information system operator who does not really need a computer, and by this I mean a user who, however much he may desire one, is unwilling or unable to pay all the prices. The real problems come when an organization cannot pay the price but does not know it.

The Benefit of a Computer

Let us now consider the other side of the coin, the reasons why a computer is needed. Here are some questions which may bring out that need.

Are you restricting needed services because you are unable to do the job by hand? (e.g. permit multiple file search, permit searching on other than the prime sort key, or perform iterative searching to help the user arrive at the best answer to his question).

Are you insisting, against all evidence to the contrary, that information users are able to formulate a mathematical statement of their needs (query statement) precisely, on their first try, without an intimate knowledge of the content of the files?

Are you providing the services your users want or the service they have learned to ask for? (i.e. are they adapting to what they think your limitations are, or are you adapting to their needs?)

Are you able to make the changes in file content that are required? Do you know, from actual test, the quality of your files?

As we can see, the "considerations in establishing a computerized file" are many and often complex. Ideal answers are rarely available, for two reasons: (1) Software suppliers and, to a lesser extent, hardware suppliers, are unable to predict accurately the performance of their products, and (2) users of information or managers of information services are unable to predict accurately the behavioral patterns of users, given a new form of information service. In other words, when the service changes, we frankly cannot predict what the changes will do to the user population.

The last of these points is the most important. Basically, we in the information business are supplying a service to human users. By changing the quality, quantity or price of that service, we are going to change the performance of the users. The value of our service should be measured by the value of that change. It is the value of the change in user performance, then, that must make the final determination of whether or not a computer is justified.

SOME INFORMATION PROCESSING TECHNIQUES FOR THE SMALL, SEMI-AUTOMATED, SCIENTIFICALLY-ORIENTED DATA CENTER*

by

David Garvin and William H. Evans
Institute for Materials Research
and
Blanton C. Duncan
Center for Computer Science and Technology
National Bureau of Standards
Washington, D. C. 20234

Abstract

Solutions are presented for several of the problems encountered in handling scientific text in machine-readable form in small data centers. The problems discussed are the selection of an adequate character set for representation of scientific text, the essential and useful features of editing routines, and batch-mode information retrieval.

Keywords: Chemistry; codes for information interchange; editing; information analysis centers; information processing; information retrieval; text handling.

Introduction

This paper describes how scientific text and data are manipulated in three information analysis centers at the National Bureau of Standards (NBS) that store their records on magnetic tape. The principal topics are the selection of an adequate character set, desirable characteristics of editing routines and the properties of a batch-mode retrieval program. We believe that our solutions can be used by almost any center that handles similar technical material.

The three centers are the Chemical Kinetics Information Center, the Chemical Thermodynamics Data Group and the Data Center for Atomic and Molecular Ionization Processes. The descriptions are drawn from current practice in these centers and from the General Purpose Document Image Code System which they share.

The remarks touch on only a few of the problems that confront the data center manager who must plan and live with automated records handling. But the topics covered are ones that deserve careful attention.

If a theme runs through all of these remarks, it is that of a general approach to text handling. Programs must be adaptable to many types of records. Any type of device that produces machine readable records must be an acceptable input device. Any type of printer must be accessible.

*Based upon a paper presented at the Forum of Federally Supported Information Analysis Centers, May 17-18, 1971, at NBS.

The adoption of a "general purpose" approach in our data centers has several bases. First, the three data centers have very different needs. Their managers know how to do the jobs with manual methods, and are unwilling to degrade those methods in favor of the machine. The possible uses of machine records files were unpredictable.

Secondly, computer hardware and software change rapidly. It appeared advisable to avoid being tied to any particular devices.

Thirdly, although the three centers have different needs, many of their records-handling problems are identical. It appeared more economical to build a basic system usable by all and then to supplement this with special programs for particular applications.

Finally, from the start there has been a strong desire to demonstrate that techniques for handling scientific text could be developed within a framework consistent with national and international standards for information interchange. That this can be done has been proved.

The remarks are addressed to two audiences. Those about the selection of a suitable character set are addressed to all operators of centers that deal with scientific text and data. The message is: accept no compromises, they no longer are necessary.

The remarks about editing and information retrieval are for a more select audience: the managers of small self-contained information analysis centers. A center with a staff of ten is large in this frame of reference.

The small center of concern is one that must arrange for all its computerized services, either by buying or renting existing packages or by having programs written specifically for it. Probably the main purpose of this center will be evaluation of data. In its first few years of operation, it will need computer techniques, but not elaborate ones.

The center that is, or can be, imbedded in a matrix of a computation center that provides many clients with a variety of text handling techniques is in a more favorable position. It can let somebody else worry about these details while it gets on with its main business. But even so, these remarks may be pertinent. They may help evaluate the available services.

Input to a Data Center File

Input to a computerized file is a very large topic. It deserves, and gets careful attention from data center managers. Planning at this point is important. The more carefully planned the input, the more effective the later use of the material. Also, input is the largest single task of a data center that collects material from an active field. As the sorcerer's apprentice learned, once you start the flow, you can't stop it.

Only a few facets of this subject are explored here. First, input should be easy. This means that the device used should be as much like a typewriter as possible. A typist, not a specialized operator, should be able to run it. She should be able to use all the techniques taught in typing classes. Today this means a typewriter that produces punched paper type or magnetic tape or a cathode ray display. The typewriters are still more flexible than the CRD's but the latter are improving rapidly. The keypunch is out.

Second, the record produced in the computerized file should be reasonably independent of the machine that produced it. This can be achieved, but it is hard work. But the work is worth it. This is because input devices have improved greatly in the past five years and will continue to change. Machine independence of records makes it practical to change input devices, and to build a modular system.

Machine independence and modular construction are illustrated in figures 1-4. Figure 1 shows a single purpose system, in this case devoted to preparing printed output. Modular construction is shown in figures 2-4. Input from various devices is converted into a common numerical code, in our case the General Purpose Scientific Document Image Code (GPSDIC). All manipulation programs process these GPSDIC records. Output to specialized printers starts with GPSDIC records not original input. It is simpler to program separately for input, output, editing and searching. One input program serves all of our input devices. It is tailored to each device by the insertion of translation tables.

Third, and most important, is the selection of a suitable character set. The set should be sufficient to permit input of the material handled by the data center, without serious approximation of the text. Our criterion is that it must be possible to input and store a scientific manuscript in the symbolism normally used. This, as it turns out, is sufficient for almost all other purposes.

This subject is only now receiving the attention it deserves from the hardware experts. The upper case alphabet is insufficient. The 88 characters on a typewriter are not sufficient for physics, chemistry, mathematics or library practice. But 188 characters can be sufficient, and, with the features described below, can be wildly extravagant.

The character set to be described is that for a General Purpose Scientific Document Writer. This was designed at NBS by B. C. Duncan[1,2]. It has been realized in a line printer and in a prototype punched paper tape typewriter. Several commercial machines come close to having the necessary features to produce all of the symbols. Usually the missing symbols can be constructed by overstrikes, as is commonly done for cent: ¢. The character set is the basis for the storage code used by our data centers. Figure 5 shows the type of text with which we must contend.

The features incorporated in the General Purpose Scientific Document Image Code are listed and discussed below.

(1) 188 primary symbols.

These are shown in figures 6 and 7. Figure 6 is the American National Standard Code for Information Interchange (1968), together with its control set[3]. These 94 characters are supplemented by another set of 94 (figure 7, left hand side). This supplement includes Greek letters, mathematical symbols, diacritical marks, line segments for chemical display formulae and a few special symbols. Thus this code is an extension of ASCII 1968.

(2) Each symbol may be used:

—as a main line symbol

—as a superscript or subscript

—in each of seven modifications (type faces)

(3) Binary combinations of symbols (overstrikes) are allowed. These may be used for

—letters with diacritical marks

—composite mathematical symbols

—as desired by the user

See figure 7, right hand side, for examples.

Taken together, these features permit one to record almost all scientific text "in the clear".

In terms of an input machine, or a line printer, the features (2) and (3) imply half line spacing and overstriking without erasure. Punched tape typewriters have these now, but they are still not generally available on cathode ray display devices. Surely they will be.

The reader who studies this set of characters will find things that he does not like. Probably, there will be widgets and wiggly lines favored in his field that are missing. Any character set will have approximations. This set attempts to provide reasonable coverage of common needs and to include acceptable alternative representations for very specialized symbols. We suggest that it is sufficient for data centers.

Use of an extended character set has an important impact on the internal operation of a data center. No longer must artificial abbreviations be introduced, Greek need not be written out, formulae are not confounded by putting subscripts on line. The instruction to the typist becomes "type it as it is". The simplification in operating procedures is very great.

Our experience shows that ASCII 1968 can be the basis for a practical text handling system. We hope that equipment designers will use the national standard code and its control set in specifying what an input or output device must be able to do. If they do, text handling systems will become more readily transferable from one machine to another and between data centers.

Editing

Records that are typed from rough copy require proof-reading and correction. These steps are part of the input process which, by successive approximations, produces acceptable copy.

Records that are part of an archival file may be selected and rearranged to make up the text of a report. Editing is also necessary at this stage to correct overlooked errors, polish the text or to insert directions needed in specialized printing programs.

One editing program in the General Purpose Scientific Document Image Code text handling system is used for both of these functions. It operates in batch-mode, but its features should be applicable to on-line editing. These features are cataloged below in two lists: the essential and the useful. The context in which the lists should be studied is the editing of an existing file. This is slightly different than the editing of material while it is being keyboarded for the first time.

Essential Editing Features

Delete lines

Insert lines

Substitute lines for existing ones

Change fragments of text

The fourth item, correction of fragments of text, may need explanation. It is used to alter a word, part of a word, or a phrase without disturbing the rest of the line. The book "1066 and All That" has an erratum that calls for the same procedures on grand scale: "For *pheasant* read *peasant* throughout". This technique is easily the most popular one in our data centers. It is almost always used when the text is complicated. The logic required for a general purpose fragment-correction routine can become involved, especially if it is to be efficient. It is well displayed in the SUBSTITUTE program in the EDPAC set[4].

Data center managers should make sure that their editing programs include these essential techniques. They should be easy to apply. The criterion is that they be easy for a typist to apply day after day, not that the center manager can figure out how to make them work.

Useful Editing Features

Change interline spacing

Reserve space between lines (leading)

Reorder lines

Justify and center lines

Make up paragraphs from uneven lines

Paginate as desired

Insert "canned" headings

Introduce typesetting commands

These will be wanted once the type of text that is to be edited escapes from the straightjacket of a collection of single lines of information. Data center managers should see to it that such features can be added easily to their editing programs.

Few of the details of the GPSDIC editing program are pertinent to this discussion. It is sufficient to note that the directions used to run the program are a series of commands each followed, if necessary, by lines of new text. The form of the commands is simple:

Delete page 3 line 4 through page 3 line 17

The required items are underlined. The form was borrowed from OMNITAB[5].

Batch-mode editing has one major disadvantage. It is not possible to check the success of the edit until the entire run has been completed. A second pass often is necessary. This is a very strong argument for the use of on-line techniques. Our system has another limitation. It requires that the editing be done in sequence from the first line to the last. The result of this limitation is that very few editing records are prepared on punched paper tape typewriters (in contrast to the preparation of text). Instead, the corrections are prepared on punched cards. It is far simpler to check out a deck of cards (and add a few) than it is to edit an editing tape.

Retrieval of Information

A common reason for creating a large scale information base in machine readable form is that, later on, one may readily retrieve selected portions for various uses. If the file of information is carefully structured, *and if* what is to be retrieved is known in advance, the retrieval scheme may be tailored to the records.

It may or may not be reasonable to hope that machine retrieval of information will be accurate and adequate, but it surely is folly to suppose that an on-going data center will know in advance what it will have to retrieve and how best to do this.

It was this uncertainty and a recognition that the format of input used by our data centers would change from year to year (and from problem to problem) that controlled the design of our first, and possibly last, search program.

The first design criterion was that any reasonable set of records should be legal input, that there should be no prescribed structure of the text. The second was that it should be possible to state the retrieval criteria for a particular search in logical Broken English. It was suspected that the formulation of correct Boolean statements would be beyond most users program. It is. Our technique, using formal grammatical phrases, is only marginally better. Both require very careful planning when the searching directions are complicated.

The GPSDIC search system designed with these criteria in mind is described briefly below. It is a granddaughter of the BLOCKSEARCH program by Mrs. C. Messina[4].

(1) The only *required* item of structure in the file is some repeating mark that divides the text into logical blocks suitable for examination separately. This "mark" need not be a special code. It can be any piece of text that regularly appears on the first or last line of a logical block of text. The maximum size of these blocks is dependent solely upon the amount of core memory available to store a block. At present we operate with a limit of forty 100 character lines per block. This appears to be adequate.

(2) The search of a logical block is made on a character matching basis: words, phrases or fragments of words in the text are compared against a search list.

(3) Either the entire block can be searched, or a part of it. In the latter case, the part to be searched is defined in the directions provided at run time. These directions may specify subsections defined by markers in the text or regions defined by character counts.

(4) Several independent searchs can be used in a single run.

(5) Either the entire block may be printed out at the end of a successful search, or only a part of it.

None of the properties stated above is unusual for a sequential search program. But the independence of the program from the structure of the records may be. This has meant that the same program is used for widely differing files, and even files constructed with no thought that they might be searched. Indeed, when asked how a file should be structured for searching by our program, we have very little advice to give.

The formal structure of the search commands is illustrated below. Words or phrases underlined here are directions about how the search is to be made. Words or phrases enclosed between slashes are items to be sought.

(1) A simple search.

Find /methane/ *and* /gas phase/ *and* /oxidation/ *end*

A search starts with *find* and stops with *end*. All three items must be present for success. The search stops as soon as one of them is found to be absent.

(2) A search with alternatives.

Find/methane/*or*/propane/*and find*/methyl radical/*or*/ethyl radical/*end*

Here one of the first pair and one of the second pair must be present. Any number of items could have been connected by *or*. The second pair could have been connected by *and* (with different results, or course). A formal Boolean statement for this search would be much longer.

(3) A sequence search.

Find /CH_4/ *followed by* /→/ *end*

This would define CH_4 as a reactant in a chemical equation (to the left of the reaction arrow).

The general structure of search directions is suggested by these examples. Intimate groups of items to be found are connected by *and* or *and not* or by *or* or *or not*. These groups are separated by "major connectives" such as *and find, or find, followed by*, and *or followed by*. The search is from the start of the list to the end. At each connective or major connective a decision is made whether irrevocable failure has occurred, or the patient is still alive. Several *find* . . . clauses may be included in one run. They are independent of each other. The word *end* in this case, appears only after the last clause.

In practice most of the records searched do have some internal structure that can be used to limit the material searched. The record shown in figure 5 is an example. The capitalized words at the left margin serve as dividers. A search for papers by certain authors can be made without scanning the entire record by using a *scan* . . . *to* . . . direction:

Scan from /AUTH:/ *to* /TITLE:/ *find* /Smith/ *but not* /Wesson/ *and scan from* /INDEX/ *to* final *find* /hydrogen atom/ *and* /ethane/ *write from* /BRIEF:/ to /AUTH:/ end.
In this example "final" is a direction that means scan to the end of the block. "First" is used in a similar manner.

The examples given above do not display all the features of the program, but they show all that is proper for a general discussion. Copies of the "instructions to users" and program listings will be provided to those interested.

The question that should concern the data center manager is this: is this type of retrieval technique appropriate? It is claimed here that this technique is *necessary*, but that it is *not sufficient*. The program described permits selection of material from an unordered file *on one pass*. Thus it permits a search of the input accumulated by a center, without the need for careful logical rearrangement of the files. This type of program, if nothing else, is a backstop that is needed. It certainly is the first technique a center should develop.

But in practice our data centers have wanted other tools. The Chemical Kinetics Information Center is the example. It maintains in hard copy an author index and an index arranged by journal reference. These classical tools are used very frequently to find specific items in the file. In addition it has a "keyword" or "descriptor phrase" index created from the (subject) index terms listed on each abstract in the file. Search strategy is often based on this listing. If there is very little information on a topic, this subject index is used as a set of "uniterm" cards for manual retrieval. If the search is likely to be complicated, the subject listings help organize a suitable set of search directions for the program described above.

The computerized files of this center are now of a size (about 10,000 records) that sequential searches are becoming costly. We intend to do most of our future batch-mode searching using an "inverted" subject index file, in the belief that this will be faster. If completely automated, this approach would mean a two pass system, which could be slower than the present one pass technique. However, this second pass (retrieval of the abstract or paper) will be manual most of the time. It usually can be done with shorter turn-around time than when using a batch-mode computer approach. Furthermore, we find that we almost always need retrieval of the hard copy. It is necessary to be sure that the article is pertinent to a request. Nothing turns off a user quite so rapidly as delivery of a batch of (to him) trash.

Concluding Remarks and Acknowledgements

In 1971 each of the ideas expressed here should be either standard practice or obsolete. The information industry has developed techniques that go far beyond those described here. They are available in large systems, or in specific installations. Small information analysis centers need a rapid and purposeful transfer of this information technology in order to make them more effective. Improved techniques must, somehow, be made available to these small businessmen at a price they can afford. The goal should be to let them evaluate data, untrammelled by a necessity to develop methods for manipulating records.

The programs and techniques which these remarks describe were developed primarily by three chemists, working as chemists, not information specialists. The system was designed by one of us (B. C. Duncan) and is based on his General Purpose Scientific Document Image Code. Most of the programs were coded by the present authors. Mrs. C. Messina, Office of Standard Reference Data, NBS, had made significant contributions in two areas: typesetting and development of EDPAC, from which routines have been adapted. Messrs. C. Albright, R. Chandler, and R. McClenon[6] added special programs. OMNITAB, developed by J. Hilsenrath, et al, has been our model for many of the control statements.

References

1. B. C. Duncan, "Prototype General Purpose Scientific Document Writer Installed," NBS Tech. News Bulletin *54*, No. 2, 35 (February 1970).

2. B. C. Duncan, "The Document Image Code," NBS Tech. News Bulletin *52*, No. 4, 86 (April 1968).

3. ANSI Standard X3.4-1968 Code for Information Interchange.

4. C. G. Messina and J. Hilsenrath, "Edpac: Utility Programs for Computer Assisted Editing, Copy Production and Data Retrieval," NBS Tech. Note 470 (1969).

5. a. J. Hilsenrath, G. G. Ziegler, C. G. Messina, P. J. Walsh, and R. J. Herbold, "OMNITAB: A Computer Program for Statistical and Numerical Analysis," NBS Handbook 101 (1968).
   b. S. T. Peavy, R. N. Varner, and D. Hogben, "Source Listing of OMNITAB II Program," NBS Special Publication 339 (1970).
   c. S. T. Peavy, R. N. Varner, and S. G. Bremer, "A System Programmer's Guide for Implementing OMNITAB II," NBS Tech. Note 550 (1970).
   d. R. N. Varner and S. T. Peavy, "Test Problems and Results for OMNITAB," NBS Tech. Note 551 (1970).
   e. D. Hogben, S. T. Peavy, and R. N. Varner, "OMNITAB II User's Reference Manual," NBS Tech. Note 552 (1970).

6. R. C. McClenon and J. Hilsenrath, "Reform: A General-Purpose Program for Manipulating Formatted Data Files," NBS Tech. Note 444 (1968).

Figure Captions

FIGURE 1. Diagram of a program dedicated to a single job: preparing and printing via computer.

This example is dedicated to the printing of a report. A manual analog is a manuscript, a typist, and the typed copy.

FIGURE 2. Diagram of a general purpose program that produces archival records *not* dedicated to any output devices. This structure is used in the GPSDIC system.

FIGURE 3. Input/output independent record manipulation and file maintenance.

All the techniques shown are independent of input and output. Ideally, they should be independent of the storage code.

FIGURE 4. Diagram of a general purpose program that uses archival records to print on any available output device. Both the records and the output are independent of the input device used.

In a modular system the choice of output device may be made long after the records are prepared.

FIGURE 5. Sample record in GPSDIC.

Output from a line printer developed at NBS to handle scientific text. The character set used is that in Ref. 1, fig. 4.

FIGURE 6. The ASCII 1968 character set and control codes. Text and rules typed on a Model 37 Teletype. The control set consists of the items in columns 0 and 1 plus "SP" (space) and "DEL" (delete). The remainder of the table shows the 94 printing graphics.

FIGURE 7. Additional characters in the GPSDIC set.

The Shift Out (SO) set of 94 is the complete array. The SO set of 32 is appropriate for a machine that can print 128 characters. The composites are binary combinations of characters. This figure is the current (1971) GPSDIC set.

The character set has been modified slightly during the past few years, to include new uses and to bring it into correspondence with the International Standards Organization R 646.

SINGLE PURPOSE PROCESSOR

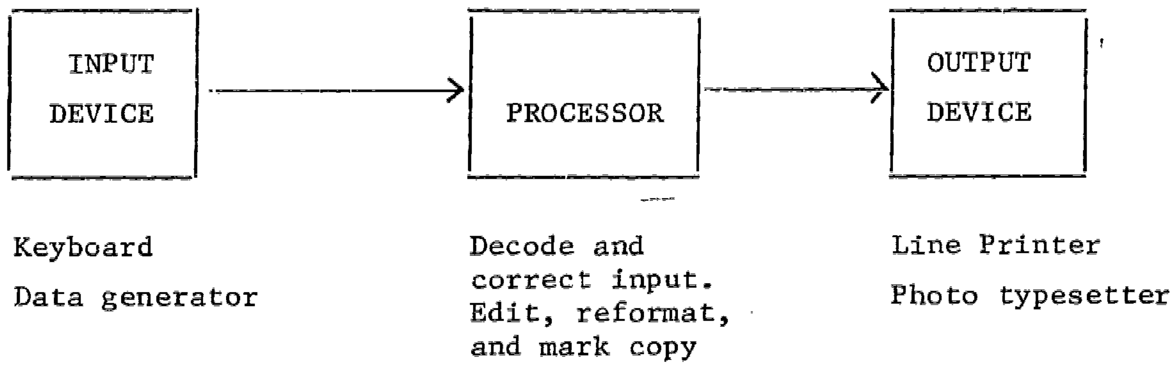


FIGURE 1. Diagram of a program dedicated to a single job: preparing and printing via computer.
This example is dedicated to the printing of a report. A manual analog is a manuscript, a typist, and the typed copy.

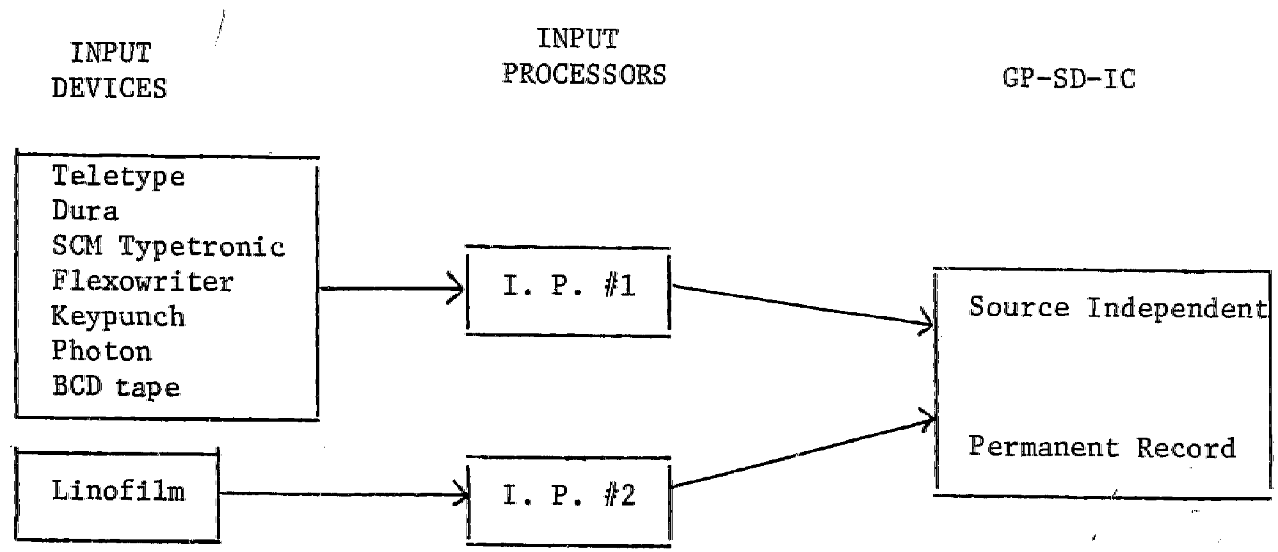


FIGURE 2. Diagram of a general purpose program that produces archival records not dedicated to any output devices. This structure is used in the GPSDIC system.

RECORDS MANIPULATION

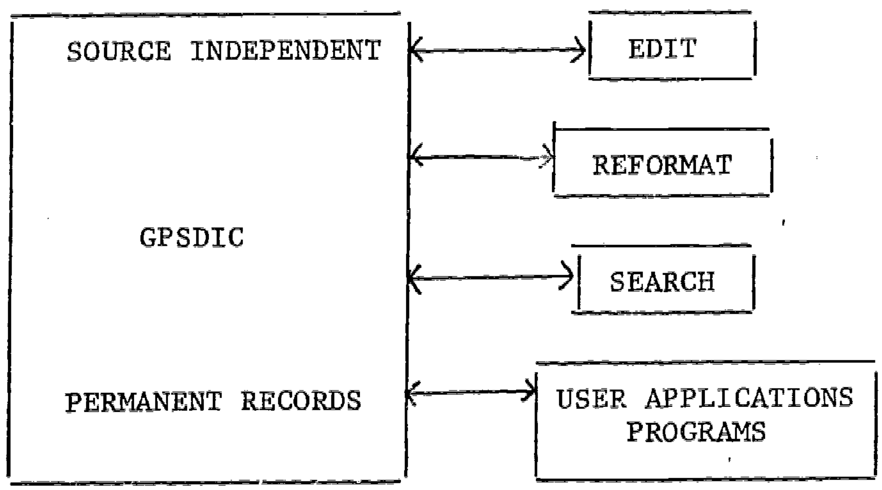


FIGURE 3. Input/output independent record manipulation and file maintenance.
All the techniques shown are independent of input and output. Ideally, they should be independent of the storage code.

63

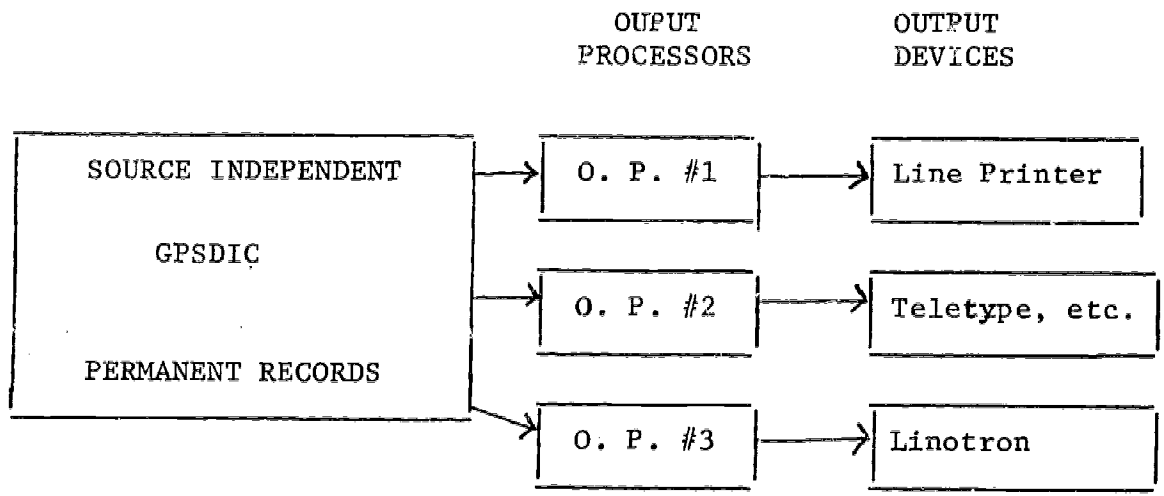


FIGURE 4. Diagram of a general purpose program that uses archival records to print on any available output device. Both the records and the output are independent of the input device used.
In a modular system the choice of output device may be made long after the records are prepared.

Book	1 Page 3 +...30...35...40...45...50...55...60—	y	Ln eedd
BRIEF:	FW/4569 Page 1 CHPLB-1968-2-143 CKIC/15403 —	2	1
AUTH:	Brennen, W., and Shane, E. C. —	5	2
TITLE:	Pressure-Dependence of the Yellow Nitrogen —	8	3
	Afterglow Intensity —	11	4
REF:	Chem. Phys. Letters (Amsterdam) 1968 2 143 —	14	5
REACT:	$N + N + M \rightarrow N_2(A^3\Sigma_u^+) + M$ —	18	6
	$N_2(A^3\Sigma_u^+) + M \rightarrow N_2(B^3\pi_g) + M$ —	22	7
	$N_2(B^3\pi_g) + M \rightarrow N_2 + M$ —	26	8
	$M = N_2$ —	29	9
	$N_2(B^3\pi_g) \rightarrow N_2 + h\nu$ —	33	10
	$N_2(A^3\Sigma_u^+) + M \rightarrow N_2 + M$ —	37	11
INDEX:	Experimental: gas: Bånd cleaved NN: —	40	12
	Bond formation NN: pressure: energy-transfer: —	43	13
	fluorescence: excitation: quenching: rate: —	46	14
	radiative: electronic: second-order: —	49	15
	chemiluminescence: nitrogen-molecule (product): —	52	16

FIGURE 5. Sample record in GPSDIC.

Output from a line printer developed at NBS to handle scientific text. The character set used is that in Ref. 1, fig. 4.

ASCII Code

	0	1	2	3	4	5	6	7
0	NUL	DLE	SP	0	@	P	`	p
1	SOH	DC1	!	1	A	Q	a	q
2	STX	DC2	"	2	B	R	b	r
3	ETX	DC3	#	3	C	S	c	s
4	EOT	DC4	$	4	D	T	d	t
5	ENQ	NAK	%	5	E	U	e	u
6	ACK	SYN	&	6	F	V	f	v
7	BEL	ETB	'	7	G	W	g	w
8	BS	CAN	(	8	H	X	h	x
9	HT	EM	)	9	I	Y	i	y
10	LF	SUB	*	:	J	Z	j	z
11	VT	ESC	+	;	K	[	k	{
12	FF	FS	,	<	L	\	l	|
13	CR	GS	-	=	M	]	m	}
14	SO	RS	.	>	N	^	n	~
15	SI	US	/	?	O	_	o	DEL

FIGURE 6. The ASCII 1968 character set and control codes. Text and rules typed on a Model 37 Teletype. The control set consists of the items in columns 0 and 1 plus "SP" (space) and "DEL" (delete). The remainder of the table shows the 94 printing graphics.

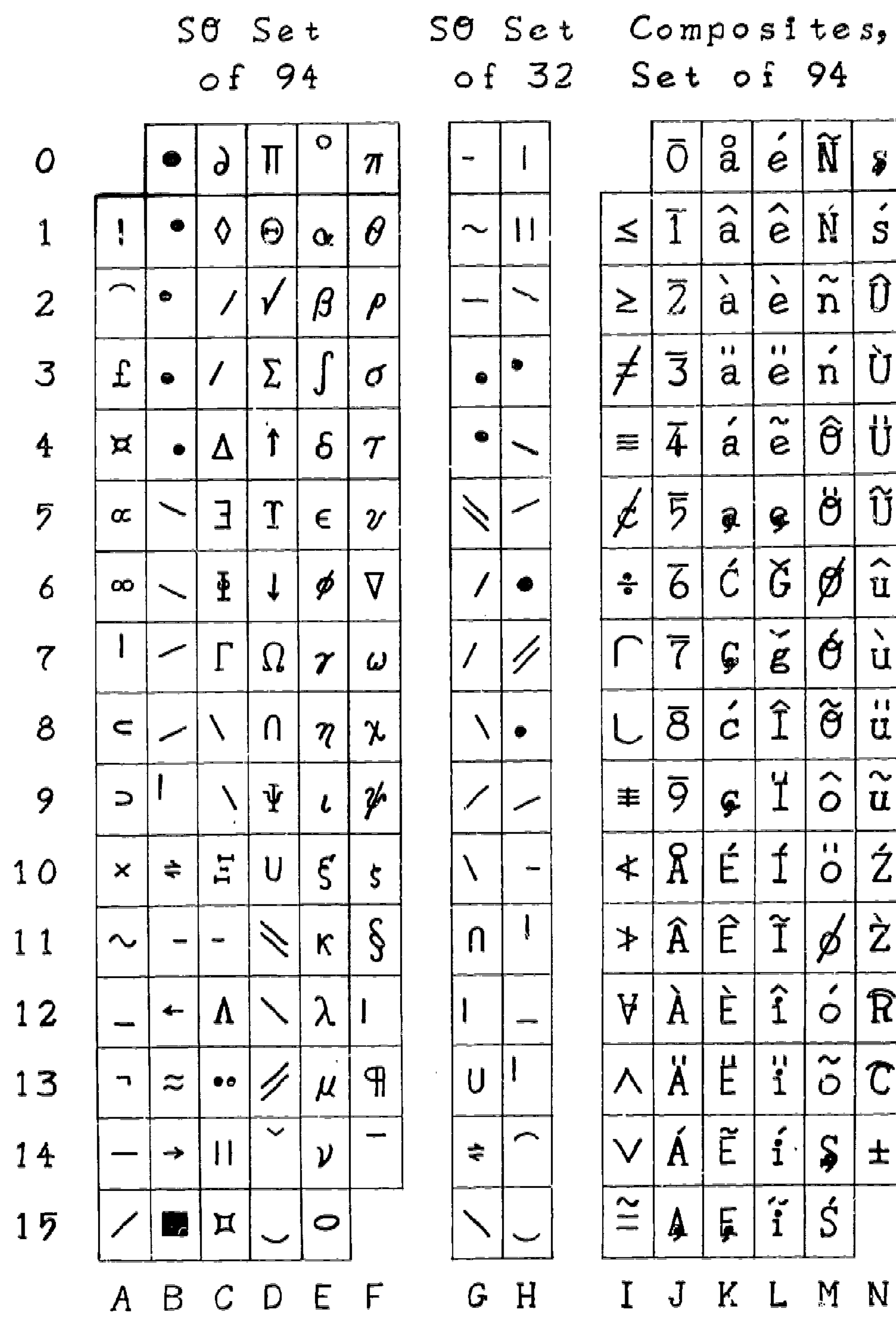


FIGURE 7. Additional characters in the GPSDIC set.

The Shift Out (SO) set of 94 is the complete array. The SO set of 32 is appropriate for a machine that can print 128 characters. The composites are binary combinations of characters. This figure is the current (1971) GPSDIC set.

The character set has been modified slightly during the past few years, to include new uses and to bring it into correspondence with the International Standards Organization R 646.

A CASE STUDY OF USER ACCEPTANCE OF AN INTERACTIVE RETRIEVAL SYSTEM, SOME THOUGHTS ABOUT CASE STUDIES, AND A THOUGHT ABOUT LEGITIMAZATION

Dr. Don H. Coombs
Director, ERIC Clearinghouse on
Media and Technology
Stanford University

Rather than just report on our experience with one particular interactive retrieval system, Lockheed's DIALOG, I would like to take a slightly broader view. By way of introduction I would like to philosophize briefly on different ways to evaluate retrieval systems. Then I will discuss DIALOG, and I'll finish up by suggesting a new concept which may have relevance to information retrieval—to all kinds of information retrieval, batch processing as well as on-line. That concept, which I propose at least half-seriously, is *Legitimazation*.

I. Introduction

There are various ways to evaluate retrieval systems. My prejudice, coming as I do from the Institute for Communication Research at Stanford, is toward measures which involve people—toward behavioral measures. I don't wish to suggest that these are the only worthwhile measures, but only that they can indeed be worthwhile.

In trying to use behavioral measures to evaluate interactive retrieval systems, we are at a primitive level—the case study. (I consider what we did with DIALOG as primarily case studies, even though there was some scaling involved.) We are at the case study stage, although there are lots of more valuable approaches than the case study. But you do case studies when there is no better way to get a grip on a problem.

The ideal alternative to the case study for evaluating interactive retrieval systems is obvious. You assemble all the available systems, a variety of data bases (although this may be ridiculously ambitious), and a relatively large number of representative users. Then, using a design to control for as many threats to validity as possible, you measure ultimate user satisfaction.

The importance of working with "representative users" is often overlooked; if you want to generalize to the universe of potential users, you would be well advised to involve a probability sample of them in your testing. Most retrieval system evaluation is done by information retrieval specialists—by systems people or, even worse, by hardware people. At first glance that's a rather good situation, like having an expert mechanic tell you how good a car is, when you're interested in buying it. But that's not a good analogy. A better analogy would be having General Motors tell you how good the Vega is—and that's precisely what those colorful, and expensive, brochures that they give away in the showroom are all about. It may be a good way to sell cars, but it's a poor way to evaluate performance.

To top off that analogy, it would be like having the General Motors dealer evaluate the Vega for you when he himself drives nothing but Cadillacs. That is my way of suggesting that the designers of a retrieval system probably are a long way from being, themselves, representative of the potential users.

Now, why am I not presenting a decent behavioral evaluation of different interactive systems here this afternoon? For at least one reason: It hasn't been possible to set up such a procedure—to make many systems available, at the same time, in comparable (and realistic) circumstances. A year or so ago we had ERIC files available on three interactive systems at the same time, from computer terminals in our clearinghouse. There was Lockheed's DIALOG, System Development Corporation's ORBIT, and Stanford University's own SPIRES. But we really didn't have anything like comparable situations, which would allow fair conclusions to be drawn.

Just to illustrate why the situations weren't comparable: The clearinghouse staff was more familiar with DIALOG, having had it installed first. And because of contractual arrangements, it was easier and cheaper to get DIALOG "up". Neither ORBIT nor SPIRES was then available with a cathode ray tube for quick visual display, and the SPIRES system was still in process of development.

Since that time there have been notable attempts to present different interactive systems in something like competitive situations, such as at recent conventions of the American Society for Information Science. One flaw has been that most of the systems were operating only with toy files, which leaves open a good many questions about real-life performance.

Recently U.S. Office of Education personnel have gone through an exercise which approaches comparison of different interactive retrieval systems. They did this in awarding a contract for such a system, to be available at a number of east coast sites. It's my understanding that Lockheed's DIALOG, the system I will be describing today, won that contract. If any of you are interested in information on that project, I refer you to Harvey Marron or Chuck Hoover at the Office of Education.

I think the reason I have gone through this introduction is so you would be sure I had no pretensions that what we did with DIALOG approached high science. We did the best we could, at the time. I find it encouraging that today we would, I think, do better.

II. User Acceptance of an Interactive Retrieval System

To turn to what we did, I need first to get on record the way DIALOG operates. Rather than a detailed description, this will be an extremely rudimentary explanation.

The commands which allow the searcher to manipulate the file on the computer are relatively simple. Each of the special characters on the terminal keyboard above the numerals (such as @ and %) stands for one command or one type of manipulation. The principal commands used in searching are the EXPAND, SELECT, COMBINE, DISPLAY and KEEP.

Briefly, the EXPAND command can be used to bring onto the CRT a "window" or a "page" of the alphabetical index where a particular term is located, giving the number of citations posted to the terms as well as the number of cross-referenced thesaurus terms listed for each. Each visible entry is marked with a reference number plus the letter "E" (E1, E8, etc.). The EXPAND command also allows the user to look at the thesaurus. (Both uses can be seen on page 1 of Fig. 1, which is an annotated terminal record.)

The SELECT command allows the searcher to set aside for future use any terms which he wants to incorporate in his search. The typed terminal record indicates the identification or set number which is assigned to each group of documents as it is set aside. (See page 1, Fig. 1.)

After selecting out every term in the file which relates to each concept in his search, the searcher can COMBINE (as on page 2, Fig. 1) the terms appropriate to each concept by adding (in Boolean logic, ORing) the terms together to create a new set. When the selected terms have been grouped according to concept, the concept sets are then COMBINED again, this time so that the sets intersect (in Boolean logic, the terms are ANDed).

The resumes for the new, narrowed set of documents can then be brought to the screen one by one, using the DISPLAY command (see page 3, Fig. 1.). The searcher may now "page through" on the CRT what he has retrieved, selecting (KEEPing) those documents which he will wish to examine further in hard copy or microfiche form. Finally, the results can be printed off-line, or they can be typed on the terminal printer. The format for this printout is also at the option of the searcher.

An attempt was made to get people in a variety of professional roles to sit down at the remote access terminal for two-hour sessions. The nine evaluators were:

1. A researcher engaged in the planning, technical aspects and conduct of educational research projects (28, M).

2. A motivational educator in private practice, working with children referred by schools, doctors, etc. (41, F).

3. A graduate student in education who will return to district level to work (25, M).

4. An M.D. engaged in psychiatric research and therapy (29, M)

5. An assistant professor of linguistics and computer science (27, M).

6. A university librarian directing library automation, and involved in developing a different on-line retrieval system (40, M).

7. A professor of education teaching and doing research in educational psychology (52, M).

8. An elementary school teacher (24, F).

9. A secondary school teacher doing graduate work to assist him in developing media programs at his school (26, M).

SEARCH TITLE: ADULT BASIC EDUCATION PROGRAMS

DATE: 07/11/69

REQUESTOR: PERSON A

The 4 column headings indicate what command was given, what set number was assigned (if a set was created), how many documents are contained in that set, and finally a description of its contents.

COMMAND-OPERAND(S)	SET NO.	NO. IN SET	DESCRIPTION OF SET (+=OR, *=AND, -=NOT)	
E-IT/BASIC ADULT ED				*The index around the descriptor, BASIC ADULT EDUCATION, was expanded (E). No entries were posted to this term.*
S-E9	1	59	IT/BASIC READING	*However, BASIC READING was there and was selected (S) to form set 1.*
E-E9				*BASIC READING was then expanded by its reference number, E9, in order to display its thesaurus entries. There was nothing of interest.*
E-IT/ADULT BASIC ED				
S-E5	2	126	IT/ADULT BASIC EDUCATION	*ADULT BASIC EDUCATION was expanded and then selected to form set 2.*
E-E5				*It was expanded again by the reference number to show the thesaurus entries.*
S-E12	3	57	IT/LITERACY EDUCATION	*The term LITERACY EDUCATION was found there and selected.*
E-IT/REMEDIAL PROGR				*Next, REMEDIAL PROGRAMS was expanded, and it and a number of alphabetically related terms were selected (sets 4 through 9).*
S-E3	4	63	IT/REMEDIAL INSTRUCTION	
S-E4	5	4	IT/REMEDIAL MATHEMATICS	
S-E5	6	61	IT/REMEDIAL PROGRAMS	
S-E6	7	58	IT/REMEDIAL READING	
S-E7	8	6	IT/REMEDIAL READING CLINICS	
S-E8	9	22	IT/REMEDIAL READING PROGRAMS	
E-E5				*REMEDIAL PROGRAMS was then expanded by reference number.*
S-E14	10	63	IT/COMPENSATORY EDUCATION	*Two more terms (sets 10 and 11) were located in its thesaurus entries.*
S-E17	11	94	IT/EDUCATIONALLY DISADVANTAGED	
E-E14				
S-E18	12	145	IT/COMPENSATORY EDUCATION PROGRAMS	*COMPENSATORY EDUCATION was then expanded by its reference number (E 4) and one more relevant terms. was located in its thesaurus expansion (set 12).*

E-IT/ADULT EDUCATIO				*The searcher now turned to the adult aspect or concept of his search. Expanding ADULT EDUCATION produced three relevant terms (sets 13, 14, 15).*
S-E2	13	8	IT/ADULT DEVELOPMENT	
S-E5	14	231	IT/ADULT EDUCATION	
S-E8	15	117	IT/ADULT EDUCATION PROGRAMS	
E-IT/ADULTS				*Expanding ADULTS produced four terms to be selected (sets 16 through 19).*
S-E4	16	184	IT/ADULT VOCATIONAL EDUCATION	
S-E6	17	1	IT/ADUT VOCATIONAL EDUCATION	
S-E1	18	40	IT/ADULT STUDENTS	
S-E5	19	87	IT/ADULTS	
E-IT/BASIC SKILLS				*And finally, BASIC SKILLS, which relates to the earlier concept, was expanded and selected.*
S-E5	20	50	IT/BASIC SKILLS	
C-3-12/+	21	504	3+4+5+6+7+8+9+10+11+12	*Set 21 was created by the union or addition of sets 3 through 12. This set then included most of the remedial or basic education terms.*
C-1+20+21	22	595	1+20+3+4+5+6+7+8+9+10+11+12	*Set 21 was then added to sets 1 and 20 to create the basic education concept group containing everything in the ERIC files indexed by one of these terms.*
C-13-15/+	23	352	13+14+15	*Set 23 was created by adding some of the adult terms together.*
C-23+18+19	24	456	18+19+13+14+15	*This process was completed by the addition performed in set 24, which now represents the adult concept in the search.*

Note that so far set 2 has been ignored, because it pre-coordinates the two concepts in the search and therefore should not be included in either concept set.

C-22*24	25	39	(1+20+3+4+5+6+7+8+9+10+11+12)*(18+19+13+14+15) *Sets 22 (the basic or remedial education concept) and 24 (the adult education concept) were next combined (C) to form an intersection, the AND in Boolean logic, with the resulting set 25 containing 39 items indexed by at least one term from each concept set.*

D-25
K-25/1
K-25/2
K-25/3
K-25/4

This set was then displayed (D) and the first four retrieved items examined one by one on the CRT. The relevant ones, in this case all that were examined, were set aside using the keep (K) command into a reference set for future attention. This reference set is arbitrarily numbered 99.

S-IT/ILLITERATE ADU 26   33   IT/ILLITERATE ADULTS

In the examination of the first four items of set 25, a new term was turned up, ILLITERATE ADULTS, which had not been located earlier. This term was now selected directly (without going through the expansion)

C-26*22   27   21   (1+20+3+4+5+6+7+8+9+10+11+12)*26

The resulting set 26 was combined by an AND operation with set 22 (basic education terms) to form set 27.

C-25+27   28   57   ((1+20+3+4+5+6+7+8+9+10+11+12)*(18+19+13+14+15))+((1+20+3+4+5+6+7+8+9+10+11+12)*26)

The results of this combination and the previous one (set 25) were then added together to form set 28. Note that the number of items in 28 is not equal to the sum of the items in sets 25 and 27. This is so because a combination creates a set of unique documents where no item is repeated a second time.

D-28
K-28/2
K-28/7
K-28/8
K-28/10
K-28/12
K-28/13
K-28/14
K-28/16
K-28/18
K-28/19
K-28/20
K-28/21
K-28/23
K-28/25
K-28/26
K-28/27
K-28/28
K-28/29

Set 28 was then displayed item by item and the relevant ones set aside in the reference set.

K-28/30-57				*After examining 30 references and keeping 22 of them, the evaluator determined to keep all of the remaining 17 and proceed further with the search.*
C-2-28	29	109	2-(((1+20+3+4+5+6+7+8+9+10+11+12)*(18+19+13+14+15))+((1+20+3+4+5+6+7+8+9+10+11+12)*26))	*Set 29 was created by subtracting the items already examined from set 2 the ADULT BASIC EDUCATION set. This avoids duplicate printing of any items.*

A second aspect of the search was begun to turn up industrial and job training programs which dealt with basic skills.

E-IT/INDUSTRIAL TRA				*INDUSTRIAL TRAINING was expanded and selected and then expanded to its thesaurus entry. This produced six additional relevant terms.*
S-E5	30	62	IT/INDUSTRIAL TRAINING	
E-E5				
S-E13	31	56	IT/INDUSTRIAL EDUCATION	
S-E14	32	14	IT/INPLANT PROGRAMS	
S-E15	33	113	IT/JOB TRAINING	
S-E16	34	10	IT/OFF THE JOB TRAINING	
S-E17	35	82	IT/ON THE JOB TRAINING	
S-E19	36	240	IT/TRADE AND INDUSTRIAL EDUCATION	
C-30-36/+	37	512	30+31+32+33+34+35+36	*The seven industrial training terms were then combined by an OR operation to produce set 37.*
C-37+16+17	38	591	16+17+30+31+32+33+34+35+36	*Set 37 was then added to sets 16 and 17 which also relate to the same general concept.*
C-38*22	39	18	(1+20+3+4+5+6+7+8+9+10+11+12)*(16+17+30+31+32+33+34+35+36)	*This sum was intersected with the basic education concept group (set 22) to get set 39.*
P-99/5		1-50	ITEMS HAVE BEEN PRINTED	*Finally three prints (P) were initiated of sets 99, 29, and 39. Format 5 which contains the indexing, cataloging, and abstracts for each document was chosen. After the print had been completed off-line at Lockheed the results were sent to the clearinghouse for forwarding to the evaluator.*
P-29/5		1-109	ITEMS HAVE BEEN PRINTED	
P-39/5		1-18	ITEMS HAVE BEEN PRINTED	

Each user was asked to come about 30 minutes before the time the system became operative, and at that time filled out a form giving information about himself. The user's questions about the system were answered, and he was shown the terminal. When the DIALOG programing was loaded in core and the ERIC document file made accessible on an IBM Data Cell, the visitor sat down at the console and performed all searching himself. He was coached throughout the session, and prompted to make use of different aspects of the system until he had some familiarity with it.

After the session was over, there was a structured "debriefing." The transcripts of these were coded to produce some relatively objective summary results, and the transcripts also provided verbatim answers.

One big question to be answered was whether individuals with no previous experience could sit down at a terminal and, in a reasonably short time, use such a system effectively. The answer was yes.

For the most part, evaluators were enthusiastic about their two-hour experiences. Before they were prompted to comment on specific aspects of the system, the visitors were encouraged to put on record whatever impressions they wished to report. The two aspects of the system which were most frequently commented on were 1) its speed, and 2) the way it "widened horizons," the way it suggested other relevant areas of information or different approaches to the information originally sought.

Some of the UNPROMPTED statements about the "horizon-widening":

> " . . . It opened up new avenues for thought."
>
> " . . . It expanded areas that I hadn't considered as being related to the subject."
>
> "It had a sort of fallout of new ideas and possibilities."
>
> "I was amazed at . . . what possibilities it offered for further learning."

After each of the nine evaluators had commented generally on DIALOG, he was asked to specify the good aspects of the system. There were a total of 44 good points singled out. (This is a little like compiling batting averages in Little League, but the total score would be 44 good points, 28 bad points listed. Of greater value than the 44/28 breakdown is the finding that there were 25 different good aspects reported, 18 different bad aspects.

Six of the evaluators noted the speed of the system and its saving of user time. The next most frequent favorable observations were that being able to combine sets was very desirable (volunteered by 5 evaluators) and that using the system had opened new avenues for thought, or "widened horizons" (volunteered by 4 evaluators). Three of the evaluators commended the system for being simple to use, easy to work with.

The evaluators then were prompted to identify what they considered bad features, but they were not asked leading questions. Delays in waiting for the system to accept and execute a command were singled out as a bad feature by four evaluators, as was the feeling that considerable experience or time was needed to master all the operating rules.

Other critical comments:

> "Too many combinations of keys are needed to input one command."
>
> "Having to build combined sets one step at a time, rather than using parentheses, and doing it with one complex statement is inconvenient."
>
> "There is a great deal of 'paging' required on the CRT, because you can only look at nine terms at a time."

Changes in the DIALOG system, made since our study, have obviated those last two criticisms.

Evaluators were specifically asked about the "pacing" of the system because some of us at the clearinghouse came to be critical of delays when it was necessary to wait to input the next command. Only four of the nine evaluators were at all critical of delay; most were so impressed by the performance of the system that any delay was of no consequence. One felt that such variation in pacing suggested that we were running the machine, rather than vice-versa. The general conclusion was evident: At least while learning to use the system, few persons are bothered by having to wait sometimes to enter commands.

As already reported, the sessions were successful in locating relevant information for the nine visitors. The question or questions which they brought to the session were answered. But two other features of the system were evident: Users were led to ask additional questions about the chosen area of investigation, and to pursue entirely different matters than those which originally concerned them.

First I'd like to deal with the "intellectual fallout" or "horizon-widening" effect of the system. Seven of the nine evaluators reported that they had asked additional and different questions about the subject which they originally were investigating, and seven said that they came upon material on different, though related, subjects which they would like to pursue at a later date.

Verbatims:

> "I began to realize that they had some articles of international scope . . . [and that] kind of opened up that area . . . ."
>
> "Well, as we began to look at the section on instructional television, there were some related topics there that I hadn't been aware of . . . . There were a couple of topics that interested me, one in the area of teacher training, which I just happened to run across, but I would like

to look into at another time, although it wasn't particularly related to this study."

"I formulated new areas to look under, for relevant research."

" . . . I stumbled on something, I stumbled onto [specific] programing languages."

". . . The biggest problem was staying with what I had originally pursued, instead of getting off on other interesting things."

"I did run across some other information [that I would want to pursue later] . . . . We've got our mind on one thing, particularly when we're researching something, one phase, and we only think to look in certain areas, and the thing that I liked about this, the machine points out that there may be some other related information . . . . This is very significant, very helpful."

I myself assume that such "horizon-widening" effects are extremely desirable. I can't imagine a research administrator with such a narrow area of concern as to object to this aspect of a system. To look at the situation from a different viewpoint, this aspect tends to bring the useful documents in the collection to the user's attention even though his preconceived ideas of what is available are incorrect.

Next, to consider the basic interactive aspect of the system: Six of the evaluators had favorable comments on the way the system made it possible to monitor and modify searches, but the other three should not be considered negative or indifferent. Most of the evaluators had no previous experience with computerized retrieval systems, and so could hardly compare DIALOG with non-interactive, batch-processing systems. Our feeling was that the evaluators reacted to DIALOG as an entity, and that the overwhelmingly favorable general comments were to a great extent the result of this very basic aspect of the system.

The individual verbatims:

"It makes a big difference, because you get a feeling of control over your search that you don't have so much when you're actually in the library. There it's hard to remember exactly which things you were going to go back and do—you have to write things down, you have to organize things. Here you have handy little systems for putting something off somewhere and you just organize in your own mind the very basic concepts."

"It has a much more organizing effect, it helps to organize in a much more effective way."

"It helped me . . . . I think the fundamental help was I had some idea of the amount of information I was handling, or would be handling if I had it printed out. And it gave me some insight, too, into the amount of research that had gone on in certain areas."

"It's like having a great mass of information at your disposal, where you can somehow set up and know where you are and how much you've looked at."

"I think it made a great difference."

Besides the nine case studies presented in this report, there are some observations on DIALOG which are a product of its use in the clearinghouse for a variety of tasks. These applications can be categorized as duplicate checking, preparing in-house projects, and answering the information requests of visitors and others users who contacted us by mail or phone. To summarize in succinct fashion, the system proved extremely valuable in such uses.

In all, 46 people had a demonstration and introduction to the system, 68 people had their requests searched by a staff member while they were present to interact with the system, modifying the search strategy as necessary, and 21 people (including the nine evaluators) were taught to use the system and had hands-on experience. These people who used the clearinghouse as a source of ERIC materials were able to do their literature searches efficiently.

Perhaps most important here is the ease of use of the system. For people who are unfamiliar with computers and who have only a limited amount of time to devote to their professional research and to learning to use a new research technique, no matter how powerful it may be, it is quite important that the technique be simple to understand. Experience in demonstrating DIALOG and instructing people in its use indicates that it is fairly simple and does not overwhelm the person unfamiliar with computers.

It is interesting to note that the real difficulty in teaching people to use DIALOG had nothing to do with the system itself. Rather, it was the concept of coordinate searching that proved to be difficult. If the individual understood how coordinate indexing worked, it took only minutes to acquaint him with the few mechanical procedures which would allow him to search the file that way. However, the linear method of searching out materials is ingrained in most people, and time is required to help them understand coordinate searching.

No idiosyncratic search strategies emerged in the nine case studies, and this was a disappointment. I have a long-time, although seldom implemented, interest in cognitive structuring, but a retrieval system tends to constrain search strategies. Any system does. It is designed to be used in certain ways, and so it is hardly surprising that people use it in those ways. A system designed specifically to investigate idiosyncratic search strategies is conceivable, but the essential flexibility and complexity would make it quite expensive.

To sum up our experience with DIALOG, it was favorable indeed. But the whole project—and especially our experience with the nine evaluators—undoubtedly was heavily influenced by Hawthorne Effect. How the nine users would have felt about the system after its novelty had worn off is something we don't know.

Anyone wishing a more complete report on our experience with DIALOG can obtain the full 90-page document from the ERIC Document Reproduction Service, P.O. Drawer O, Bethesda, Maryland 20014, as document number ED 034 431 (on fiche for 65¢, in hardcopy for $3.29).

III. Legitimazation

In closing, I'd like to suggest half-seriously that mechanized retrieval systems are serving a new function—that of *Legitimazation*.

Let me give you an example of Legitimazation—a worst case, at least as far as ethics are concerned. Some U.S. Office of Education research contracts specifically require that literature searches be completed, for what I imagine are obvious reasons: because the investigators should have a good idea of what has been done already before they get their own projects underway.

Several times, when we had DIALOG available in the clearinghouse, we were approached by educational researchers—or their graduate assistants, because that demonstrates how important literature searching is considered to be—and we were asked to perform an exhaustive search of the ERIC files for relevant material. In each of these cases the search was required as part of their Office of Education contract. And in each case, there was a great sense of urgency—because everything else about the project had been completed, and the report already had been written.

Now that's Legitimazation in the worst sense: Using a retrieval system just so you can say that you used it. It's like not wanting to know how to cure sick people, but wanting a M.D. certificate to put on your wall in a nice frame.

Why does machine search lend itself to legitimazation so well? Because it's easier—or seems to be easier—to describe what was done. For example, "The complete ASDEC file was searched for relevant documents using the Quest III system running interactively on our 360 Model One Million." That has great specificity, compared to "A graduate student spent three weeks in the library." People know that's no good, because they know about graduate students and they know about libraries. Being able to cite a mechanized search, in contrast, is like putting a certificate on the wall from a good-sounding medical school.

This makes legitimazation sound all bad, which I don't think is the case. There's a legitimate use of Legitimazation, if you will. And that is akin to someone buying insurance. If you've ever been in a position to help someone search a file, and found lots of relevant documents, you may have observed that when you laid the 300 abstracts on him, the person didn't smile. The systems people smiled, because look at all the relevant things their system produced. But the poor user didn't smile. Either he wanted the three or four *most relevant* documents, or else he wanted just Legitimazation—he wanted to find that there weren't any relevant documents, so that he could go ahead with his work and not worry about something like it having already been done. Or worrying about how it meshed into any big picture.

It's all very well to say that research has to be conducted in a framework of previous research, so that findings can be hooked up, but it's another thing entirely to complicate someone's life with more potential hookups than he cares to deal with. (I am not speaking of how science and technology should operate, I am speaking of how people do operate.)

What our man, our last example, wants is assurance that he's out in the clear and hasn't overlooked anything. He's willing to pay for that assurance—for that insurance—in money and in time for a machine search. In return for paying that premium, he is protected against disaster; if someone has done exactly what he's up to, the blame falls not on him but on the Quest III system running on a 360 Model One Million.

Now why spend time mentioning Legitimazation? Because if that is a real function of a retrieval system—if people don't want information, many times, at all, but do want insurance—then that should be taken into consideration in evaluating retrieval systems. Most of our measures are based on the assumption that the user wants great masses of output, and often, I think, that's not true.

Let me put it another way: If we set up a committee to evaluate retrieval systems and the committee-members have certain standards in mind, and the superior system is chosen and provided to users—

The system is more likely to be successful if the standards of the committee are similar to the standards of the potential users.

I think Legitimazation is one of the functions desired by users. Maybe we should change our standards, or maybe we should change our users. Changing one is probably easier than changing the other, but I'm not arguing for a particular course of action. I'm just suggesting that some attention be paid to the situation.

INPUT TECHNIQUES FOR TECHNICAL INFORMATION

by
Joseph Hilsenrath
Office of Standard Reference Data
National Bureau of Standards

Abstract

A summary is presented of recent progress at NBS in the automation of book production through the development of techniques for computer-assisted phototypesetting. The strength of the system rests on general-purpose edit-insertion programs and other general-purpose programs which accept a variety of input media. The programs take existing files on punched cards or computer tapes; or Magnetic Tape Selectric Typewriter (MTST) cartridges; or files keyboarded on-line to a time-shared text editing system; and transform them to match the requirements of the phototypesetting system at the U.S. Government Printing Office (GPO).

Examples are shown of finished text consisting of upper and lower case Roman and Greek characters, subscripts and superscripts keyboarded on a variety of input devices. The examples are from input on punched cards, from a 44 key Selectric terminal and from a "scripting" teleprinter capable of typing 126 characters in two colors in inferior, superior or main line positions.

Keywords: Computer-assisted printing, computer input, electronic typesetting, input techniques, keyboarding conventions, phototypesetting, text automation.

1. Introduction

Little did the planners of this Forum realize when they assigned me the seemingly mundane topic of input that they would really be giving me *carte blanche* and that I would take the opportunity to sound off on a number of my pet peeves and talk about some of my favorite people.

I find myself, increasingly in recent months, in the position of a doctor who is asked to prescribe a cure for a patient who sends a relative to the office with a description of his symptoms. My advice invariably is, come back with the patient, let me examine him, and then I can prescribe.

My assignment today doesn't allow me time to ask any of you to describe his data handling headaches. So, what remains for me to do is to describe to you a few of our more or less miraculous cures, even though I doubt that such a recital is any more ethical in the computing game than it is in the medical profession. You are still entitled, I believe, to a brief explanation of the experience and the biases which underlie my comments this afternoon.

*Based on a talk presented at the Forum of Federally Supported Information Analysis Centers, May 17-18, 1971.

When we joined the Office of Standard Reference Data about four years ago, we resisted the advice of colleagues to issue a cookbook for data and file organization and for bibliographic formats. We saw no profit in that kind of effort because generalities are of little use; to be more specific would require us to learn your job even better than you know it yourself; and finally, any *dicta* on our part would restrict the whole operation to the ingenuity of one person, or a small group of people. We chose instead to build a tool kit of programs which could take any systematic file arrangement and play games with it, that is rearrange it at will to any one of a number of alternate arrangements or formats. We've had enough experience with a number of generalized programs to convince us that economical solutions of varied data handling, and typesetting problems as well, lie in general rather than in special purpose programs.

We now have a number of interesting instances where an existing program, not at all intended for the new job at hand, solved that job more elegantly and more efficiently than we could have solved it had we addressed ourselves to the solution of that problem directly. Some of the publications that have gone through our typesetting programs at the Bureau have produced dramatic savings (about $3000 in one issue alone). If we did not have on hand a general purpose program able to cope with the specific requirements of that publication, the cost of writing and debugging an *ad hoc* program would easily have exceeded the savings from computer-assisted typesetting.

Before I discuss our own programs, you should know what experience we have had with programs developed by others. We had some good experience, a few years ago, with IBM's Text-90 [1] system, and can therefore recommend its successor, Text-360, for many types of reports, especially as the input now connects with a terminal on-line. But even card input to Text-360 is a viable and attractive way of using a computer for document preparation. Those of you who have 360's, are advised to look into this text editing and formatting package. It has excellent page make-up facilities and can be used to feed a phototypesetting process.

If I continue a bit further with a recital of our experiences, I should mention that we make extensive use of on-line keyboarding and text editing, via an IBM 2741 terminal, into a number of commercially available text-editing services that support IBM's Administrative Terminal System (ATS). IBM called it DATATEXT, a local outfit in Washington called it VIPCOM, a company in New York calls it Word One. These are all minor variations of ATS. If you have an IBM 360, (Model 50 or up), and put in an ATS system, most of your text handling headaches will be solved overnight at miniscule cost. A number of universities have made it available on their machines. As an example, the University of Iowa offers ATS to its staff at $2.00 per hour connect time, plus an appropriate charge for storage. Even at commercial rates of $3.00 to $3.50 an hour, which we now pay, the system is viable. We use it extensively in keyboarding ordinary reports and for very fancy computer-assisted typesetting as well. There are a number of examples in the exhibits here which you might look at later. They represent a rather small fraction of the display we have prepared for those of you who come back on Wednesday to see some of the work that has gone through our automated systems. There will also be demonstrations of a variety of inputting techniques, both on-line and off-line, utilizing MTST's, 2741 terminals and a variety of Teletypes.

Most of the phototypesetting production, that we've been involved in at the Bureau of Standards, has gone through the Government Printing Office on the Linofilm machine. That work is

Numbers in brackets indicate references at the end of the paper.

fed through one or another of the typesetting programs of the Government Printing Office (GPO) for which our programs provide input tapes. We are also making increased use of the Linotron 1010, a much faster, and more reliable machine. Most of that work goes through the GPO's Master Typography Program.

For the last few years we've been developing our own software which often permits us to bypass the typesetting programs at the Government Printing Office and allows us to drive the Linotron directly. This has been a fairly exciting development which has cost us relatively little, about one woman-year, and has proved rewarding to the GPO, as well as to us. By us, I mean primarily the Data Systems Design Group, which I lead in the Office of Standard Reference Data, and the staff of the Computer-Assisted Printing Section in the Office of Technical Information and Publications, with whom we work very closely. If any of you have notions about the limitations of the Linotron for scientific text, please spend a little time with Carla Messina or Rubin Wagner before you leave Gaithersburg and learn how they have been able to tame the Linotron to do their bidding. Mrs. Messina's software, which has been installed at the GPO, now makes it practical and efficient to set on the Linotron 1010 complicated technical material containing as many as 1020 different characters.

On the way to the GPO, to paraphrase a popular title, we flirted briefly with a Photon typesetter. By flirting, I mean that we put a few small publications through that machine. We are now experimenting with a Stromberg Carlson 4060 computer-driven microfilm device at the Goddard Space Flight Center. We also have a low priority interest in seeing whether our programs can be used to drive other electronic typesetting devices. That interest stems from our desire to be of service to information analysis centers that do not have access to GPO facilities, and must rely on commercial phototypesetting services.

2. Punched Card Input

Much of our information, and yours, has already been generated on magnetic tape, from punched cards, so that the question of keyboarding afresh is not a problem. A problem never-the-less remains if one wishes to get away from the upper case character set so characteristic of what Dr. Blanton Duncan calls Stone-Age printout. We have a potent medicine for that problem in the form of a program called SETLST which is described and listed in NBS Technical Note 500 [2].

The programs KWIND and SETLST accept punched cards or magnetic tape records normally intended for line printers and produce a magnetic tape properly flagged and transformed to interface with one or another of the typography programs at the GPO. The result is graphic arts quality in upper and lower case and typeset in mixtures of typefaces (bold, italic, bold italic, small caps, etc.). Figures 1 and 2 show the typographic variety in products of this program produced from a magnetic tapes that contained only capital letters to start with.

In other applications of this program, words (See Figure 3) such as ALPHA, BETA etc., have been replaced by the greek letters α, β, etc. When the material is in tabular form, even the letter G standing alone in a fixed field can be recognized by the program to produce its greek equivalent. The generality of the SETLST program arises from the fact that it gets its information on how to format a specific job from a set of control cards supplied at run time. The typographic information is supplied only on the control cards. It is not contained in the program. We have other general-purpose programs which accept card input and achieve fancy output. Figure 4 shows a portion of a table of spectroscopic data phototypeset from punched cards containing just digits and capital letters.

In spite of these striking examples, we recommend punched cards for input only in special circumstances. Punched cards require either special paraphenalia (EAM equipment) or a batch mode computer facility. We now favor on-line systems for keyboarding as these are becoming increasingly available at reasonable cost.

Our recent efforts have, therefore, been directed to developing keyboarding techniques for typewriter-like devices that provide readable copy while capturing the character stream in machine-readable form, either on paper tape or (preferably) on the disk of a time-shared computer.

It should be emphasized that neither the computer nor our programs achieve the illustrated transformations on their own. Nor are such transformations always practical even if possible. If the data is not itself flagged, transformations are feasible only when the data base is suitably structured or otherwise systematic. Isolated exceptions to systematic transformations can also be handled if they are known to the person who is operating on the file. What is significant about our approach is that the details of the transformations are supplied in the form of *control cards* tailored to the job rather than *programs* so tailored.

3. Keyboarding on a Scripting Typewriter

The difficulty of keyboarding scientific text on primitive devices was illustrated most dramatically by the New York Herald Tribune on January 31, 1929 and again the next day when they published the full text - equations and all - of Albert Einstein's paper on the unified field theory. The mathematical equations were translated into words which were cabled along with the text. The equations were reconstructed from the words, were written by hand, and were printed as line drawings.

This unprecedented and still unequaled journalistic scoop itself attracted enough attention so that the editors were moved to explain in detail how the formulas crossed the Atlantic Ocean over ordinary telegraph cables, since ". . . . cable codes are equipped only for the transaction of human affairs in ordinary arrangements of letters and numbers . . . and not for . . . complex arrangements of Greek, Roman, and Gothic letters used in mathematical formulas."

The problem, stated so well in 1929, has remained unsolved for over 40 years. Certain abstract journals still spell out Greek characters and reduce mathematical formulas to a linear notation. Only in the last two years have there become generally available on the market, machines capable of generating and transmitting a code structure that can handle scientific text in its full-blown glory (to borrow a phrase from Dr. Garvin). In Figure 5 we see an excerpt from the Einstein paper as it appeared in the Herald Tribune and as it would be keyboarded on, and transmitted by a Model 37 Teletype today. The transmission can be to another Teletype device or directly to a computer. We have pieces of this text stored on the computer at Dartmouth College and can retrieve them at will. The next time I have an opportunity to retell this story, I should be able to show how this portion of the Einstein manuscript looks when listed on the high speed printer, developed at the Bureau of Standards, that Dr. Garvin alluded to, and how it looks after it is phototypeset.

We now have in process two major publications - an article for a mathematical journal and a book on statistical designs - which have served as a test bed for one of our newer systems for automated publication. In this system material is prepared on the Model 37 Teletype which has forward and reverse half-line indexing for subscripts and superscripts; can type 126 different characters (including the Greek alphabet), and punches a paper tape consonant with the typed copy. After all corrections have been made in the paper tape, it is converted to magnetic tape from whence it is run on the computer into the GPSDIC system which produces a computer output on the extended character printer (GPSDIC train). This computer printout contains sufficient information to serve in place of a conventional galley proof. Errors that are discovered at this stage can be corrected in the batch mode. When the galley is deemed satisfactory, the material is run through a number of programs developed by Mrs. Carla Messina for justification (without hyphenations) and for processing to produce a magnetic tape ready to mount directly on the Linotron 1010 at the GPO.

Experience gained with these pilot publications has confirmed our basic preference for on-line keyboarding over paper tape operations and especially for on-line editing instead of off-line paper tape correction followed by batch mode editing.

The availability of "scripting" teletypewriter devices interfacing with suitable teleprocessing computers will, I believe, be recognized soon to offer solutions to many text-processing problems that have heretofore been characterized in the literature as "unsolved".

These devices for which a few determined pioneers have been waiting nearly 10 years should have an important impact on computer usage beyond text processing. They make it possible to enter mathematical problems into the computer in natural mathematical notation for direct computations along lines that have been spelled out clearly in the literature since 1963. I refer to the work of M. B. Wells at the Los Alamos Scientific Laboratory [3,4], H. J. Gawlik at the Royal Armament Research and Development Establishment [5], and M. Klerer at the Hudson Laboratories [6,7].

In these systems it is often sufficient to feed the computer the *statement* of the problem rather than its *solution*. In Figure 6 we see an example of a problem stated in terms of words and symbols natural to the discipline in which the problem arises. In the MIRFAC system that problem statement is all that the computer requires to obtain the solution. This is not an isolated instance. The system handles problems of much greater mathematical complexity with equal facility. Klerer and May have written a compiler which is uniquely suited to the solution of mathematical problems involving complex display formulas. In Figure 7 we see again a computer program which is simply the statement of a clearly defined mathematical problem.

Now that suitable input devices are available at a reasonable cost, we would hope to see such compilers implemented on more ubiquitous computers, so that the time we engage in "computing without programming" will exceed the time we spend in "programming without computing."

4. Scientific Text on 44 Keys

Since our efforts to automate book production at NBS started when the Model 37 Tel ype was in its early development, we settled for an input device which though it was more primitive in its character set (88 characters) and had no scripting capability, could, however, be connected to a commercial on-line text-editing service. We considered, that the advantage of an economical on-line text-editing system outweighed the recognized "disadvantage" of using a 44 key typewriter, and devised a simple and now clearly viable keyboarding convention for handling scientific text without compromising the notation. The use of an existing software-hardware combination (IBM's ATS) allowed us to turn our full attention to the design and implementation of a comprehensive software package as an interface between the archival tape produced by the ATS system and the typography programs at the GPO. As the original motivation for our mechanization was the preparation of a large abstract publication, the software system builds author and keyword indexes using the typesetting flags to identify items to be indexed.

An example of the notational complexity which is achieved routinely by the system, which for the lack of a better name we have dubbed the *44 key system*, is shown in Figure 8. The typographic information is conveyed in this system in two ways. The systematic use of boldface for titles and volume numbers and of italics for journal abbreviations, and for the variety of indentions is controlled by preceeding each of these portions of the text by a different number of tabs. Other typeface changes occurring within the title or the abstract are provided by a system of overstrikes. Thus, any character overstruck by a / turns on the italic face; an equal sign turns on the boldface, etc. A comma, and a double quote used in this fashion produce respectively subscripts and superscripts and a right parenthesis produces greek characters. Figures 9 and 10 afford a comparison of two keyboarding techniques.

The typing convention for the *44 key system*, which has been in productive use for nearly four years by the staff of the Computer Assisted Printing Section and a number of Technical Divisions, has produced approximately 5000 typeset pages of published output. The manner in which the system handles the notational complexity of NBS manuscripts, coupled with the advantage of on-line operation account for much of the acceptance that the system has gained at NBS. For those groups at NBS who do not share our preference for on-line keyboarding and editing, the Computer Assisted Printing Section is equipped to convert MTST (Magnetic Tape Selectric Typewriter) cartridges to computer readable magnetic tape to feed into our edit insertion programs. The keyboarding convention is the same for both the on-line use of ATS and the off-line use of MTST machines.

Successful as the 44 key system has been for abstract bulletins and conference proceedings, it is not a system we would recommend to individual authors for preparation of manuscripts involving chemical and mathematical expressions. For that purpose a scripting typewriter with a fuller character set, provides the author with a manuscript fully as legible as we have all become accustomed to seeing come from the hands of a capable typist on a conventional typewriter augmented perhaps by special keys, or more recently on a typewriter with interchangable type spheres.

Such copy is in fact produced on the Model 37 Teletype. Unfortunately, currently available ATS text editing services do not accept input from ASCII coded terminals. Until we have the ability to connect a Model 37 Teletype to an economically viable text editing system, we must cope with paper tape corrections off-line. When we are able to do on-line editing from a Model 37 as easily as we now can with a 2741 terminal, we will be able to achieve manuscript automation literally at the authors desk. Such automation at the source is now technically feasible at NBS and shows promise of substantial savings in time and money.

References

1. Hilsenrath, J., and Waibel, K., "Computer Assisted Text Preparation", Technical Report TR-67-47 (July 1967), Computer Science Center, University of Maryland, College Park Md. 20742. Also available from the National Technical Information Service as AD657457.

2. Messina, C. G., and Hilsenrath, J., Edit-Insertion Programs for Automatic Typesetting of Computer Printout, NBS Technical Note 500 (April 1970). Available from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (Price 60 cents).

3. Wells, M. B., "MADCAP: A Scientific Compiler for a Displayed Formula Textbook Language", Comm. ACM, Vol. 4, pp. 31-36 (1961).

4. Wells, M. B., "Recent Improvements in MADCAP", Comm. ACM, Vol. 66, pp. 674-678 (1963).

5. Gawlik, H. J., "MIRFAC: A compiler Based on Standard Mathematical Notation and Plain English", Comm. ACM, Vol. 6, pj 545-547 (1963)

6. Klerer, M., and May, J., "An Experiment in a User-Oriented Computer System", Comm. ACM, Vol. 7, pp. 290-294 (1964).

7. Klerer, M., and Grossman, F., "Editing and Type Composition of Two-Dimensional Mathematical Text via Computer", IEEE Transactions of Engineering Writing and Speech, Vol. EWS-11, No. 2, pp. 53-64 (August, 1968).

National Bureau of Standards KWIC Index of Engineering Standards

Standard Ball Mill Test for Friability of	Abrasive Grain (1965)	ANS	B74.8
Standard Test for Bulk Density of	Abrasive Grains (1964)	ANS	B74.4
Standard Test for Capillarity of	Abrasive Grains (1964)	ANS	B74.5
Standard Sampling of	Abrasive Grains (1964)	ANS	B74.6
Grading of	Abrasive Microgrits (1967)	ANS	B74.10
Coated	Abrasive Products (1955)	USC	R89
Grading of Abrasive Grain On Coated	Abrasive Products (1967)	USC	PS8
	Abrasive Wear (1968) HS J965	SAE	J965
Standard Safety Code for the Use Care, and Protection of	Abrasive Wheels (1964)	ANS	B7.1
Silver Polish,	Abrasive + Dip-Type (1966)	IRC	1.15
Markings for Identifying Grinding Wheels and Other Bonded	Abrasives (1958)	ANS	B5.17
Malleable Iron	Abrasives (1961)	SFS	2161
Cast Steel	Abrasives (1966)	SFS	20T66
Specifications for Tumbling Chip	Abrasives (1967)	ANS	B74.11
Oil Lavandin	Abrial (1960)	EOA	163
	Abridged Volume Correction Tables for Petroleum Oils (1951)	AWP	F2
Standard	Abridged Volume Correction Tables for Petroleum Oils (1960)	ARE	XYZ339
	ABS Lavatories for Recreational Vehicles (1967t)	IPM	TSC16
Acrylonitrile-Butadiene-Styrene	(ABS) Plastic Drain, Waste and Vent Pipe and Fittings (1965	USC	CS270
Acrylonitrile-Butadiene Styrene	(ABS) Plastic Pipe (schedules 40 And 80) (1969)	USC	PS18
Acrylonitrile Butadiene Styrene	(ABS) Plastic Pipe (standard Dimension Ratio) (1969)	USC	PS19
rd Method of Te-. for Recovery of Asphalt From Solution By	Abson Method (1965) ASH T170 Standa	AST	D1856
Z11.181 Standard Method of Test for Ultraviolet	Absorbance and Absorptivity of Petroleum Products (1965) ANS	AST	D2008
ntimeters (1968t) Methods of Test for	Absorbance of Polyethylene Due To Methyl Groups At 1,378 Ce	AST	D2238
Recommended Practice for Calculation of	Absorbed Dose From X or Gamma Radiation (1968)	AST	D2568
) ANS K065.15 Standard Method of Test for	Absorbed Gamma Radiation Dose In the Fickle Dosimeter (1963	AST	D1671
Water	Absorbency of Bibulous Papers (1964)	TAP	T432TS

Figure 1. A portion of a QWIC index produced on the Linofilm at the GPO from a tape generated at NBS from a master tape which contained only capital letters. Note the initial caps and the circled exceptions.

146. Ehmann, W. D., Huizenga, J. R.
Bismuth, thallium and mercury in stone meteorites by activation analysis.
Geochim. Cosmochim. Acta, **17**, *125–135 (1959).*
(ENGLISH). ARGONNE NATIONAL LAB., LEMONT, ILLINOIS.

147. Eichholz, G. G.
Activation assaying for tantalum ores.
Nucleonics, **10**, *No. 12, 58–61 (1952).*
(ENGLISH). RADIOACTIVITY DIVISION, DEPT. OF MINES AND TECHNICAL SURVEYS, OTTAWA, CANADA.

151. Facchini, U., Orsoni, L.
A method for the determination of uranium content in minerals by using the fission of U–235.
Nuovo Cimento, **6**, *241–254 (1949).*
(ITALIAN) (ENGLISH SUMMARY). CENTRO INFORMAZIONI STUDI ESPERIENZE (C.I.S.E), MILANO, ITALY.

Figure 2. A portion of a test run of a bibliography produced on the Linotron 1010 at the GPO direct from a tape generated at NBS using the program SETLST. Here the source consisted of fixed field records in all caps. While the need for so much typographic variation is doubtful the exercise illustrates the flexibility of the general-purpose programs employed here.

ELEMENT	QUANTITY	TYPE	ENERGY MIN	ENERGY MAX	DOCUMENTATION REF VOL PAGE	DATE	LAB	OCT. 1, 1967 PAGE 7 COMMENTS	5 BORON SERIAL NO.
B	Diff Elastic	Eval Rept	4.3 6	1.4 7	NAA-SR-11980IV	4/67	AI	CAMPBELL+ LEG COEFS TABLE C-M SYSTEM	38734
B	N Production	Expt Rept	6.0 4	1.8 6	ANL-5567	6/56	ANL	LANGSDORF 5ANGS LEG COEF ONLY LABSYS	39629
^{10}B	Evaluation	Eval Rept	1. –4	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING TOT SEL SNE SIN ABS ND NT NA	38857
^{10}B	Total X-Sect	Eval Rept	1. –4	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 94 POINTS TABLE+CURVE	38856
^{10}B	Elastic	Eval Rept	1. –4	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 94 POINTS TABLE+CURVE	38855
^{10}B	Diff Elastic	Eval Rept	5.5 5	1.5 6	NAA-SR-11980IV	4/67	AI	CAMPBELL+ LEG COEFS TABLE C-M SYSTEM	38733
^{10}B	Diff Elastic	Eval Rept	5. 5	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 14 PTS LEG COEFS F1-F8TBL+CRV	38848
^{10}B	Nonelastic	Eval Rept	1. –4	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 94 POINTS TABLE+CURVE	38854
^{10}B	Tot Inelastic	Eval Rept	1.2 6	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 66 POINTS	38853
^{10}B	Diff Inelast	Eval Rept	1.2 6	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 4PTS ENERGY DISTR OF NEUTS	38847
^{10}B	Absorption	Eval Rept	1. –4	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 94 POINTS	38852
^{10}B	(n, d)	Eval Rept	4.8 6	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 20 POINTS	38851
^{10}B	(n, t)	Eval Rept	1.2 6	1.4 7	ORNL-TM-1872	0/67	ORL	IRVING 67 POINTS	38850
^{10}B	(n,α)	Eval Rept	1. –4	1.4 7	ORNL-TM-1872	0/57	ORL	IRVING 94 POINTS	38849

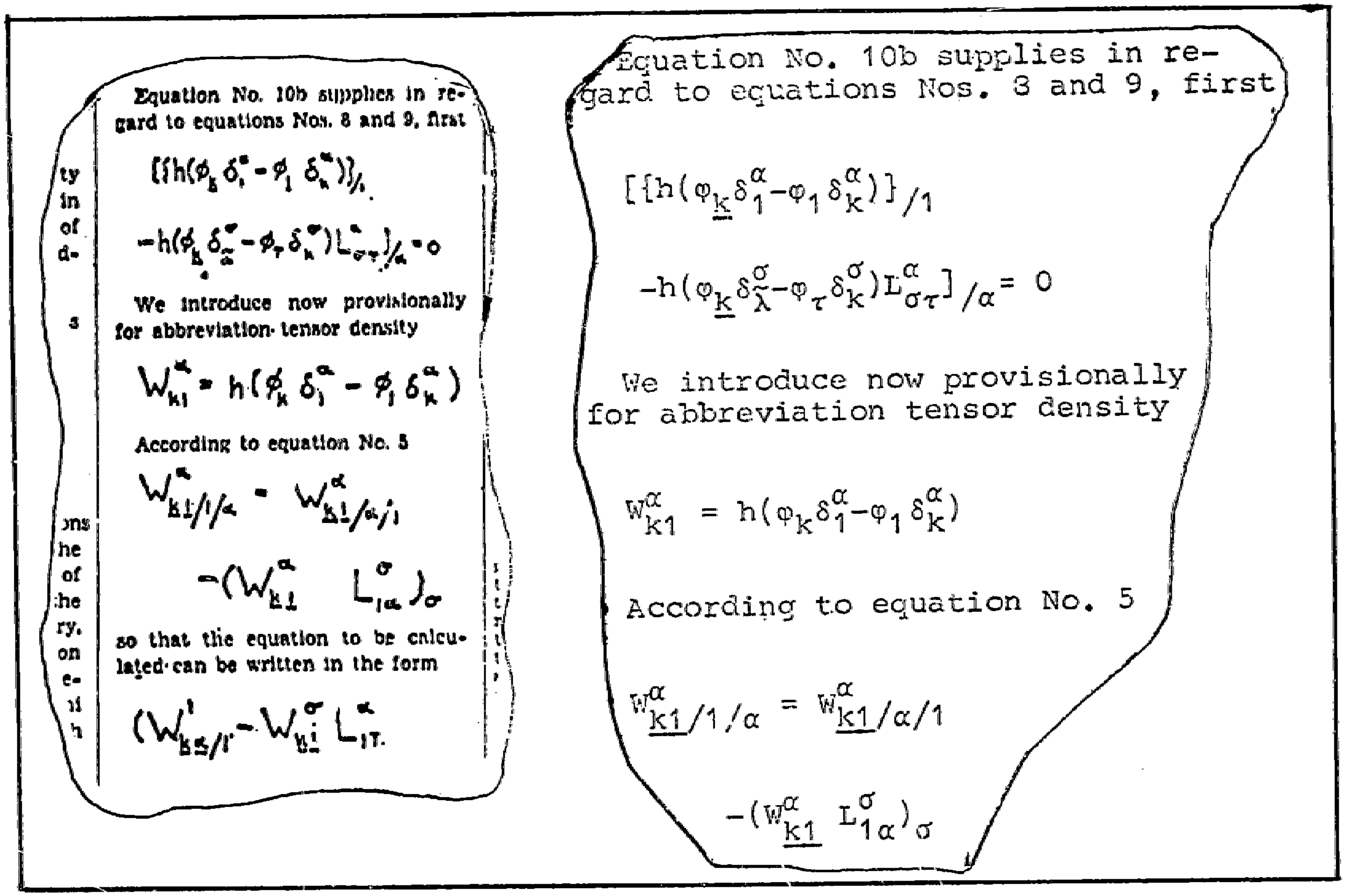
Figure 3. A portion of a test run of a Computer Index to Neutron Data (CINDA) showing the extent of transformation achieved in the left hand portion of the line by the program SETLST operating on a magnetic tape record written in all caps and with greek letters spelled out.

TABLE 3. *Observed and classified lines of* W I

Wavelength Å	Intensity		Wavenumber (cm^{-1})		Classification
	Arc	Spark	Observed	o−c	
2746.734	40		36396.11	+0.02	15_3— 518°_4
2747.005	50	2	36392.52	+0.06	$^3P_1$— 496°_2
2747.155	15	3	36390.54		
2747.826	40	2	36381.65	−0.21	$^5D_4$— 426°_3
				+0.16	$^5D_2$— 397°_2
274[illegible]					$^3G_4$— 528°_3
274[illegible]					18_3— 553°_4
274[illegible]					20_2— 573°_3
274[illegible]					$^5D_3$— 411°_4
274[illegible]					19_2— 556°_2
274[illegible]					
274[illegible]					
27[illegible]					$^3G_3$— 496°_2
27[illegible]					[illegible]826_5— 561°_4
27[illegible]					$19_{[illegible]}$— 561°_4

2746.734	40		36396.11	+0.02	153 -	51840
2747.005	50	2	36392.52	+0.06	3P1 -	49620
2747.155	15	3	36390.54			
2747.826	40	2	36381.65	-0.21	5D4 -	42630
				+0.16	5D2 -	39720
2748.312	20	15S	36375.21	-0.04	364 -	52830
2748.577	30	3	36371.71	+0.02	133 -	55340
2748.767	5		36369.20	-0.10	202 -	57330
2748.844	80	20	36368.17	+0.03	5D3 -	41140
2748.997	25	2	36366.15	+0.15	192 -	55620
2749.538	1	2	36359.01			

Figure 4. A portion of a spe. :roscopic table set from information supplied on ordinary punched cards. Note how differently the lines containing pure numerics are treated from those that contain mixtures of letters and numbers. This specialized treatment was handled by a general-purpose program called SETAB.



Equation No. 10b supplies in regard to equations Nos. 8 and 9, first

$$[\{h(\phi_{\underline{k}}\delta_1^\alpha - \phi_1\delta_k^\alpha)\}/_1$$

$$-h(\phi_{\underline{k}}\delta_\lambda^\sigma - \phi_\tau\delta_k^\sigma)L_{\sigma\tau}^\alpha]/_\alpha = 0$$

We introduce now provisionally for abbreviation tensor density

$$W_{k1}^\alpha = h(\phi_k\delta_1^\alpha - \phi_1\delta_k^\alpha)$$

According to equation No. 5

$$W_{\underline{k1}}^\alpha/1/\alpha = W_{\underline{k1}}^\alpha/\alpha/1$$

$$-(W_{\underline{k1}}^\alpha \; L_{1\alpha}^\sigma)_\sigma$$

so that the equation to be calculated can be written in the form

$$(W_{k\underline{s}/l}^{l} - W_{\underline{k}\underline{i}}^{\sigma} L_{1\tau}^{\alpha}$$

Equation No. 10b supplies in regard to equations Nos. 3 and 9, first

$$[\{h(\phi_{\underline{k}}\delta_1^\alpha - \phi_1\delta_k^\alpha)\}/_1$$

$$-h(\phi_{\underline{k}}\delta_\lambda^\sigma - \phi_\tau\delta_k^\sigma)L_{\sigma\tau}^\alpha]/_\alpha = 0$$

We introduce now provisionally for abbreviation tensor density

$$W_{k1}^\alpha = h(\phi_k\delta_1^\alpha - \phi_1\delta_k^\alpha)$$

According to equation No. 5

$$W_{\underline{k1}}^\alpha/1/\alpha = W_{\underline{k1}}^\alpha/\alpha/1$$

$$-(W_{\underline{k1}}^\alpha \; L_{1\alpha}^\sigma)_\sigma$$

Figure 5. A portion of the Einstein paper on a Unified Field Theory printed in the Herald Tribune on February 1, 1929. Today such technical text can be captured in machine readable form on a "scripting" teleprinter in the manner shown at the right. Such a single keyboarding can serve three purposes: transmission over communication circuits, storage in a computer, and input to a phototypesetting system.

begin

1 read a, b, and y from tape

2 $r=\int_0^y \exp(-a^2x^2)\cos bx\,dx$

3 print r to 8 figs, a to 8 b to 8 y to 8

Figure 6. A short problem statement written in the MIRFAC language, which serves as a complete program. Note that the problem solver need not tell the computer how to evaluate the integral as the compiler already knows how to integrate accurately.

$$A = \int_{1}^{n} \int_{1}^{\sqrt{y}} \frac{e^{-\alpha xy} \sum_{i,j=1}^{n} \left(A_i B_j \ \mathrm{SIN}^{i+j} x\right)}{\dfrac{\mathrm{LOG}_2 |y|}{\sqrt{\dfrac{3.07}{\beta + \frac{x}{\gamma}} \ \mathrm{TAN}^{-1} \dfrac{y}{x}}} + \prod_{k=1}^{n} \left\{ C_k \left[x(ky + \alpha) \right] \right\}} \, dx \, dy.$$

Figure 7. Improbable as it may seem, the above is a complete program which the Klerer-May compiler accepts and returns the value of the integral. The typing of this material is not much more difficult than typing the corresponding expression in a manuscript.

March-April 1968

Constant pressure flame calorimetry with fluorine. II. The heat of formation of oxygen difluoride, R. C. King and G. T. Armstrong, *J. Res. Nat. Bur. Stand. (U.S.),* **72A** (Phys. and Chem.), No. 2, 113-131 (Mar.-Apr. 1968).

Key words: Bond energy (O—F); flame calorimetry; flow calorimetry; fluorine; heat of formation; heat of reaction; hydrogen fluoride (aqueous); oxygen; oxygen difluoride; reaction calorimetry; water.

The heats of the following reactions were measured directly in an electrically calibrated flame calorimeter operated at one atm pressure and 303 °K.

$$OF_2(g) + 2H_2(g) + 99H_2O(1) \rightarrow 2[HF \cdot 50H_2O](1)$$

$$F_2(g) + H_2(g) + 100H_2O(1) \rightarrow 2[HF \cdot 50H_2O](1)$$

$$\tfrac{1}{2}O_2(g) + H_2(g) \rightarrow H_2O(1)$$

The reactants and products were analyzed for each of the reactions. From these heats we calculated the corresponding heats of formation, as follows:

$$OF_2(g)\Delta H^\circ_{f298.15} = +24.52 \pm 1.59 \text{ kJ mol}^{-1} (+5.86 \pm 0.38 \text{ kcal mol}^{-1})$$

$$HF \cdot 50H_2O(1)\Delta H^\circ_{f298.15} = -320.83 \pm 0.38 \text{ kJ mol}^{-1} (-76.68 \pm 0.09 \text{ kcal mol}^{-1})$$

$$H_2O(1)\Delta H^\circ_{f298.15} = -285.85 \pm 0.33 \text{ kJ mol}^{-1} (-68.32 \pm 0.08 \text{ kcal mol}^{-1})$$

Figure 8. Sample entry from NBS Spec. Pub. 305-1 keyboarded on a 44 key terminal on-line to a time-shared text editing system. See Figure 9 for the keyboarding convention used in this work.

March-April]968

Constant pressure flame calorimetry with fluorine. II. The heat of formation of oxygen difluoride, R. C. King and G. T. Armstrong, J. Res. Nat. Bur. Stand. (U.S.), 72A (Phys. and Chem.), No. 2,]]3-]3] (Mar.-Apr.]968).

Key words: Bond energy (O_F); flame calorimetry; flow calorimetry; fluorine; heat of formation; heat of reaction; hydrogen fluorine (aqueous); oxygen; oxygen difluoride; reaction calorimetry; water.

The heats of the following reactions were measured directly in an electrically calibrated flame calorimeter operated at one atm pressure and 303 °K.

ΘF2(g) + 2H2(g) + 99H2Θ(1) ¥ 2#HF] 50H2Θ@(1)

F2(g) + H2(g) +]00H2Θ(1) ¥ 2#HF] 50H2Θ@(1)

]?2Θ2(g) + H2(g) ¥ H2Θ(1)

The reactants and products were analyzed for each of the reactions. From these heats we calculated the corresponding heats of formation, as follows:

ΘF2(g)DH°f298.]5 = +24.52)].59 kj mol"] (+5.86) 0.38 kcal mol"]]

HF] 50H2Θ(1)DH°f298.]5 = _320.83) 0.38 kj mol"] (_76.68) 0.09 kcal mol"]]

H2Θ(1)DH°f298.]5 = _285.85) 0.33 kj mol"] (_68.32) 0.08 kcal mol"]]

The uncertainties indicated are the estimates of the overall experimental errors. The value of the average O_F bond energy in OF2 was calculated to be]9].29 kj mol"] (45.72 kcal mol"]].

Figure 9. Sample entry as keyboarded on an IBM 2741 into an on-line text editing service. Note the use of overstrikes to obtain grid changes, subscripts, and superscripts. This system is in daily use for the production of NBS Spec. Pub. 305 and its supplements and numerous conference proceedings. The next figure shows the same material keyboarded on a scripting typewriter with 126 printable characters.

March-April 1968

(1)

Constant pressure flame calorimetry↓with fluorine. II. The heat of formation of oxygen difluoride,nR. C. King and G. T. Armstrong, iJ. Res. Nat. Bur. Stand. (U.S.),b72An(Phys. and Chem.), No.↑2, 113-131 (Mar.-Apr. 1968).

(2) (3)

Key words: Bond energy (O-F); flame calorimetry; flow calorimetry; fluorine; heat of formation; heat of reaction; hydrogen fluoride (aqueous); oxygen; oxygen difluoride; reaction calorimetry; water.

The heats of the following reactions were measured directly in an electrically calibrated flame calorimeter operated at one atm pressure and 303^K.

$OF_2(g) + 2H_2(g) + 99H_2O(l)$ (sYn) $2[HF$ (g^n) $50H_2O](l)$

$F_2(g) + H_2(g) + 100H_2O(l)$ (sYn) $2[HF$ (g^n) $50H_2O](l)$

$O_2(g) + H_2(g)$ (sYn) $H_2O(l)$

The reactants and products were analyed for each of the reactions. From these heats we calculated the corresponding heats of formation, as follows:

$OF_2(g)\Delta H^{\wedge}_{f298.15}$ = +24.52 (g:n) 1.59 KJ mol^{-1}(+5.86 (g:n) 0.38 kcal mol^{-1})

HF_1 (g^n) $50H_2O(l)\Delta H^{\wedge}_{f298.15}$ = -320.83 (g:n) 0.38 KJ mol^{-1}(-76.68 (g:n) 0.09 kcal mol^{-1})

$H_2O(l)\Delta H^{\wedge}_{f298.15}$ = -285.83 (g:n) 0.33 KJ mol^{-1}(-68.32 (g:n) 0.08 kcal mol^{-1})

The uncertainties indicated are the estimates of the overall experimental errors. The value of the average O-F bond energy in OF_2 was calculated to be 191.29 kj mol^{-1}(45.72 kcal mol^{-1}).

Figure 10. Sample entry as keyboarded on a Model 37 Teletype. On this terminal Greek characters as well as subscripts and superscripts appear in natural form for easy proofreading. The numbered and circled symbols n, i, b, s, and g are keyboarded in red. They signal changes respectively to the following grids: Roman (normal), Italics, Bold, Symbol, and Greek. The indentations in the copy are achieved with multiple tabs which control the systematic type face changes and other formal characteristics of the typeset copy shown in Figure 8.

COMPUTER USAGE IN A LARGE DATA CENTER

James I. Vette
National Space Science Data Center
Goddard Space Flight Center
Greenbelt, Maryland

1. Introduction

In order to present the various ways in which computers are used at the National Space Science Data Center (NSSDC), it will be necessary to give a brief description of the total activity of the Center. (More detailed information can be obtained from other documents.[1-6]). In that way, one can see what we mean by a large data center. I'm sure that there are some in the audience that are associated with larger facilities but compared to the general IAC's identified in the COSATI Directory, NSSDC represents a large data center.

NSSDC is responsible for the acquisition, organization, storage, retrieval, announcement and dissemination of the scientific data obtained primarily by satellites. To a lesser extent, we are involved with the results from experiments carried out on sounding rockets, probes, high-altitude aircraft and balloons. The size of the data base involved is given in Table 1. It can be seen that data are stored on magnetic tapes, punched cards, microfilm, photographic films, and prints, as well as hard copy.

One of the main functions of NSSDC is to provide data and information to qualified users and to refer others to appropriate sources for the services they seek. Our user community are generally scientists, engineers, college level teachers, and students who wish to use the data in some scientific investigation or for some instructional purposes. The casual seeker of knowledge about the space program and its scientific results is referred to the appropriate sources for public information. A measure of the activity of the reproduction of data and data products are given in Table 2. We will only be concerned in this talk with that reproduction where computers are utilized.

In addition to serving as a data and information center, NSSDC also performs as an IAC in analyzing and synthesizing some of the vast quantities of data in its archive so that new and useful forms of the data are available. In this analysis work the computer is used extensively.

2. Availability of Computer Systems

Before discussing in some detail the specific uses of computers in performing the functions of NSSDC, the various computers readily available to the Center will be listed. There are ten large general purpose computers and numerous smaller ones at Goddard Space Flight Center (GSFC)

VOLUME OF DATA AT NSSDC (12/31/70)

FORM	CUMULATIVE
Sheets and Bound Volumes, Sheets	166,724
Digital Magnetic Tapes, 1/2 inch x 2400 feet	11,328
Microfilm, 100-foot Rolls	16,177
Photographic Films:	
9-1/2-inch width, linear feet	18,000
70-mm width, linear feet	310,482
16-mm width, linear feet	8,840
35-mm width, linear feet	759,769
4 x 5 inch, each	9,186
8 x 10 inch, each	2,410
16 x 20 inch, each	93
20 x 24 inch, each	8,005
Photographic Prints:	
9-1/2-inch width, linear feet	9,000
70-mm width, linear feet	22,000
8 x 10 inch	7,035
11 x 14 inch	500
16 x 20 inch	93
20 x 24 inch	3,200
Punched Cards	37,700

1970 NSSDC REQUEST OUTPUT

MEDIUM	UNIT	NUMBER OF REQUESTS COMPLETED	TOTAL AMOUNT OF OUTPUT
Digital Magnetic Tapes	2400′ Reels	123	655
Punched Cards	Cards	65	77936
Computer Printout	Sheets	223	64700
Microfilm	Reels	202	2520
Hard Copy	Pages	356	52276
Photo LUNAR ORBITER		181	
Positives or Negatives	Each/Feet		4352/2584
Black & White or Color Prints	Each		5778
35 mm x 100 feet	Reels		114
SURVEYOR		16	
Positives or Negatives	Each		75
Black & White or Color Prints	Each		81
GEMINI		47	
Positives or Negatives	Each		21
Black & White or Color Prints	Each		97
NIMBUS		178	
Positives or Negatives	Each/Feet		729/7445
Black & White or Color Prints	Each		2078
MARINERS 6 and 7		39	
Positives or Negatives	Each/Feet		153/4050
Black & White or Color Prints	Each		5533
35 mm x 100 feet	Reels		5
APOLLO		239	
Positives or Negatives	Each/Feet		496/22278
Black & White or Color Prints	Each		6487
35 mm x 100 feet	Reels		58

COMPUTER PRODUCTION-1970

		Computer Time (Hours)	Man-Years
A. Processing	IBM 360/75	12	4
	IBM 7094	519	
B. Requests	IBM 360/75	30	3
	IBM 7094	379	
C. Information System	IBM 7094	645	3
D. Analysis	IBM 360/91	5	3
	IBM 360/75	18	
	IBM 7094	92	

PROGRAM DEVELOPMENT-1970

		Computer Time (Hours)	Man-Years
A. Processing	IBM 360/75	2	3 1/2
	IBM 7094	101	
B. Requests	IBM 360/91	1	1 1/2
	IBM 360/75	3	
	IBM 7094	42	
C. Information System	IBM 360/75	17	8
	IBM 7094	493	
D. Analysis	IBM 360/91	2	2
	IBM 360/75	3	
	IBM 7094	13	

where NSSDC is located. Four of the large computers are used by the Data Center for various tasks. An IBM 7094 MOD II running in the conventional batch processing mode is located in the Data Center building and is heavily utilized. An IBM 360/91 and a 360/75 both operating in a multiprogramming, variable task mode (MVT) are also used to a lesser extent. There are terminals at the Data Center which allow for remote job entry to these two computers. In addition an interactive system employing APL (A Programming Language) is available through a typewriter terminal to an IBM 360/95, which is the largest computer at GSFC. In addition plots and microfilm outputs are available through an S-C 4020 and S-D 4060. There are Cathode Ray Tube (CRT) Terminals with light pens (IBM 2250's) available for the 360 computers on a limited basis for special development work.

3.0 Computer Usage

We will discuss the computer usage in four functional categories: (a) processing of data into the archive, (b) responding to requests for machine sensible data, (c) storing and retrieving information from the information system, and (d) analyzing data. For the calendar year 1970 we show in Table 3 computer time used both in computer program development and in the production running of these programs on the various computers. In addition approximately 1000 terminal hours were logged on APL for analysis of data. The approximate amount of effort in man-years is also given in Table 3 for each category. For program development the effort is for computer programming and for production work this represents tape handling, job submission, setting up the various computer runs, and handling the resultant outputs. This latter effort does not include the operation or maintenance of the computer facility since this is not the responsibility of the Data Center.

We will now discuss in more detail the type of work accomplished through the use of the computer in these four categories.

3.1 Processing

The data received in machine sensible form are nearly always on digital magnetic tape although occasionally punched cards are used. Analog tapes and punched paper tape are practically never used as storage media for the type of data NSSDC archives; consequently the necessary equipment to handle these is not available at NSSDC. Although the raw data from the satellites are collected by various tracking networks operated by NASA, USAF, foreign countries, and ESRO, the reduced and analyzed data which NSSDC uses are collected directly from the principal investigators in charge of the individual experiments and responsible for the first or prime analysis of the data. Consequently the data has been processed by a wide variety of digital computers and the magnetic tapes we receive are coded in forms appropriate for these computers. Unfortunately there is a great degree of incompability between the various computers in terms of the character size (6 or 8 bits), parity (odd or even), word size (16,24,36,48,60 bits), as well as the specific meaning of a string of bits in terms of a number or a letter. For those not so familiar with computer jargon it is really analogous to different cultural languages (not to be confused with programming languages such as COBOL, FORTRAN, etc.).

In order to handle this problem it is necessary to translate from one computer language to another. Fortunately the problem is really one of transliteration since there is no syntax involved. The processing of the incoming data is used to accomplish the following functions: (a) verify that all the data are readable from the tape, (b) verify that the format of the data has been correctly specified and documented by the sender, (c) make an index or catalog of each tape and (d) produce a new self documented tape in the "language" of our local computer which includes the index and format information. A generalized Data Base Management System has been under development to perform these tasks in a straight forward manner. The heart of this system is really a problem oriented language which allows our people to specify easily how much data is to be retained from the original tape and what the organization of the data will be on the new standard tape. The system also includes a set of programs which can operate on the standard tape to provide various checks, produce specified plots or printouts, do statistical analyses and produce the necessary index for each tape. In addition the system contains the necessary programs to produce an output tape from the standard tape in any of the common computer "languages" so that the user who has requested specific data will have no trouble in entering this directly into his own computer without having to perform the transliteration process.

3.2 Requests

The processing of the machine sensible data described in the preceding section has prepared this data so that it can readily be retrieved in part or whole and outputted in a variety of forms for the greatest convenience and ease of use by the requester. These outputs include computer printouts of the data in various tabular formats, plots of selected data, and a magnetic tape that is compatible with the requester's computer. In addition there are some programs which convert the data on tape to specialized outputs which are extremely useful. An example of one of these outputs is shown in Figure 1. This is known as a grid print Mercator map which provides a black body temperature as a function of position from the earth viewing infrared radiometers on board the Nimbus Meteorological Research Satellites. These maps can also be produced in stereographic projection about any point on the earth's surface in scales of 1- 10, 20, or 30 Million. This particular map shows Typhoon Marie, a storm in 1966 which was observed by the Nimbus II Satellite. In addition a data population map can be produced which gives the number of data points used to determine the average temperature of the grid print map. Occasionally special computer programs are written to select data on the basis of criteria specified by the user or to perform certain averaging of the data. However, most requests are for data covering specific time intervals. One program was written to determine when certain satellites would intersect specified field lines of the earth's magnetic field.

3.3 Information System

In order to keep track of the numerous supporting information that is necessary to supply to users along with the scientific data itself a number of computerized files are used. These files constitute the major part of the total information system of NSSDC. A whole range of reports can be printed out periodically from these files. In addition specialized inquiries can be made with the coding of simple computer programs.

The Automated Internal Management File (AIM) is used to store information about the satellites, experiments and data sets. There are about 50 different items connected with each of the three types of entries. As of December 31, 1970 this file accounted for 1380 satellites, 1824 experiments and 975 data sets. This file is used to produce the catalog of the Data Center's holdings as well as brief descriptions of various satellites and their experiments which appear in published compilations from time to time. In addition management information about the file and the status of acquiring and processing the data are available.

A second file is the Technical Reference File (TRF) in which information about all the documents (published and unpublished) concerning the satellites, experiments, rocket and balloon flights and appropriate aircraft flights is kept. Besides the author, title, and bibliographic notation, an internal classification of the document and location is produced. Keywords are assigned by our staff of space science professionals to relate the article to the appropriate satellite, experiment, specific disciplines, geophysical events, and other items of interest to our users. This is an extension, for a small subset of documents, to the extensive indexing, keywording, abstracting, storing and retrieving of the aerospace literature performed by NASA through its Scientific and Technical Information Facility (STIF) and through the AIAA. The TRF is used to produce various types of bibliographies as well as provide management information about the scientific output of various investigators and satellite missions.

There are several other computerized files which will only be mentioned briefly here. The bookkeeping connected with our request business is maintained in a computerized file called Request Status and History (RASH). In addition there is a ROCKET file for keeping track of all the rockets launched throughout the world carrying space science experiments. A distribution file is used to maintain the names and addresses of people in various categories and one output of this file is printed gum labels for mailing purposes. A Data Set Inventory System that is used to keep track of all our data products, their location and status is in the process of being completed. In addition an Extraterrestrial Photographic Information Center File is used to supply supporting information about our photographs including descriptors about some of the subject matter contained in the pictures.

None of the information system files described is extremely large. The total number of characters in some of these are given in Figure 2 and the number of transactions per month for AIM is given in Figure 3. The importance of the information system to the operation of the Data Center can be judged from Table 1 where one sees that more computer time and program development have been used than any other area. However, during the present year the processing category will require the maximum computer usage as we begin to process the large quantities of data now coming in.

We are in the process of putting a portion of our information files in an on-line terminal operated system that is commerically available. We hope to determine from this experiment our full requirements for an on-line system and to measure the change in efficiency of our operation using this service.

3.4 Analysis

Most of the data collected by NSSDC cannot be understood directly in terms of simple physical processes since there are many competing physical phenomena occurring during the measurements. The analyses conducted by the Data Center emphasize the selection of data from a large number of experiments in order to produce tractable models of the various environmental conditions that exist in space. In many cases these models are strictly empirical; in other cases theoretical ideas provide parameters which can be determined from the data. In one sense, such models represent data compression and a theory which explains fairly completely a given class of observations represents the maximum compression. The results of such syntheses are data products generally useful to a broader class of users than those capable of working with the basic observational data.

Since this analysis work involves lengthy computations as well as the handling and display of large amounts of data, computers are used extensively in this work. Optimization of instrument parameters, transformation of coordinate systems, transformation of physical quantities, non-linear regression analysis, correlation of selected physical quantities, time series analysis, orbit computations, and graphical displays are the main functions accomplished by computers in carrying out these tasks.

Several specific examples will be given. Energetic protons are trapped or contained in the earth's magnetic field. In order to obtain a fairly complete mapping of these particles with energies above 50 Million electron volts (MeV), 21 different experiments were studied. Some synthesized results are shown in Figure 4. The "best value" particle flux contours are given in a magnetic coordinate system which takes the real shape of the earth's magnetic field into account. More details on the trapped radiation model environment are given in references 7-12.

Another example is shown in Figure 5 where the orbits of several different satellites, which were operating at the same time, are displayed in a coordinate system (solar-ecliptic) in which certain experimentally identifiable boundaries, the bow shock and the magnetopause, remain stationary. Although some of this work is reflected by the computer usage shown in Table 1, much of it is accomplished using the interactive APL terminal. One of our scientists has used this system to determine on-line the optimum values to assign to any given broad band X-ray detector[13]. The parameters specifying the detector are inputted, the forms of the X-ray spectrum can be chosen, and the answer is returned immediately. Different parameters and spectral forms can be used to cover all possible situations.

In addition some efforts are underway to improve the general data manipulation and display problems we have through the use of problem oriented languages with the existing computer tools available to us at the present time. As one can see the computer plays a vital role in our Data Center. We are looking forward to the time when new high density storage devices and interactive time shared computers with graphic displays can be used to give us new capability in the storage, retrieval, display, manipulation, and analysis of our growing data base.

References

1. Vette, J. I., "The Operation of the National Space Science Data Center," NSSDC 67-41, Oct. 1967.

2. Karlow, N. and J. I. Vette, "Flow and Use of Information at the National Space Science Data Center," NSSDC 69-02, Jan. 1969.

3. Karlow, N. and J. I. Vette, "Data Information System at the National Space Science Data Center," *Proc. Amer. Soc. Info. Sci., 6, 231, 1969.*

4. Vette, J. I. and N. Karlow, "Data Management at the National Space Science Data Center," *Jour. Spacecraft & Rockets, 7, 1234, 1970.*

5. Vette, J. I., "The Availability of Satellite Data for Scientific Analysis," Trans. Amer. Geophys. U. (EOS), *51*, 667, 1970

6. Fava, J. A., C. K. Michlovitz, and N. Karlow, "Responding to Request at the National Space Science Data Center," NASA GSFC-X-601-70-162, May 1970.

7. Vette, J. I., "Models of the Trapped Radiation Environment - Vol. I: Inner Zone Protons and Electrons," NASA SP3024, 1966.

8. Vette, J. I., A. B. Lucero, and J. A. Wright, "Models of the Trapped Radiation Environment - Vol. II: Inner and Outer Zone Electrons," NASA SP-3024, 1966.

9. Vette, J. I. and A. B. Lucero, "Models of the Trapped Radiation Environment - Vol. III: Electrons at Synchronous Altitudes," NASA SP-3024, 1967.

10. King, J. H., "Models of the Trapped Radiation Environment - Vol. IV: Low Energy Protons," NASA SP-3024, 1967.

11. Lavine, J. P. and J. I. Vette, "Models of the Trapped Radiation Environment - Vol. V: Inner Belt Protons," NASA SP-3024, 1969.

12. Lavine, J. P. and J. I. Vette, "Models of the Trapped Radiation Environment - Vol. VI: High Energy Protons," NASA SP-3024, 1970.

13. Wende, C. D., "The Normalization of Solar X-Ray Data from Many Experiments," NASA GSFC-X-601-71-166.

Figure Captions

Figure 1 Grid print map of Typhoon Marie. The black body temperature derived from the High Resolution Infrared Radiometer on the Nimbus II satellite is printed in degrees Kelvin as a function of geographic position using a Mercator projection.

Figure 2 The size of various information files at NSSDC.

Figure 3 The number of monthly transactions for the Automated Internal Management [AIM] File. A transaction is either a new entry, a correction, or a deletion to the file.

Figure 4 A B-L flux map for protons. The contours are for omnidirectional flux; this flux gives the number of protons above 50 Mev that would enter a sphere of cross sectional area of one squarecentimeter in one second. The B coordinate is the intensity of the earth's magnetic field and the L parameter is a quantity that labels a given field line. The L parameter can be interpreted as the distance from the center of the earth that the field line crosses the geomagnetic equator. This B,L coordinate system is used extensively in the study of the earth's radiation belts.

Figure 5 Orbits of satellites Vela 6A, Vela 5A, and HEOS-A1. The orbital paths of each satellite during the period April 8-12, 1970 are projected onto the X-Y plane of the solar-ecliptic coordinate system. The distance units are earth-radii. In that system the earth is at the origin, the X axis points toward the sun, and the X-Y plane is the ecliptic plane. The magneto sheath is the region between the bow shack, shown by the short-dash line, and magnetopause, shown by the long-dash line. These two boundaries are caused by the interaction of low energy protons from the sun [the solar wind] and the earth's magnetic field.

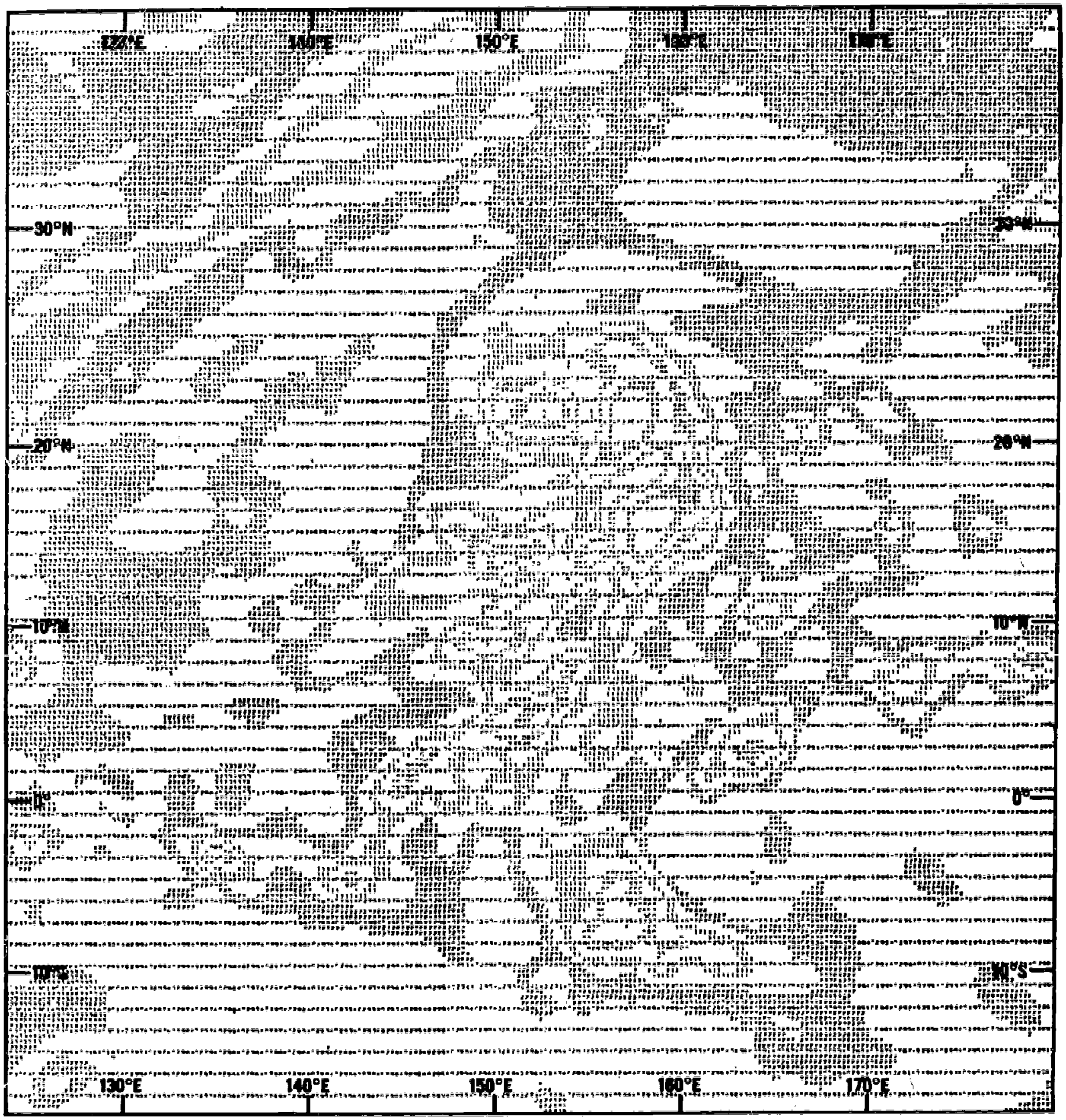


Figure 1 Grid print map of Typhoon Marie. The black body temperature derived from the High Resolution Infrared Radiometer on the Nimbus II satellite is printed in degrees Kelvin as a function of geographic position using a Mercator projection.

NSSDC AUTOMATED FILES—CHARACTERS PER SYSTEM

AS OF DECEMBER 31, 1970

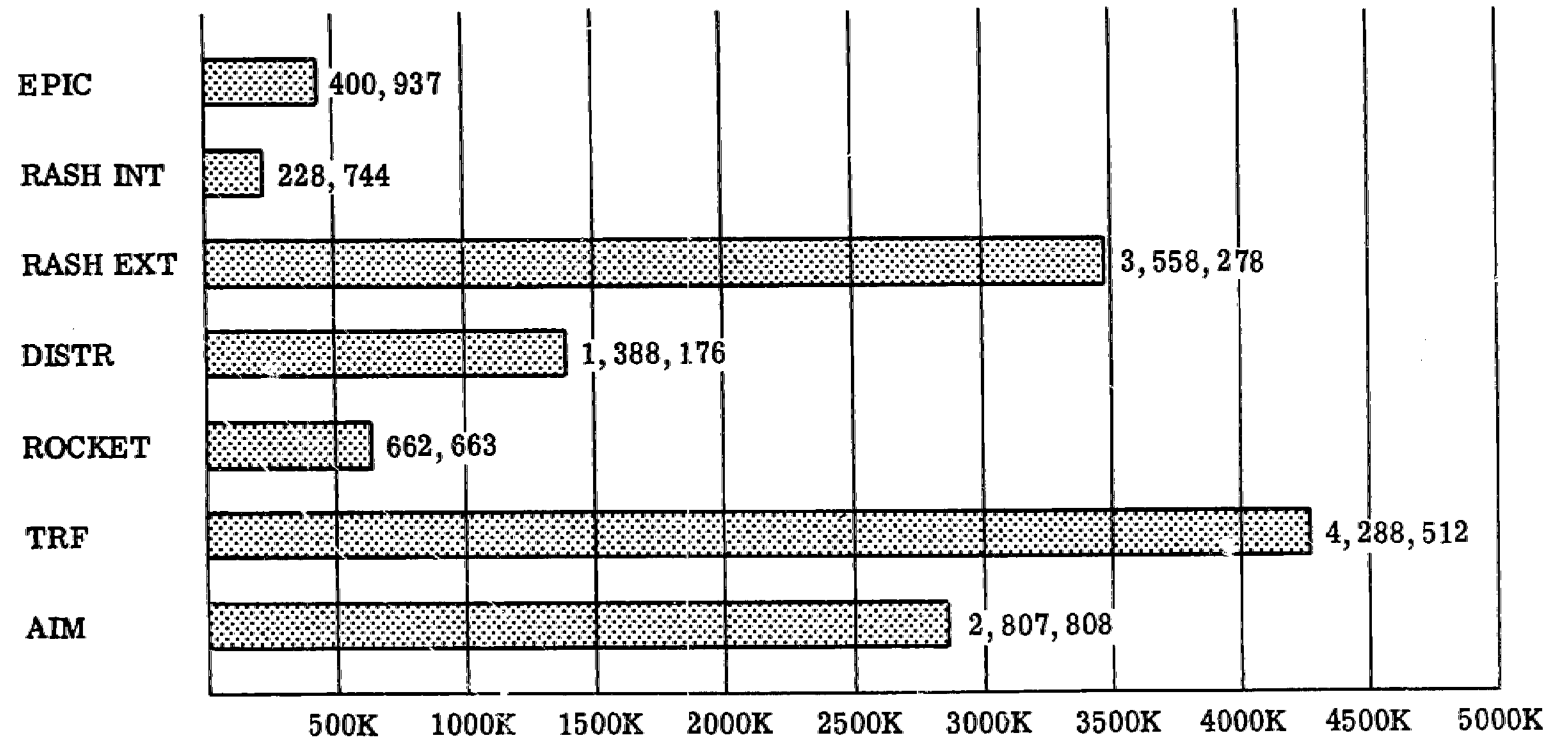


Figure 2 The size of various information files at NSSDC.

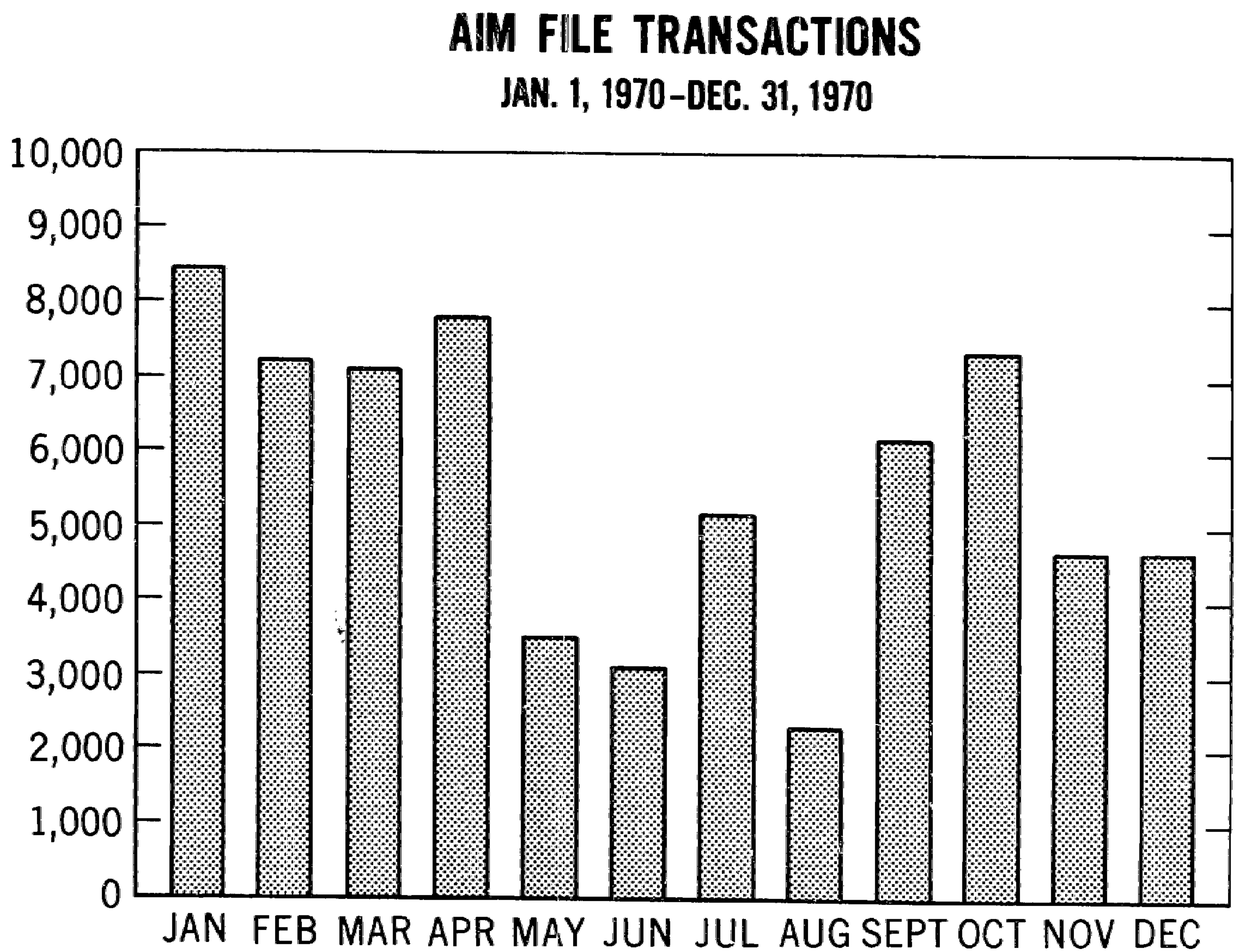


Figure 3 The number of monthly transactions for the Automated Internal Management (AIM) File. A transaction is either a new entry, a correction, or a deletion to the file.

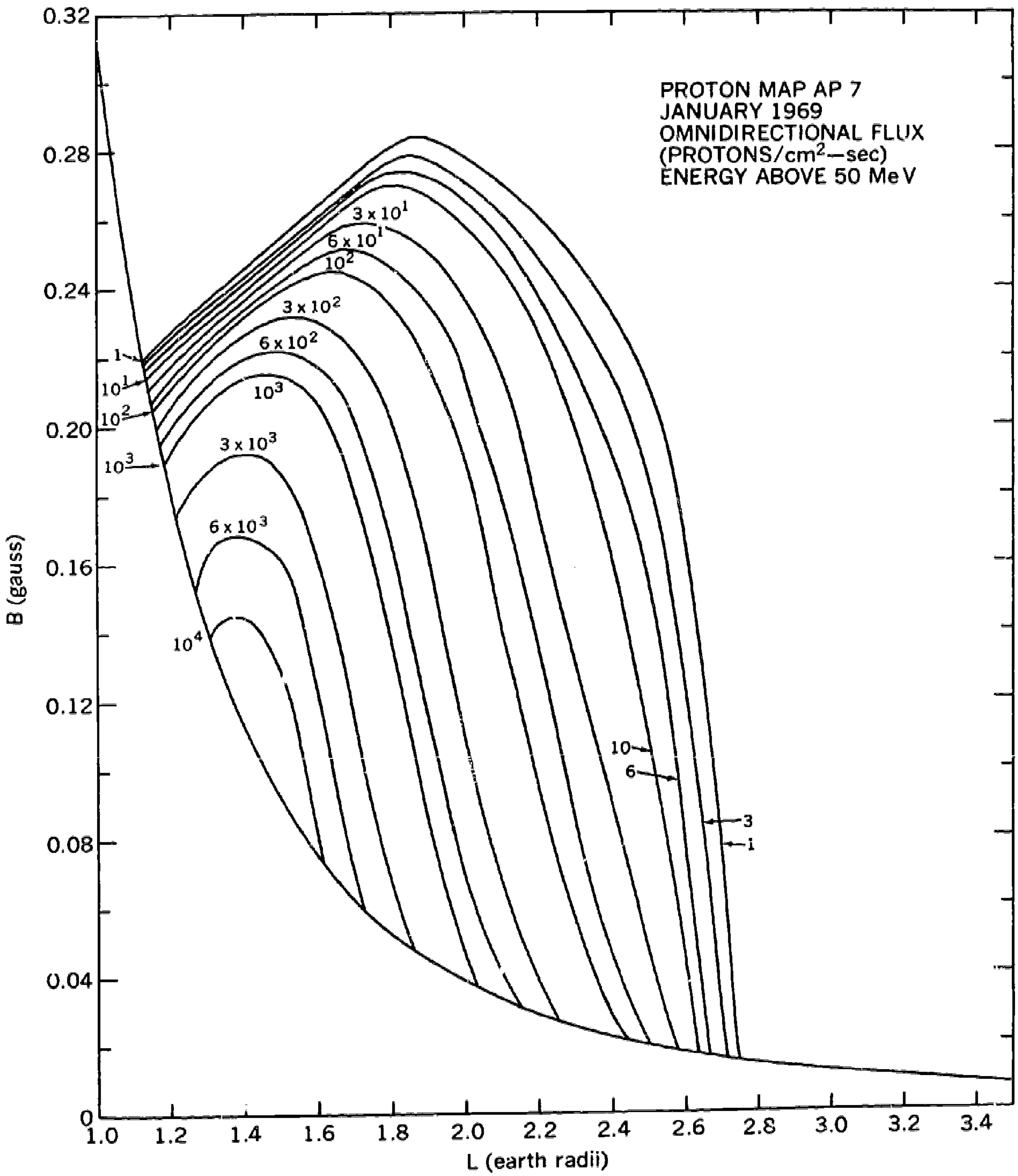


Figure 4 A B-L flux map for protons. The contours are for omnidirectional flux; this flux gives the number of protons above 50 Mev that would enter a sphere of cross sectional area of one squarecentimeter in one second. The B coordinate is the intensity of the earth's magnetic field and the L parameter is a quantity that labels a given field line. The L parameter can be interpreted as the distance from the center of the earth that the field line crosses the geomagnetic equator. This B,L coordinate system is used extensively in the study of the earth's radiation belts.

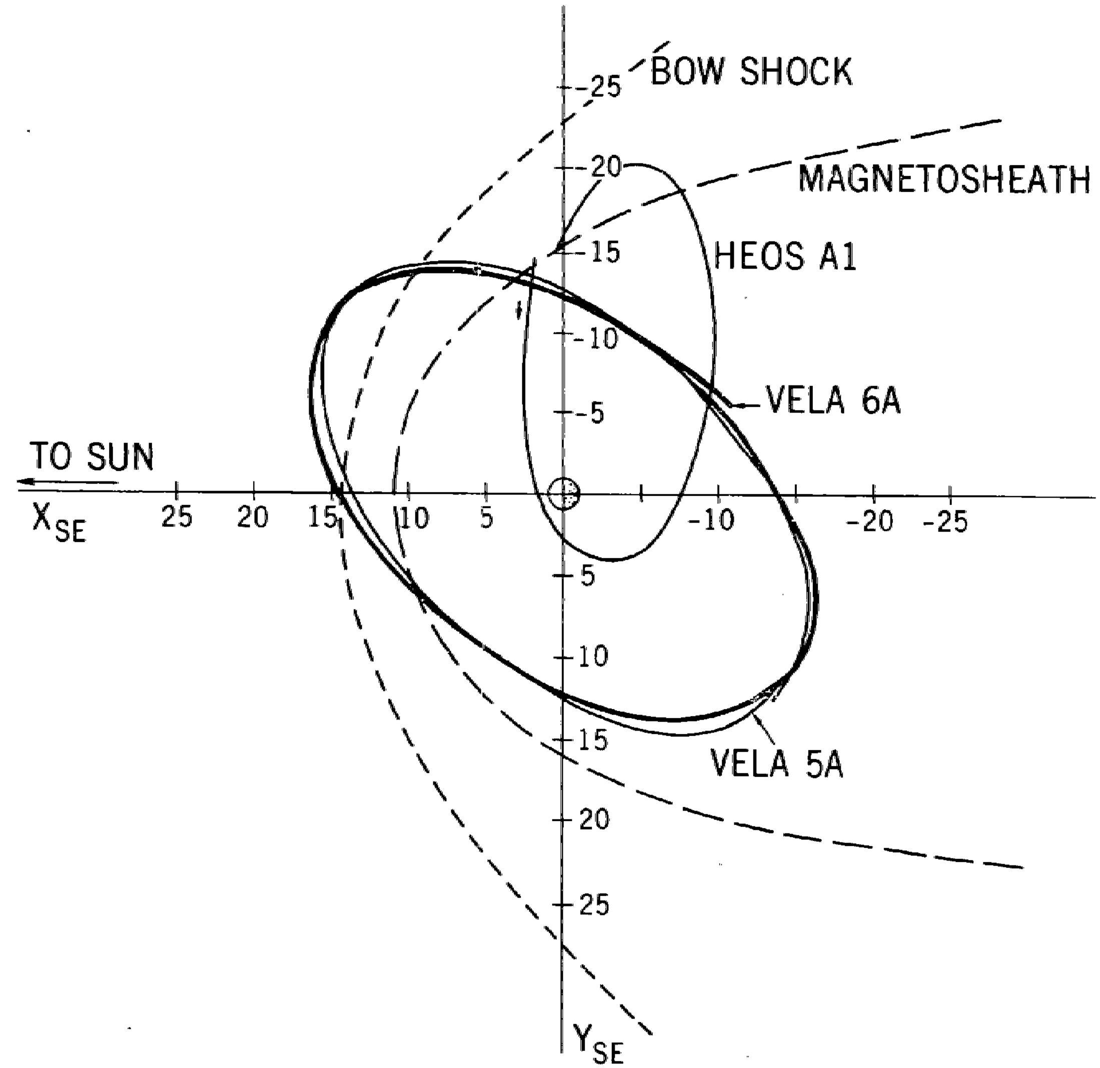


Figure 5 Orbits of satellites Vela 6A, Vela 5A, and HEOS-A1. The orbital paths of each satellite during the period April 8-12, 1970 are projected onto the X-Y plane of the solar-ecliptic coordinate system. The distance units are earth-radii. In that system the earth is at the origin, the X axis points toward the sun, and the X-Y plane is the ecliptic plane. The magneto sheath is the region between the bow shack, shown by the short-dash line, and magnetopause, shown by the long-dash line. These two boundaries are caused by the interaction of low energy protons from the sun (the solar wind) and the earth's magnetic field.

THE NATIONAL STAKE IN BETTER TECHNICAL INFORMATION

James H. Wakelin, Jr.,
Assistant Secretary of Commerce for Science and Technology

It has been eight years since publication of the bible of the information analysis profession, the Weinberg Report, entitled *Science, Government, and Information.*

Since then, we have made progress in all three elements of that title—in science, in Government, and in information—but our progress has been halting.

First, science and technology are no longer the glamor fields they were then. In fact, I am very much troubled by the attitudes of anti-science and anti-technology we see so frequently.

Second, Government is admittedly poorly organized to cope with society's demands upon it. This is particularly true of Government's agencies concerned with science and technology. That is why President Nixon recently proposed to the Congress to reorganize the Executive Branch around the great contemporary purposes of government. His plan would bring government closer to the people, simplify program coordination and conflict resolution, and permit clearer assignment of authority and accountability.

Third, information multiplies and accumulates too rapidly for our over-burdened scientists and engineers to process it—or for our antiquated governmental institutions to use it efficiently. Data-processing technology and data-producing organizations have outpaced the supply of human beings qualified and organized to handle it.

Despite our massive social problems—of education, environment, cities, transportation, housing, and the like—I remain optimistic that we can solve them, using information to do so. I believe that there is a latent respect among the American people for science and technology. Society can achieve its loftiest ambitions, but it requires these tools to do so. Anti-science and anti-technology attitudes can be made to yield to persuasion, because, in the end, science and technology are essential to the achievement of society's goals. We need a massive infusion of confidence, which I believe can come from enlightened young people. Perhaps some of the commencement speakers who are just around the corner will tell us how. Consider the remarks which follow to be my commencement speech to you.

The Changing Role of the Information Analysis Center

The publication of the Weinberg Report in 1963 was a milestone. While laboratories performing some of the information analysis functions had existed for a century or more, true IACs were proliferating at that time.

Information Analysis Centers—IACs—not only have proliferated in the last decade, they have subtly changed. Alvin Weinberg emphasized transfer of information as an inseparable part of research and development, in these words:

> All those concerned with research and development—individual scientists and engineers, industrial and academic research establishments, technical societies, government agencies—must accept responsibility for the transfer of information in the same degree and spirit that they accept responsibility for research and development itself.

Transfer of information is, of course , an integral part of the R&D process, yet there is more to the information analysis function than that. In forwarding the invitation from Ed Brady to make these remarks tonight, Lew Branscomb put a different and significant focus on your profession when he wrote:

> I consider that the major "information problem" the nation faces is not information manipulation or transmission but the quality of available information and its interpretation for appropriate use.

In those two quotes we see symbolized the changes of the past eight years: You long ago went beyond the stage which might be called *the transfer of information* to a new level of concern over *the quality of information*. All scientists and engineers are involved in the transfer of information. But the IACs are both "meccas" and "mechanisms." They are the meccas for comprehensive, quality information and its interpretation. And they are the best mechanisms we have for feedback, completing the loop to assure appropriate information appropriately used.

Relevant information has always been a *sine qua non* for any expert. But judging the relevance of information other than that which he generates is always difficult. The researcher can place confidence in the IAC for several reasons. The principal reason is that the operation of IACs usually takes place in a research atmosphere. This permits top-level experts to take part in analysis activities, while continuing to do research. It provides means for checking crucial values, for confirming experimental techniques, or for assessing purity of materials. It brings together experts in related fields.

I can see only one possible disadvantage in this research atmosphere. If it is too "ivory tower," that makes it easy for IAC staff members to consider that the specialists in their field are the primary users of their information. That would be an error, or at least an over-simplification. It is the *non-expert*, or the expert from another field, who has the greatest need for IAC output and services.

The greatest benefit leverage-factor for IAC services is found when the services are used in direct application to practical problem-solving. It is no derogation of basic research to state that the benefits achieved for society are realized much faster through problem-solving than through basic research.

The role of the IACs is constantly changing. The IACs must plan to provide services and output which can be used by engineers and applied scientists as well as by basic research workers. You must provide more and more outputs which will directly contribute to solving major problems relevant to national needs. Some IACs are best operated by Government, some by universities, some by research institutes, and some by professional societies. I was happy to note just recently that the American Nuclear Society is establishing an Information Center on Nuclear Standards.

Six Major Needs in Information Analysis

In his speech to this forum three and a half years ago, Donald Hornig, now president of Brown University, spoke of "the responsibility of the information analysis center to try to ensure that significant information from all sources is incorporated into the body of related information stored in the center." By the word *significant* he implied that the IAC staff must use *judgment* to decide what information is worth storing for future use and *ingenuity* to try to locate *all* information that is worth accumulating.

But this responsibility must not be passed entirely to the IACs. It must be shared by active workers in each field. In compiling my list of six major needs in information analysis, therefore, I borrow the first from Don Hornig:

1. *We need to involve a much larger proportion of the total technical community in information analysis activities*—as users, as participants, or just as supporters, for we are all potential users and participants. In each of those roles, we should promote the concept of the IACs through word-of-mouth advertising. Of course, you are primarily interested in your professional field and the constituency your center serves. But why not take every opportunity to inform both your professional colleagues and your customers (so to speak) that yours is an information analysis center. They might call on you with unrelated problems if they knew that an IAC is an organization with a unique capability for acquiring, selecting, storing, retrieving, evaluating, analyzing, and synthesizing a body of information in any clearly defined specialized field, perhaps in theirs.

2. *We need some new IACs.* I cannot believe that the present roster of Federally supported and other IACs covers all the areas of science, technology, and scholarly interests in which mechanisms of this sort could contribute to solving national problems. Lew Branscomb has suggested that NBS, which now operates several IACs in the Standard Reference Data program, shall probably establish others. These might help fulfill the Bureau's responsibilities in fire research, environmental technology, building technology, and other areas. I can imagine unfilled needs in information for policy analysis at the highest levels of Government. If and when new IACs are established throughout the nation, in whatever institution, you who constitute the reservoir of knowledge on how to set up and run them should offer your full assistance to the newcomers. This is something I know you will do.

3. *We need a strengthened National Technical Information Service to support IACs and to fill gaps including functions which are not properly those of IACs.* NTIS, as many of you know, was established late last year to bring together many of the technical information functions of the Department of Commerce. It publishes Federal publications and data files and makes them available to the business, scientific, and technical communities. This is different from, but a supplement to, information analysis, and I do not see NTIS taking over any of the functions of your agencies or your centers. To use my earlier expression, I see it as a "mechanism," not a "mecca." I believe that NTIS can support all of you by providing certain services and products far more efficiently and economically than you could. Tomorrow afternoon Mr. Harry Pebly of Plastec will discuss such a proposition.

Two other useful services of NTIS are the subscription and the standing order services. For example, for $10 per year you can subscribe to "Aerospace Medicine and Biology," a continuing bibliography of studies on the biological, physiological, and environmental effects of space flights—a joint publication of NASA, the Library of Congress, and the American Institute of Aeronautics and Astronautics. For $22 per annum you can subscribe to "Air Pollution Abstracts," published monthly by NTIS for the Air Pollution Control Office of the Environmental Protection Agency. And for the same price you can subscribe to the semi-monthly "Selected Water Resources Abstracts." There are Asian serials, Communist China serials, Eastern European serials, USSR serials, and others covering translations of technical documents from throughout the world. You can receive, free, pamphlets about these and other NTIS services by visiting the information center in the lobby of the Commerce Building or by writing the National Technical Information Service, Springfield, Virginia 22151.

NTIS and the IACs can be mutually supportive, and it is a challenge to both to develop mechanisms of support. I know that Mr. Knox and Dr. Brady are studying this matter very carefully.

4. *We need greater appreciation at the level of assistant secretary—not just in the Department of Commerce, but in all departments—of the role and value of IACs.* Every department has an Assistant Secretary for Research and Development, by that title or something close to it. He is a key constituent of yours who should be familiar with the range of services which IACs supply to his department, or which they should be supplying. Yet I dare say that, with the possible exceptions of Defense and AEC, not one of your IACs has ever been visited by an Assistant Secretary of a supporting Federal department. If you have, then the thanks go to Andy Aines and COSATI for the high-level support they have provided over the past several years. Support should work two ways, so I would make my next recommendation that:

5. *We need, at the IAC level, to support COSATI,* for much of what has been accomplished by IACs we owe to the leadership and coordination it provides to the field. I was Assistant Secretary of the Navy for R&D when COSATI was formed, so I have enjoyed a close view of it since its inception and I can't praise Andy and the agency representatives on COSATI highly enough.

6. *We need strengthened international cooperation,* for all of the areas in which you operate are international concerns of science, technology, or scholarship. I know that many of you already work closely with your counterparts in other nations, and I encourage you to intensify these efforts.

Three Problem Areas

In approaching the end of my remarks, I would like to discuss briefly three problem areas with which I am concerned in my new responsibilities in the Department of Commerce. Yet these are national problems, not my Department's alone. Regarding each, let me ask you, "What could you contribute to the solution of this problem?"

The first is:

International voluntary standardization and certification On April 28 the Department of Commerce sent to the Congress an Administration bill designed to promote exports through strengthened international voluntary standardization and certification activities. This bill would assign to the Secretary of Commerce the principal Federal responsibility for assuring that U. S. interests are adequately represented in this field. It also would authorize the Secretary to enter into grants or contracts with nonprofit organizations. By helping to write new internationally agreed-upon standards, we think that the U. S. will make sure that such standards reflect U. S. engineering practice. The legislation also will enable the Government to cooperate with U. S. industry in reaching international agreements. And it will create an official U. S. link in the international standards making process. I think this is an area in which an IAC, probably located at, and operated by, the National Bureau of Standards, will be essential.

The second problem is:

Environment. I am told by ecologists that the life sciences boast, throughout the world, 52,000 journals. These are primary publications, which publish at least one, but in many cases two or more, articles on ecology per issue. To a great extent, this knowledge is being addressed to other specialists within the same discipline, and to the peer groups within those disciplines. It seldom reaches across disciplinary boundaries, and even less frequently across national boundaries. There is a time lag between the acquisition of new knowledge in ecology and its practical application to problems of environmental management. As we accelerate national and international programs of environmental management, this lag will impede the development of improved ways of anticipating, assessing, and solving the problems of environmental deterioration. I find this a second area in which one or more IACs will be essential. The nature and location I leave to your imagination, and to your future planning.

Coastal Zone Management. As some of you may know, I have spent the past year serving the Honorable Russell W. Peterson, Governor of Delaware, as Chairman of the Governor's Task Force on Marine and Coastal Affairs. Seven distinguished citizens of Delaware constitute this task force. In mid-February we presented to the Governor and the Legislature a preliminary report. During the next several months we will complete the final report, with recommendations on the major resources of the state including water management, fisheries, and wildlife; recreation including parks, boating, and sportfishing; and an extensive treatment of environmental quality including, but not limited to, waste disposal, pesticides, protection of the beaches and shoreline; and the problems created by mosquitoes and biting flies. The preliminary report was issued because of the urgency of certain decisions facing the state concerning the use of its coastal zone. One of its central recommendations was that a comprehensive baseline study of the principal water bodies of Delaware's coastal zone be performed, in cooperation with New Jersey, Maryland, the Delaware River Basin Commission, and the Federal Government.

What this suggests to me is that many information analysis centers are needed to provide scientists, engineers, and policy-makers with baseline data on all of our coastal zone problems which are quantifiable—and most of them are. Think of the opportunities for inter-disciplinary approaches! Such compilations and analyses will require a joining of scientific, engineering, sociological, economic, legislative, and communications skills. Such approaches usually will encompass a region of two or more states. They often will involve international cooperation. Examples are eutrophication and other problems of the Great Lakes, air pollution from the automobiles and factories of both

Detroit and Toronto, and the many marine problems of the Atlantic, Pacific, and Caribbean coasts. I find it fascinating to speculate about the possibilities for real public service when new or redirected information analysis centers focus their attention on coastal zone or other environmental problems.

* * * * *

I conclude most of my speeches, particularly one like this where so many of my new associates in the Department are present, with a little tribute to them. I was pleased to find, on my arrival in this position the first of March, literally hundreds of highly competent, highly motivated people, in the National Bureau of Standards and throughout the Department. The Secretary and I both appreciate what a priceless resource the Nation has in this staff. We are challenged to experiment, to innovate, and, if necessary, to create new institutions within the Department to expand our technological horizons. We want to find out how these people can better use technology in its proper role to accomplish the mission of the Department. Its proper role, in our view, is to serve people through meeting the needs of business, industry, the environmental community, and other nations through the free enterprise system. I am convinced that the information analysis centers share with us part of the responsibility of serving people everywhere—not just with transfer of information, for many others, such as NTIS, do that. You have the major responsibility for assuring the quality of information and its interpretation for appropriate use. We'd like to tap into your network, into your plans, yes, even into your dreams and aspirations. For we too, have a national stake in better technical information.

THE ROLE OF SECONDARY SERVICES AND INFORMATION ANALYSIS CENTERS

Dr. Russell J. Rowlett, Jr., *Editor*,
Chemical Abstracts Service

I have completely re-oriented my thoughts for this presentation in light of the opening day's discussion, because I feel, as a representative of a discipline-oriented secondary service, that I must say some things that need saying even though they are things which perhaps you will not wish to hear.

Let me begin by looking at the six different components of the information world as I see them. First, of course, is the information *creator*. This is the author or group who starts the whole publication process by preparing the research or technological report or patent. Second is the *primary record*, the publicly available document produced from the original report by the publisher. Third are *document archives*, the storage of the original report by a library which makes it available to those who seek it. Fourth are *data archives*, as in an Information Analysis Center, which most of you here represent. Fifth is the *secondary record*, produced by Chemical Abstracts Service and other of the so-called "abstracting and indexing" services. Finally, at the end of the pipeline we have the information *consumer*, the user of the information as processed by the other five information components.

There is the capability for logical, progressive build-up of an information framework among these six components with a minimum of overlap. Traditionally, however, they have all operated separately and independently. Under the old framework, this was possible because the human intellect — largely the consumer's intellect — bridged the inconsistencies and inaccuracies, of which there are many. For example, as was mentioned here yesterday, *Chemical Abstracts*, *Engineering Index*, and *BIOSIS* found in the very first step of an overlap study that there was difficulty in identifying the "document packages," that is, the journals which the three cover. This was so because the bibliographic citations were inconsistent. The three have never cooperated to arrive at standardized document identification, because in the past, the users of the three services could intellectually bridge the differences. The library community, even more so than the secondary services, continues to rely upon the human intellect to bridge inconsistencies and inaccuracies.

But the growing use of machine handling plus today's economic pressures demand that all of the components of the information world work together for a standardized and consistently identified data package. In principle, there exists in the primary literature a single coverage policy — one paper by one author or group of authors is published only one time. A similar policy needs to be promoted and established in secondary abstracting services. We are working toward this end. The scientific community can no longer afford to have several secondary services analyzing, abstracting, and indexing the same document. And, in my opinion, information analysis centers should be working toward the same elimination of duplicate intellectual effort. We need a standardized identifier for the bibliographic citation and the bibliographic package, and we need a routine procedure for the user to obtain this package from his local library.

The components of the information world must cooperate, not only for elimination of the duplicate intellectual analyses which go on at the primary and secondary publishing stages, and, in my opinion, at the information analysis center; but also, for elimination of the duplicate input keyboarding of identical data. The American Chemical Society is working toward these ends. When we have the primary publications available in machine language, the secondary services will use directly the titles, abstracts, citations, references, and eventually even chemical structures.

Let me emphasize that before *Chemical Abstracts* is willing to eliminate any of its coverage of chemistry in the overlap areas of biology, physics, medicine, etc., we want to be certain we have built bridges into our indexes which will allow the user to go directly from the discipline-oriented secondary service of chemistry to the discipline-oriented secondary service of biology, physics, or engineering. If we do not use the same index terms, then cross-references should guide the user without any doubt. Only when this is accomplished will we be ready for what has been called "mutually exclusive coverage."

Andy Aines asked here yesterday morning what the IAC's and the secondary services could do to help each other. In my opinion, the IAC's should start with the abstracts and index entries available from the discipline-oriented secondary services and build upon them. They should not set up duplicate abstracting and indexing services.

Let's look very quickly at the nature of the secondary information services. First, they are *document*-accessing services. They are not *data*-accessing services. The secondary services provide access to the primary documents, the primary literature. But the abstract is not a surrogate. It has never been and is not today the purpose of the secondary services to replace the primary literature. Their document-accessing function might be compared to the enrichment of an ore. The secondary service selects the ore, refines it, and has it ready to process, but does not actually extract the pure metal. Completion of the process requires separation of specifically needed data from closely associated material. In my opinion, this is the objective of Information Analysis Centers.

The secondary service focuses on new information, on facts not fancy. This new information is not evaluated in the secondary service, and the user, knowing that the accepted values reported by the secondary services are not always authentic values, must make selections based on his own individual needs and experience. It is the place of the IAC's to determine which accepted values are authentic and which are pertinent to the particular task of a specialized user.

IAC's can, in addition, provide a level of data identification which is not possible in a general discipline-oriented secondary service index. Some time ago a CAS survey of the types of data reported in chemistry and chemical engineering showed that chemists and engineers are capable of measuring almost 1100 different chemical properties, uses, applications, activities, etc. Yet scientists continue to demand a guarantee that, every time a particular thermodynamic property related to their individual interests is recorded in a primary paper, CAS carry an index entry for that property. Such a guarantee is absolutely impossible! But, it is possible for secondary services to do a better job of indicating the properties, and the combinations of uses and applications that are recorded in original papers. We need your help in this area. We are going to conduct an experiment soon in which we will code a number of selected abstracts according to the kinds of properties, uses, and applications that are measured. We cannot code for all 1100 properties but we can code for a couple of dozen. We need suggestions on which properties and activities will be of most relevance. In this way the secondary services will also aid the IAC's in their analytical task, for, I repeat, it is my opinion that their task *begins* with the abstracts and the index entries which are available from the discipline-oriented secondary services.

Let me now turn to subject coverage. Here continuity has been considered most essential. We are concerned about continuity of subject coverage across a given field of chemistry, chemical engineering, biology, physics, etc., but we also are concerned about continuity throughout a period of time. Science is a living subject, and the body of information grows with the advance of research understanding. Years ago when spectroscopy was new, its practitioners convinced Dr. E. J. Crane that he should put an entry into the Chemical Abstracts indexes every time someone reported spectra measurements. Today these spectra entries total 30 columns in a six-month volume index. Spectroscopy is no longer an unusual interest, but these entries have been continued to maintain continuity.

The definition of that which has wide utility within the entire scientific community, but which is not so specific as to be useless to a large percentage of those who subscribe to a discipline-oriented service, is a significant problem. Too much detail wastes the money of the subscriber to the general secondary service, and yet this is the detail which is needed to operate an information analysis center. Reaching a happy medium should be our goal. Recognizing that we cannot do your whole job, you should be able to start from what we have done and build from there. At the same time, those who utilize a secondary service in a special framework should not have to redo the work of the service.

Another coverage problem relates to negative data. Here again we need the guidance of those who are using secondary publications. Acceptance of negative data in a secondary service is often based on arbitrary standards. Do you look for an actual quantitative measurement? What do you index in a paper which reports only plus and minus for the activity of a chemical? We need to know what level of negative data the user actually requires. But communicating with subscribers and users is a difficult job for a secondary service. Subscription lists consist of purchasing agents, librarians, secretaries, etc., rather than people who actually use the service. We need communication from users and from IAC's.

Suggestions are needed not only on subject content but also on timeliness. An IAC should have the abstracts and index entries within a time frame that fits its needs. Briefly let me describe just what CAS is doing to improve timeliness. We now receive almost all of the basic journals of chemistry and chemical engineering by airmail in page-proof form. For the *Journal of Organic Chemistry*, an ACS periodical, we are actually using manuscript prior to primary publication. According to the agreements which have been negotiated with the West German Chemical Society and the United Kingdom Consortium, the ACS will eventually also be provided with their primary documents in manuscript form after they have been accepted for publication. These international input centers are performing volume, in-depth indexing simultaneously with abstracting. We are thus eliminating intellectual duplication and are handling a given document in one professional operation.

In the period of over two years that CAS has abstracted and indexed from the original manuscripts of the *Journal of Organic Chemistry*, many significant errors made by the authors in chemical structures, molecular formulas, etc., have been detected. We find such errors by input of the structures to the Registry System. The corrections are returned to the primary journal office in time to be incorporated into the original published record. Thus are the records of the primary journals corrected. There are also errors in the secondary service records. At *Chemical Abstracts* we feel some subscribers just love to look for errors, and we don't like to disappoint them. Today we must handle on the average 1400 abstracts every day. This means prepare, process, edit, and index 1400 abstracts plus about 15,000 index entries, of all types, each and every day. Even our staff sometimes does not realize that we probably process daily more characters than most metropolitan newspapers

Secondary services are subject-oriented in such a way as to have predictable access routes. Certainly you demand this for the kind of use you make of secondary index services. A printed service must depend on a hierarchy, an order within the index, because it is searched by human intellect. It is an organization of access keys. This is particularly relevant to the CAS handling of chemical substances in which an already rigid control of vocabulary has been reinforced in recent years by the development of the CAS Chemical Registry System supported by the National Science Foundation. Today this Registry System enables staff to retrieve 65 to 70 percent of all the *CA* index names and molecular formulas, edited and verified by the computer without professional intervention. This makes possible a significant dollar savings. The names are input just as they appear in primary journals, and the complex chemical nomenclature is retrieved.

It is fortunate that chemistry has a unifying factor ' the molecular structure, which is a two-dimensional structure with a third dimension that can be interpreted. Thus, it has been possible to develop an information system based on this mathematically interpretable structure. Other scientific disciplines are not as fortunate and have not been able to develop systems around such a central structure. Physicists use *Chemical Abstracts* because they desire to search by solid-state compounds, thin films, etc. Actual surveys show that more non-chemists than chemists use *Chemical Abstracts*.

But the secondary services cannot provide all of the data values which you need for complete analysis. We can only lead you to the sources where you can find the data values. We can give you some but not all of the indicated data.

The complete identification of a given concept often results only from a combination of facts. Such facts find their full expression along different axes which are broad in form in several scientific disciplines. It is irrational to expect all of these scientific disciplines to put down separately the same concept. Yet there must be a compatibility between services. This is the point we are urging on our colleagues in the other secondary services: that we build index bridges and clearly indicate them.

Since the needs of information analysis centers appear to include data from several secondary services, the proper index bridging is of great importance to an IAC. With these guideposts, you will be able to go from one secondary service index to another. CAS works by collective index periods, and for the ninth collective period which begins in 1972, we are studying ways to build such bridges between the *CA* indexes, MEDLARS, AIP Classification Scheme, BIOSIS Basic Index, Nuclear Science Abstracts, etc. We are not going to be able to accomplish everything at once, but building the bridges will be a beginning.

In conclusion, let's look at the needs of an Information Analysis Center vs. those of the information consumer. In my opinion, a specialized information center is always interested in exhaustive records of a given type of data. The information consumer on the other hand is often interested in only an accepted value. There are two general kinds of users for *Chemical Abstracts*. One is the man who is looking for an exhaustive record. You see him in the library running his finger down the entire page. No matter where you put the entry in the index he will find it, even if he has to look cover to cover. He is making a patent infringement or domination search, or he wants to find every fact in a new research area. The other man has a pot boiling in his laboratory. He wants one value, a melting point, a frequency line of a spectra, etc. He tears into the library, grabs an index and finds one reference. If the data agree wi what he's got, fine. If he can't find anything, equally fine; he's found something new in the lab. There are differences in the ways of serving these searchers. In an information center, you want an exhaustive record, but a secondary service must also satisfy the user who needs only the accepted value.

How can a secondary service help to point out to information analysis centers where specific types of data are to be found? I mentioned our experiment in coding. I hope that will be helpful. We need additional experiments. Perhaps you have ideas. Another thought, can a secondary service such as CAS indicate to information consumers the particular subject fields covered by an information analysis center? We have a *CAS Source Index*. It was formerly the *List of Periodicals*. It includes the library holdings of almost 400 libraries all over the world; almost 30,000 journals are included. Is there some way within the confines of this Source Index that we can indicate where a user should go for the type of study that an information analysis center can do on a given subject area? I think it is possible. We would like your comments.

I have tried very rapidly to review the components of the information world, to indicate some of the problems and some of the interfaces between secondary services and information analysis centers. I realize I have only scratched the surface. I look forward to your questions.

"THE USE OF ABSTRACTING AND INDEXING SERVICES AT THE ERIC CLEARINGHOUSE ON LIBRARY AND INFORMATION SCIENCES (ERIC/CLIS): A CASE STUDY"

by

J. I. Smith
Associate Director
ERIC Clearinghouse on Library and Information Sciences

INTRODUCTION

The role of the ERIC Clearinghouse on Library and Information Sciences (ERIC/CLIS), operated by the American Society for Information Science for the U. S. Office of Education, centers about three major functional elements:

(1) A Clearinghouse center which acts as a catalyst, focal point and agent for the acquisition, document processing (cataloging, abstracting, and indexing), announcement, and dissemination of fugitive reports and journal literature (in effect, a type of secondary service),
(2) An information service center which handles an increasing number of inquiries, and serves as a referral, or switching center, to existing sources of information in the library and information sciences, and
(3) An information analysis center which identifies burning issues of current need within our scope of coverage, and responds to these by the synthesis and analysis of information from the past and current literature.

Although my main discussion will focus on the information analysis activities of ERIC/CLIS, which gives us a reason for either using or not using abstracting and indexing services, I would first like to briefly describe the ERIC system so that you can fully appreciate our rationale and methods of operation.

ERIC (the Educational Resources Information Center) is a nationwide system established to serve the field of education through the dissemination of information on educational resources and research materials.

The total system functions on both a decentralized and centralized basis, and consists of the following components:

(1) The management group within the U. S. Office of Education, called ERIC Central;
(2) A network of clearinghouses, each with its own subject area of responsibility (ERIC/CLIS is one of these Clearinghouses; its subject area being library and information science);
(3) A central document processing and reference facility, currently operated by Leasco;
(4) A central source for obtaining copies of documents in microfiche and hard copy. This service, called the ERIC Document Reproduction Service (EDRS), is currently operated by Leasco Information Products Co.; and
(5) A centralized contractual endeavor to produce an index of the journal literature in education, currently operated by CCM Information Corp.

"THE USE OF ABSTRACTING AND INDEXING SERVICES AT THE ERIC CLEARINGHOUSE ON LIBRARY AND INFORMATION SCIENCES (ERIC/CLIS): A CASE STUDY"

by

J. I. Smith
Associate Director
ERIC Clearinghouse on Library and Information Sciences

INTRODUCTION

The role of the ERIC Clearinghouse on Library and Information Sciences (ERIC/CLIS), operated by the American Society for Information Science for the U. S. Office of Education, centers about three major functional elements:

(1) A Clearinghouse center which acts as a catalyst, focal point and agent for the acquisition, document processing (cataloging, abstracting, and indexing), announcement, and dissemination of fugitive reports and journal literature (in effect, a type of secondary service),
(2) An information service center which handles an increasing number of inquiries, and serves as a referral, or switching center, to existing sources of information in the library and information sciences, and
(3) An information analysis center which identifies burning issues of current need within our scope of coverage, and responds to these by the synthesis and analysis of information from the past and current literature.

Although my main discussion will focus on the information analysis activities of ERIC/CLIS, which gives us a reason for either using or not using abstracting and indexing services, I would first like to briefly describe the ERIC system so that you can fully appreciate our rationale and methods of operation.

ERIC (the Educational Resources Information Center) is a nationwide system established to serve the field of education through the dissemination of information on educational resources and research materials.

The total system functions on both a decentralized and centralized basis, and consists of the following components:

(1) The management group within the U. S. Office of Education, called ERIC Central;
(2) A network of clearinghouses, each with its own subject area of responsibility (ERIC/CLIS is one of these Clearinghouses; its subject area being library and information science);
(3) A central document processing and reference facility, currently operated by Leasco;
(4) A central source for obtaining copies of documents in microfiche and hard copy. This service, called the ERIC Document Reproduction Service (EDRS), is currently operated by Leasco Information Products Co.; and
(5) A centralized contractual endeavor to produce an index of the journal literature in education, currently operated by CCM Information Corp.

CLEARINGHOUSE ACTIVITIES

Each of the Clearinghouses in the network performs similar functions within its particular subject area. These functions include:

(1) Identification and acquisition of the so-called "fugitive" reports, papers, speeches, etc., which are not published through commercial channels, and the identification of core journal articles, within its subject field;

(2) Evaluation of the documents received;

(3) ataloging;

(4) Abstracting and indexing;

(5) Forwarding the document resume forms (which contain the cataloging information, indexing terms and abstract) to the central document processing facility, along with a copy of the document itself; and

(6) Sending journal article resumes to the contractor for production of the journal index.

The Clearinghouses retain hard copies of each document received for their own library collection, and also maintain a complete file of microfiche of all the documents which have processed by the central document reproduction service. These activities provide each Clearinghouse with a fairly extensive bibliographical base of fugitive document literature.

INFORMATION SERVICES

The document resume forms which are forwarded to the central processing facility by the Clearinghouses, are put into machine-readable form by the facility and the resultant tape is used to produce a monthly abstract publication called *Research in Education (RIE)*. The journal resumes are also put into machine-readable form, for production of the monthly journal index called *Current Index to Journals in Education (CIJE)*. Each Clearinghouse automatically receives copies of each of the two monthly publications for ready reference use. The input tapes for both of these publications, which are updated on a quarterly basis, provide machine-searching and retrieval capabilities.

Building a data base of the document and journal literature in the field of education is a part of the mission of the ERIC system. This data base, along with the broad announcement and document availability service provided by the system, makes ERIC a complete information system, unique in the field of education.

I mentioned earlier that ERIC was both a decentralized and centralized system. The decentralized portion of ERIC consists of the Clearinghouses which work independently with their respective user communities in the way that best meets the needs of those communities, and, at the same time, conform to rules and guidelines established by the Office of Education for processing documents to meet system standardization requirements. The Clearinghouses then feed the processed information into the central document and journal article processing contractors for the production of the magnetic tapes, from which the two monthly announcement publications mentioned earlier are produced. The tapes themselves are made available for machine searching by system users and may be purchased from the central processing facility.

ERIC/CLIS, along with the other Clearinghouses, has a limited capability for providing direct service to users. Financial constraints prevent us from providing great numbers of bibliographies, or separate listings of the documents we process. To most effectively reach the members of our user community under these restrictions, our announcements are limited to those appearing in our newsletters, which contain brief descriptions of the documents which have been processed in the Clearinghouse. Also in order to provide broader dissemination of information to audiences with specialized interests who may not have access to the ERIC publications, ERIC/CLIS sends its document resumes to twelve library and information science journals. Each editor of those journals selects and publishes those abstracts which are relevant to the readership of that particular journal. Last year, ERIC/CLIS reached nearly 60,000 people on a continuous basis through this announcement mechanism, thus providing them with information in their specific subject areas of interest. The point, is that these information service activities are extremely important as a means of keeping ERIC/CLIS constantly before the eyes of the library information science community, and we accomplish this, under our financial limitations, by "piggy-backing" other existing services in our field. As a matter of fact, our input into the ERIC system is used by abstracting and indexing services as part of their coverage.

INFORMATION ANALYSIS ACTIVITIES

In my opinion, the main, and most exciting, aspect of a Clearinghouse is that each of us also serves as an information analysis center for respective user communities. We have constant contact with our users through our acquisitions program and our information services activities, our staff is currently up-to-date on all developments through the document processing of the fugitive and journal literature, and we have a data base consisting of input from the 20 different centers, which means 20 different subject areas within the field of education. Thus, we are very much aware of what is needed in our fields by way of bibliographies, state-of-the-art reports, literature reviews, short papers, etc. We do not get into data compilations or quantitative evaluations, as many information analysis centers do. Instead, the information analysis publications of ERIC/CLIS are aimed at providing information in direct response to the needs of managers, practitioners, research workers and users of libraries and information centers, by synthesizing and evaluating existing knowledge in response to those needs.

These special publications are produced by commissioning an expert in the field to write the report, paying him an honorarium, supporting him with bibliographies and a machine search of all relevant data in the ERIC system in his subject area, obtaining copies of papers, and providing minimal funds for typing the manuscript and for local expenses. In other words, our basic involvement entails the provision of bibliographic support to the authors. Most of these authors are not part of the ERIC/CLIS staff. The reason for this is simply one of economics. Our operation is just too small to allow a staff member to take time from his (or her for the sake of women's rights) other duties to write such papers, whereas the staff, working as a whole, can provide much more effective support to outside authors who are knowledgeable in a particular topic.

The overall scheduling for our information analysis products is generally as follows:

(1) Collect subject areas of interest and concern which have been expressed in letters, personal communications, and the literature;

(2) Evaluate these subject areas to establish priority needs through consultation with users and advisory boards;

(3) Select the target areas to be considered and identify potential authors who have capabilities in these areas;
(4) Communicate with potential authors and establish agreements for honoraria, bibliographic support and date of completion, as within six months of agreement date;
(5) Send letter of confirmation;
(6) Send author's guide;
(7) Initiate machine search of the ERIC data base, and send copy of search results to author;
(8) Send updates of that search on a monthly basis to the author;
(9) Receive rough draft of report and submit to other authorities in the field for review and appraisal;
(10) Send comments from reviewers to authors, and begin editing the report (the editing is done by an ERIC/CLIS staff member);
(11) Prepare the paper for final typesetting and printing, and
(12) Distribute copies of the report.

As you can see, bibliographic support keeps everybody quite busy and when you have approximately 50 different information analysis projects in motion at the same time, you can well imagine the amount of work required to support these projects. We also provide bibliographic support to the authors of the *Annual Review of Information Science and Technology*, as well as review publications in the library and information science field.

THE USE OF ABSTRACTING AND INDEXING SERVICES

Although the fugitive documents contained in the ERIC data base are invaluable as a source of information, we find that they are by no means the complete answer in providing authors with the broad range of literature needed to write a good review, or compile an extensive bibliography. Therefore, we find it necessary to go to other sources in our field which cover the journal literature extensively. Although we routinely process approximately 20 core journals in the field, there are about 500+ journals in the library and information science fields, which means that we do not approach complete or exhaustive journal coverage, by any means.

In addition to the data base we have created for the ERIC system there are about four main abstracting and indexing services in the library and information science fields.

The problems in using these secondary services are as follows:
(1) The terminology and classification schemes of each service differs substantially;
(2) The material contained within the publications of these secondary services, for the most part, is quite old;
(3) There is a significant overlap in the journals covered by these secondary services, and
(4) The indexes have not been computerized and, only manual searching can be done, which is, of course, time consuming.

The apparent disadvantage of using secondary services as part of our bibliographic support is, however, ironically enough, a blessing in disguise because the field of library and information science is indeed being thoroughly covered by the abstracting and indexing services in the field. For one, this means that we can concentrate to a larger extent on the fugitive literature, for which we alone provide service of this kind in the field. It also means that our authors have access to these secondary tools, which is why we commission only those people who are active and knowledgeable in the field, since we assume that these authors are not only aware of these secondary services but are, in fact, using them.

The main point is that we have, first of all, identified the secondary services within our field, evaluated each of them for possible application to our own bibliographic support activities, have used them on a very selective basis, and are making efforts to make changes. We also scan approximately 20 primary publications, especially those which have reviews in them which may be of interest to our authors, but as yet we have not incorporated these efforts into any kind of organized bibliographic activity. One of our special projects is to develop a complete alphabetized listing of all terms in our field with a reference as to how that particular term is used by a particular secondary service. This, we hope, will be of benefit to us when we do retrospective searching outside of our own data base. Another special project we have initiated is that of incorporating all of the references cited in the fugitive documents and compiling a type of citation index to these references. We hope to put these references into machine-readable form and combine them with the data base we already have. In effect we are building our own data base, rather than using abstracting and indexing services except on a manual or scanning basis.

THE ASSOCIATION OF SCIENTIFIC INFORMATION DISSEMINATION CENTERS

For those of you who are interested in using abstracting and indexing services, I suggest that you make contact with the Association of Scientific Information Dissemination Centers (ASIDIC). The purposes of this Association are:

(1) To promote the applied technology of information storage and retrieval, as related to large data bases containing bibliographic, textual, and fact information;

(2) To share experiences in information handling through meetings, seminars and workshops;

(3) To recommend standards for data elements, formats and codes; and

(4) To promote research and development to provide a more efficient use of existing and varied data bases.

Membership in this group is held by organization, not by individuals, and information centers which meet the following criteria are eligible for full membership:

(a) Center operations are computer-based;

(b) Data bases from two or more suppliers are processed; and

(c) A minimum of 100 user-interest profiles are processed on a continuing basis.

There is also an associate member status available for suppliers of machine-readable data bases, and for other organizations which have an interest in the affairs of the Association, but do not meet the criteria for full membership.

The members of ASIDIC essentially reprocess tapes procured from organizations such as Chemical Abstracts Service, Engineering Index, the Institute for Scientific Information, Biological Abstracts, and others for their individual information purposes. These member centers have developed program packages that are capable of searching multiple data bases. A few of the centers which belong to ASIDIC are: the University of Georgia; the Illinois Institute of Technology Research Institute, Chicago; the University of Iowa; the University of Nottingham, England; the National Science Library in Ottawa; and the University of Pittsburgh.

Any of you who are interested in providing searching services for your centers from some of the major tape services available might want to contact ASIDIC for further information. The President is Dr. James L. Carmen of the University of Georgia, and the Secretary, who sends out the newsletters and other information, is Miss Diana Follmer, 3M Center, St. Paul, Minnesota 55101. You may want to contact her to be put on the mailing list.

USES OF ABSTRACTING AND INDEXING SERVICES IN IACs

Robert E. Snider, *Director*
Air Force Machinability Data Center

A case study of the use of abstracting and indexing services by the Air Force Machinability Data Center (AFMDC) disclosed very limited utilization of these services. I will explain why we at AFMDC have not been able to justify a more extensive use of them.

First, however, I will describe the scope of operations of our Center for the purpose of showing that AFMDC is somewhat unique to the centers that deal with chemistry, metallurgy or electronics.

The Air Force Machinability Data Center is located in Cincinnati and is operated by Metcut Research Associates Inc. under a contract with the Air Force Materials Laboratory. At AFMDC we collect, evaluate, store and disseminate material removal information including specific and detailed machining data for the benefit of industry and government. A strong emphasis is given to engineering evaluation for the purpose of optimizing data being disseminated.

Data are being processed for all types of materials and for all kinds of material removal operations such as turning, milling, drilling, tapping, grinding, electrical discharge machining, electrochemical machining, etc.

AFMDC is using a computerized system for storage and retrieval of some 26,000 coded documents related to the material removal processes.

As I stated earlier we have not been able to make extensive use of abstracting services. One of the primary reasons being that there seems to exist a language barrier between abstractors and the terminology used within the material removal industry.

Charles T. Meadow in his book entitled "The Analysis of Information Systems"([1]) said:

> "Almost all index languages in use are to some extent artificial. A natural language, although hard to define, is easy to illustrate. English, French, and German are natural languages. They are the languages that people naturally speak. Index languages are invented, not for general communication, but for a very special form of communication—that of enabling indexers and library searchers to communicate with each other and, in a sense, with the documents of the library. The particular role that the language is to play will vary with the library, the collection, and the users. Selection or design of an index language is probably the single most difficult step in designing an information retrieval system; in our opinion, the biggest single reason for this is our general inability to predict the performance of human beings when faced with a communication system different from that with which they have become familiar. Our approach here is to present some basic principles for the design and use of these languages, leaving it to the designer of an individual system to apply them to each local condition."

[1]Meadow, Charles T., *The Analysis of Information Systems*, John Wiley & Sons, Inc. New York, 1967

In past experience in trying to establish an AFMDC Interest Profile with some abstracting services, we have encountered somewhat the same experience that an aerospace engineer found when he approached his computer supervisor and said that he would like to find some reports on research conducted on bearing materials applicable to high flying aircraft. Two uniterms were selected for the computer search "bearing" and "high altitude." Only one document was cited and upon recovery of this document it was found to have the title "Child Bearing in the Himalaya Mountains."

In the manufacturing community much of the terminology, although natural to English language, have different meanings and thus cause monitors of AFMDC's interest profile and abstractors to cite many documents that are not relevant to our needs. For example, simple words such as turning, milling and tapping can relate to various fields. The word milling in one report may be describing milling as performed in the ore industry; in another it may be talking about the basic powder metallurgy industry. In the material removal field, milling is a term used for cutting material on a milling machine by chip making and it can also be used in a nonconventional machining operation of material removal by chemical attack called "Chemical Milling."

Thus at AFMDC, we have established our own abstract review using personnel familiar with the manufacturing industry and trained in the knack of recognizing documents relevant to our needs.

At the present time, we are searching the following abstracts for acquisition and in so doing have developed a knowledge of where within many of these that our field of interest normally appears. Certain areas of these abstracts have proven to be the most productive to AFMDC:

Metal Abstracts (ASM)
International Aerospace Abstracts (IAA)
United States Government Research & Development Reports (USGRDR) (U.S. Dept. of Commerce)
Scientific and Technical Aerospace Reports (STAR)
Aerospace Research Applications Center (ARAC)
Current Awareness Programs from DCIC, etc.
NASA Technical Briefs, etc.
Materials Information Bulletin - AFML (Contracts)
Some periodicals have abstracts section, especially foreign
Publication listings in society journals

I would like, in closing, to say that we at AFMDC certainly appreciate the valuable time saving contribution abstracting services are providing the information community involved with the majority of fields of science. However, AFMDC could utilize them to a fuller extent if both abstractors and interest profile monitors in some abstracting services were more oriented to the material removal industry.

A PROFILE OF SCIENTIFIC-TECHNICAL TAPE INFORMATION SERVICES

John M. Gehl and Vladimir Slamecka
School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Ga. 30332

Introduction

We wish to trace the outline of some of the main features of scientific-technical tape services which have been developed during recent years. In preparing this profile, we have drawn almost exclusively on Kenneth D. Carroll's *Survey of Scientific-Technical Tape Services* published by the American Institute of Physics in September 1970. Although not quite up-to-date, this survey suffices our purpose of exhibiting commonalities and variations of characteristics of these services.

To begin, we quote briefly from the motivation section of the Carroll report, which puts the results of the AIP survey in perspective:

> "During the past few years there have been an increasing number of tape services entering the information resources market. Each of these services makes available to a library or information center, on a continuing basis, computer-readable data which can be utilized in as many diverse services as the center's programs and clientele require. As these services increased, it was sometimes a problem for libraries and information centers to keep up with all the various data bases available, the subject areas covered by the tapes, whether the organization offering the tape performed in-house services upon request, or if software was available to the subscriber. In the preliminary survey reported here, we have tried to compile a directory of current tape services, listing for each service the general characteristics of the data base."(Page 2)

In recognition of problems such as those mentioned, the Carroll study solicited information from representatives of all known commercially available tape services (including two services offered by Federal agencies: ERICTAPES, from the Educational Resources Information Center; and U. S. Government R&D Reports, from the National Technical Information Service). Information obtained from these inquiries is shown in the report under the following categories:

Name of Tape

Source

Contact-Representative for further information

Characteristics of the data base (including: subject matter;
types of source items input; methods of subject analysis or indexing; searchable data elements; availability of abstracts; and time span available)

Frequency of tape issue

Average number of source items per tape

Subscription cost or leasing details

Software availability

Type of in-house service offered

Publications produced from base by originator

Fifty-six services are listed in the Carroll report. (One of these appears in name only - the IEEE Annual Index Tapes, a service which at the time was in the final stages of development.)

The "Developers"

The principal sources for commercially available scientific-technical services are learned and professional societies, publishing firms, and commercial organizations. The CCM Information Corporation offers six separate tape services, Chemical Abstracts Service offers seven, Derwent Publications, Ltd (England) also offers seven, and Predicasts, Inc. offers five. Three tape services are offered by organizations within the U. S. Government — the two already mentioned, and MARC, the machine-readable cataloging distribution service offered by the Library of Congress. Four of the services are offered by institutions located outside of the United States: in addition to Derwent Publications, already mentioned, they are the Institute of Electrical Engineers (England), Shirley Institute (England), and the International Atomic Energy Agency (Austria). Finally, we might note that one service is directly affiliated with a university: that service is Petroleum Abstracts, produced by the Information Services Department of the University of Tulsa, Tulsa, Oklahoma.

Virtually all of the organizations studied use their data bases to produce one or more publications; these are either bibliographies, indexes, abstracts, thesauri, keyword supplements, patent review books, data books, or similar products under different names. For example, the National Information System for Physics and Astronomy, which offers a tape service called SPIN (Searchable Physics Information Notices), publishes, among other products, the current awareness journal *Current Physics Titles*.

Subject Coverage

The subjects covered by tape services span almost the entire range of scientific knowledge, though coverage is not equally balanced from subject to subject. Chemistry and chemical engineering, for example, are specifically covered by eleven different tape services; of these, one focuses on marketing information, another on patent information, and two others on those portions of chemistry and chemical engineering which are directly pertinent to the petro-chemical and petroleum refining industry.

Nor is petroleum the only industry which receives explicit attention. Another example we could cite is the pulp and paper industry, which is the subject of three tape services, all of which are produced by the Institute of Paper Chemistry. These services comprise the tape equivalents of the *Abstract Bulletin*, of the *Author and Patent Indexes* for that publication, and of that publication's

Keyword Supplement. Yet another example of an industry-oriented tape service is that offered to the textile industry by the Specialized Information Service Data Base, produced by Shirley Institute.

Two of the services now available are concerned with polymers, plastics and macromolecules; one simply with plastics; and one with plastics and electrical/electronics engineering.

One service covers diodes, transistors, microwave tubes and integrated circuits; one covers physics, electrical/electronics engineering, computers and control engineering; and one covers electrical/electronics engineering, computer science, and applied physics. The subject matter of another is physics and astronomy.

Three services are concerned with the mathematical sciences; five with metallurgy, farming, agriculture, or the earth sciences; six with biochemistry, virology, or the life sciences. No less than seven are focused on statistical or financial information.

Finally, an additional seven tape services provide broad or interdisciplinary coverage. These include: the CCM Corporation's Current Index to Conference Papers in Engineering; COMPENDIX, a service of Engineering Index, Inc., which covers all fields of engineering and certain fields of applied science and management; PANDEX, which provides broad coverage of scientific, technical and medical journals; U. S. Government R&D Reports, a service whose coverage includes not only scientific and technical subjects but social sciences as well; ERICTAPES, which are concerned with providing coverage of varied aspects of education; the Institute for Scientific Information's Combined Source and Citation Data Tape, which offers broad interdisciplinary coverage of journal literature, including the primary journals of basic and applied science, engineering and technology, medicine, psychology and psychiatry, and the behavioral sciences; and that same Institute's Source Data Tape, which provides similar coverage.

One last service which we will single out for special attention is the INIS Output Tape, which is produced by the International Atomic Energy Agency and which covers nuclear science and technology.

Volume of Data and Periodicity

At this point it may be appropriate to give some idea of the volume of information provided by these services. The rather crude measuring unit for this purpose will be the number of items cited per tape. Of the total number of services for which information on this question is available, approximately one half cite more than 5,000 source items on each tape. Two of these in fact cite more than 20,000 such items. One is ICRS, the Index Chemicus Registry System tape, which cites 4,000 abstracts and 17,000 Wiswesser Line Notations on each monthly tape, for a total of 21,000 items; the other is Predicasts Corporation's F&S Index of Corporations and Industries, which include approximately 25,000 source citations on each of its quarterly tapes.

Considering the wide variety of topics covered by these tape services, it is not surprising to find that the number of tapes issued each year is quite different from service to service. Virtually every conceivable time interval is represented - weekly issues, three issues a month, biweekly issues, semimonthly, monthly, eleven issues a year, quarterly, every four months, semiannually, and annually. Approximately 75% of all of the services issue tapes either on a monthly or an even more frequent basis. Of all the services, only one offers services on the basis of the frequency of issue requested by the particular subscriber; that service is the textile information service of the Shirley Institute in England.

Combining now the information available on both the average number of source items cited on a tape, and the frequency of tape issue, we may conclude that, for 45 services for which sufficient data was available on this question, almost half of those services cite more than 25,000 source items annually. Of this group, seven cite more than 200,000 items annually, and of those seven, there are two which cite more than 300,000.

Cost

The cost for all this information is not always low. However, more than 75% of the services are offered at annual costs of $2,500 or less. Apparently the most expensive service is the Institute for Scientific Information's Combined Source and Citation Data Tape; the subscription cost for this service is $20,000 a year. On the other extreme is the service offered by the International Atomic Energy Agency, which is free to member states. The subscription cost of the Library of Congress MARC tapes, which provide, on a weekly basis, current English-language monograph cataloging data, is $800 a year.

Three of the services covered by the survey base their subscription charges on the subscriber's gross assets. Two of these are offered by the American Petroleum Institute, and the third by the University of Tulsa's Petroleum Abstracts service. As an indication that some subscribers do indeed have greater assets than others, we may note that average yearly costs for a subscription to the University of Tulsa service range from $200 a year to one hundred times that: $20,000.

Sources of Data

We ought at this point to get back to the question of what the information purchased from these services is all about. We have already discussed the subjects covered; let us now turn to the question of the scope of these services. Doing so, we may note that - as we might have expected - a large portion of current data bases are devoted to coverage of the journal literature. More than three out of four services are such that 50% or more of their data bases are devoted to journal coverage, and almost two out of three are such that journal coverage accounts for at least 80% of their total data base volume.

The number of journals covered by the services varies from precisely one - which is the case for the Mathematics Computation Magnetic Tape, which is produced by the American Mathematical Society and which is comprised entirely of the contents of the one journal *Computational Mathematics* - to the 4,500 journals which contribute in an average year to the input of BA Previews, the tape produced by the BioSciences Information Service of *Biological Abstracts*. More than three out of four of the services for which information on this question is available provide coverage of at least 500 journals.

A quite large percentage of this journal coverage is accounted for by English-language literature. Only one of the services has characterized its data base as predominantly (i.e., more than 50%) non-English. That service is the one offered by the American Geological Institute. Its subject is the earth sciences (including areal, economic, engineering, extraterrestrial and marine geology; geochemistry; geochronology; geohydrology; geomorphology; and so forth); and it reviews 1,600 journals for input, only 40% of which are in the English language.

One data base characteristic which varies considerably from service to service is the percentage of journals which are entered in entirety, as contrasted with the percentage which are reviewed and entered into the data base only selectively. Approximately half of the services fall into the former category. Examples of services which enter journal literature into the data base on a selective basis are the Virology index; Search-Data, a service offered by Compendium Publishers to provide abstracts with citations of original sources of marketing information in chemical and allied fields; and Expansion and Capacity Digest, one of the market-oriented tape services offered by Predicasts, Inc. Examples of data bases into which journals are entered in their entirety are the SPIN tape; PANDEX; the *Chemical Abstracts* service's Basic Journal Abstracts; and the Institute for Scientific Information's Combined Source and Citation Data Tape.

Approximately one out of three of the scientific-technical services for which information is available indicated that at least some part of their data base is devoted to coverage of the report literature, but only three out of twenty devoted more than 10% of their data base to such coverage. The three services devoting the highest percentages of their data bases to coverage of Government-sponsored research described in the reports literature are: the International Atomic Energy Agency tape (25%); WORLDCASTS (25%); and, obviously, the U. S. Government R&D Reports tape (69%).

Another source of input which deserves separate attention is the patent literature. Nine services devote at least 25% of their data base to such coverage; of those nine, there are four which are devoted exclusively to that purpose.

Three of the scientific-technical tape services are devoted exclusively to the coverage of papers presented at conferences. The three services, all of which are offered by the CCM Corporation, are: the Current Index to Conference Papers in Chemistry; the Current Index to Conference Papers in Engineering; and the Current Index to Conference Papers in Life Sciences. Each of these indexes provides coverage of approximately 1,200 meetings in the field specified. The chemistry tape service reports on 7,500 individual papers, the engineering service on 32,000 papers, and the life sciences service on 15,000 papers.

The data bases maintained by three others of the services are devoted to statistical or historical data. They are: D.A.T.A. Book Files, a service which is offered by a division of Computing and Software, Inc., and which is concerned exclusively with manufacturer-supplied information on diodes, transistors, microwave tubes, and integrated circuits; the Growth and Acquisition Guide tape, a service offered by Predicasts, Inc., to provide tabulated information on acquiring and acquired companies by four-digit line SIC, line of business, and latest annual sales given for both; and COMPUSTAT, a financial-oriented service which provides information on public-held companies. COMPUSTAT is offered by Investors Management Science and is comprised of data taken from company annual reports, government reports, daily news services and company contacts.

Just as we have previously seen that commercially available tape services provide coverage of virtually every subject, so we now find that they monitor virtually all forms of printed information - books, journals, reports, monographs, theses, patents, newspapers and even private information sources.

Aspects of Information Control

Having now a general picture of the characteristics of the data bases themselves, we may want to review briefly some of the techniques used for controlling those data bases. How is all of this information moved about? How is it indexed for effective storage in the data bases, and how is it searched for and retrieved?

We come first to the question of subject analysis and indexing. Of the approximately 30 services for which information on this question is available, roughly half assign an average of from five to ten indexing terms to each item in the data bases. However, several of the services use a far larger number of terms to describe an item. These services are: the American Petroleum Institute's Index to API Abstracts of Refining Literature, for which an average of 35 index terms or descriptors are assigned each entry; the IFI/Plenum Data Corporation's Uniterm Index to U. S. Chemical and Chemically-Related Patents, for which the average number of index terms or descriptors assigned per item is also 35; Investors Management Sciences' COMPUSTAT service, which on an average assigns approximately 60 index terms to each item; and SEARCH-DATA, of the Compendium Publishers International Corporation. The SEARCH-DATA service assigns on the average approximately 100 index terms to each item included in the data base.

The indexing terms and descriptors referred to in the above figures included both controlled descriptors and free-language terms. Of 36 services for which an answer to this question could be determined, nine (i.e., one out of four) relied entirely on free-language indexing, whereas 27 (i.e., three out of four) specified the use of controlled descriptors. However, of those 27 services which used controlled descriptors, approximately 50% employed free-language indexing as well.

Relatively few of the services used classification schemes. Those which were used include: UDC (used by the tape service of the American Geological Institute), the American Mathematical Society's Subject Classification Scheme, the classification system designed for *Mathematics of Computation - Midwest Research Institute*, the National Agriculture Library Classification, the Subject Headings for Engineering system, the IEE/IEEE INSPEC system, the International Atomic Energy Agency's INIS classification schedule, and the U. S. Patent Office's classification codes.

About 40% of the data bases contain abstracts or their equivalent.

Techniques for searching the various data bases differ considerably from one tape service to the next. Of the numerous services which cover authored material (journals, reports, etc.), all but one allow searching of the file for the name of the first author and (in most cases) all other co-authors as well. (The exception is CITE, a service of Engineering Index, Inc., devoted to applications technology in plastics and electrical/electronics engineering; CITE's tape records include a "searchable" segment, composed entirely of index terms, and a nonsearchable "display" segment that identifies author, title, and citation.)

In addition, some services allow a search based on the institute with which an author is affiliated, his location, the sponsor of his work, and/or the publisher of his book. Other services offering variations on "authorship"-search include those which allow searches based on corporate authors, editors, patent assignees, or manufacturer names.

Besides author's name, another searchable data element allowed by almost all systems is, of course, the title of the article, report, or other authored document. Bibliographic information also offers a legitimate search device for most of the services considered. For journals, such information usually includes such searchable items as journal title, CODEN, journal volume, issue, and page numbers; for other material contained in the data base, a search may be conducted on such items as: conference name; report number; patent number; specially assigned document accession number; and so forth.

But beyond such standard items as author, title, and basic bibliographic information, most services allow searching of the data base in various other ways; the extent of this variation can perhaps be indicated by a simple (but certainly not exhaustive) enumeration of some of the data elements upon which a search of the data base content may be conducted for one or more of the tape services. Such searchable data elements, then, include: descriptors (with or without links and roles); keyword phrases; words in a document's abstract; the language in which a document is written; primary and secondary subjects of a document; indexing terms and title enrichment terms; and classification codes.

For an example of searchable data elements allowed by statistics-oriented tape services, one could cite WORLDCASTS, which allows searching on any of the following items: industry-product SIC; country; event; event code; year; earliest year first; quantities; smallest quantity in given year first; unit of measure by type of unit; source (publication); quote (the name of the person making the forecast).

As a final example of a searchable data element, we might mention the provision in CCM Corporation's Virology Index which allows for searching exclusively either for review articles or for articles of a non-review type.

Hardware/Software

Of course, to perform a search on any of the data elements specified for any of the tape services, a subscriber needs not only the suitable computer hardware, but appropriate software as well. With reference to approximately one out of three presently available tape services, a tape subscriber will be required to develop his own software. In such a case, the subscribing institution will use the tapes strictly as additional input to its own system, and will use its own software and its own search strategy.

However, in the remaining cases, the institution which produces the tape either already offers supporting software to its subscribers, or will develop whatever software is required for an interested customer. In addition, some tape services have indicated that, although they do not themselves offer supporting software, various suitable search programs are available elsewhere on the commercial market.

Attempting to offer only the briefest profile of the characteristics of some of the programs available in the various packages, we will merely recite the principal features of just one important group of services (those provided by the CCM Corporation), and will supplement this recitation with an equally brief consideration of some of the typical ways in which other services deviate from that paradigm.

CCM Corporation, then, provides COBOL programs both for print-out and for SDI; the appropriate computer configuration is an IBM 360 with disk operating system and a 32,000-word core memory. The system uses 7- or 9-track magnetic tape with 800 bits per inch. Records are written in either fixed-field or MARC II format, and coding is EBCDIC, BCDIC, or ASCII.

The most obvious ways in which descriptions of the capabilities of other services differ from the CCM services pertain either to the programming languages used or to the basic choice of computer equipment. Other languages in which software is written for these tape services include assembly language, autocoder, FORTRAN, and FORTRAN with ALC.

Other hardware configurations are found in systems which require more core memory (e.g., the Uniterm Index to U. S. Chemical and Chemically Related Patents, which uses a 256K memory), or different operating systems (although some services offer various operating system alternatives, such as the Index Chemicus Registry System, which comes in DOS/TOS and OS versions). Only two of the tape sources use equipment other than IBM's; they are Predicasts, Inc., which uses a UNIVAC 1108, and Investors Management Sciences, whose information retrieval program will run on a Control Data Corporation 3300 computer (as well as on an IBM 360).

Other Services and Charges

However, not all of the benefits which a subscriber may obtain from the tape services are dependent upon his own search capabilities and his own computer equipment. To the contrary, a number of the organizations producing scientific-technical tapes now offer, or are planning to offer, in-house search services.

One such obvious service is the capability for conducting retrospective searches, through the data base, based on a subscriber's search profile. The methods for calculating the charges for retrospective searches understandably vary from service to service. Retrospective searches performed in conjunction with the six tape services offered by *Chemical Abstracts* are based on a flat fee (ranging from $2,100 to $4,400) plus an assessment for the cost of actual computer time used to conduct the search. The University of Tulsa calculates charges for in-house retrospective file searching on the basis of $10 for each hour of search time in addition to $1.00 for every pertinent reference found. *Biological Abstracts* (BA Previews) charges $150 a search as does IFI/Plenum Data Corporation. (The latter also offers reduced rates for contracts of 50 searches.) The American Society of Metals (Metals Abstract Index Data Base) charges $250 for a search of this kind. A final arrangement we might mention is one devised by the American Geological Institute, which calculates its fees on the basis of a rate of $10 per query per 50 items retrieved.

Since "retrospective" searches have acquired that name for the very good reason that they proceed backwards into existing literature, it is important that we determine just how far back a researcher may "look" when he relies on these tape services. The basic answer to this question is that the time span available depends considerably on which particular service is of interest to the subscriber; some services provide coverage of their subject matter going back further than 1960; others go back no further than January 1970. As a general indication, we may note that approximately half of the tape services do not begin coverage of their subjects until 1968 or later.

The other principal in-house service available from producers of scientific-technical tapes is SDI - Selective Dissemination of Information. Although a number of SDI services are still in development or early implementation stages, as many as five have been operating for one or more years. The pricing policies of SDI services are suggested by the following two basic kinds of rate structures in effect during 1970:

1. The SDI service associated with the American Mathematical Society's Mathematical Off-print Service offered title listings, abstracts or off-prints; costs for these services were: 5¢ per title selected, 25¢ per abstract selected, and 45-85¢ per off-print selected (depending on article length).

2. The American Society for Metals offered current awareness services at $250 a year per search profile. *Biological Abstracts* (BA Previews) offered CLASS (Current Literature Alerting Search Service) at $100 a year per search profile. An identical rate ($100/year/profile) was adopted both by the *Keyword Supplement* of the Institute of Paper Chemistry and by SEARCH-DATA.

Conclusion

Thus we can conclude that a number of tape services exist, and that they provide coverage of different areas of knowledge, at different levels of depth, from different viewpoints, using different information control and search techniques - and making different demands on a user's pocketbook. Furthermore, we confirm that the difficulties several university-based information services report with attempts to pool and efficiently use several tape services are neither imaginary nor understated: the range of variety of different characteristics is indeed very broad among the tape services we have compared. And it is equally easy to understand the feeling of indecision of a prospective user attempting to select the one tape service which optimally meets his situation.

The premise which underlies the utility and validity of a comparative survey such as we have presented is the necessity and sufficiency of the parameters in terms of which such comparisons are made. We fear, however, that we cannot defend this premise; we do not know whether the parameters of comparison are useful for either of the two major clients interested in surveys - those attempting to select the best service for their needs, and those seeking to pool several tapes for a wider and more efficient service. Nor do we have any evidence that a much larger number of parameters (such as prepared by Schwartz[1]) can be employed to construct a decision-making algorithm for either category of potential users, even if one assumed the unlikely situation that such detailed descriptions of data bases can be obtained and made public.

Thus while paying attention to monitoring the characteristics of tape services, perhaps we ought to be giving more thought to the idea of surveying the customers, actual and potential. What experiences and recommendations do they have? What categories of parameters are of importance to them? What is the level of minimum compatibility, or desirable compatibility? Are there guidelines for the design of tapes and services which a users' association might wish to impress or impose upon the producers? In proliferating diversity of technical design are we indeed concerned with the management of information as a national resource?

These are among the many thoughts occurring in the margins of a simple survey of information tape services.

[1]J. Schwartz. "A Checklist for the Examination of Data Base Systems." New York University, 1970 (Mimeographed). 6 p.

NSIC COMPUTERIZED INFORMATION TECHNIQUES*

William B. Cottrell
J. R. Buchanan
Nuclear Safety Information Center
Oak Ridge National Laboratory
Oak Ridge, Tennessee

TABLE OF CONTENTS

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

1. Introduction

The recent spectacular growth of the nuclear power industry and the accompanying increase in problems and concerns regarding nuclear safety have given rise to a deluge of information and data in all types of documents and formats. The Nuclear Safety Information Center (NSIC) is helping resolve the dilemma faced by scientists, engineers, and others in the field by refining and collating the information into more readily digested forms of output. (1, 2)

The purpose of this paper is to describe for the COSATI Forum of Federally Supported Information Analysis Centers, Washington, D. C., May 17-19, 1971, the computerized techniques which NSIC is using in its mission. Equally outstanding but non-computerized functions of NSIC such as state-of-the-art reports, the journal, *Nuclear Safety,* and technical consultation are outside the scope of this paper and will not be covered here. Prior to discussing the individual outputs, the Center's information system and computer hardware and programs will be discussed. Following this, there will be a prognosis on what is foreseen in IAC computerized activities.

2. Information System

NSIC was established in 1963 by the USAEC Division of Reactor Development to collect, analyze, and disseminate nuclear-safety-oriented information throughout the nuclear community. (1, 2) The Center's subject scope is divided into the 21 categories listed in Table 1.

Table 1. Information Categories

1. General Safety Criteria
2. Siting of Nuclear Facilities
3. Transportation and Handling of Radioactive Materials
4. Aerospace Safety
5. Heat Transfer and Thermal Transients
6. Reactor Transients, Kinetics, and Stability
7. Fission Product Release, Transport, and Removal
8. Sources of Energy Release Under Accident Conditions
9. Nuclear Instrumentation, Control, and Safety Systems
10. Electrical Power Systems
11. Containment of Nuclear Facilities
12. Plant Safety Features
13. Radiochemical Plant Safety
14. Radionuclide Release and Movement in the Environment
15. Environmental Surveys, Monitoring, and Radiation Exposure of Man
16. Meteorological Considerations
17. Operational Safety and Experience
18. Safety Analysis and Design Reports
19. Radiation Dose to Man from Radioactivity Release to the Environment
20. Effects of Thermal Modifications of Ecological Systems
21. Effects of Radionuclides and Ionizing Radiation on Ecological Systems

Some of the many NSIC activities established to accomplish these objectives are listed in Table 2.

Table 2. NSIC Services

1. Preparation and publication of state of-the-art reports
2. Cooperation in preparation of *Nuclear Safety*, a bimonthly technical progress review
3. Preparation of abstracts of nuclear safety literature
4. Publication of topical indexed bibliographies
5. Selective dissemination of information (SDI)
6. Answering technical inquiries
7. Preparation of special retrospective bibliographies
8. Compilation of information on current research and development
9. Provision of technical consultation
10. Collection of documents for review by qualified visitors

NSIC's director is also director of the ORNL Nuclear Safety Research and Development Program and editor of the technical progress review *Nuclear Safety*. NSIC's assistant director is responsible for the overall supervision of the Center's activities. The professional staff at NSIC consists of over 30 technical specialists, the information specialists, and editors. The technical specialists are scientists or engineers who divide their time between working on research and development problems and working at the Center. In this way, they are current both in what is going on in their specialty and in the documentation of new results. While working at NSIC, they write state-of-the-art reports, abstract documentary material, and act as consultants to people in the nuclear industry.

Two of the technical staff work principally with the computerized aspects of the Center, [3, 4] set up special searches, process the Selective Dissemination of Information profiles, and handle file maintenance. The NSIC programming and processing are done at the Computing Technology Center. Most of the information processing was initially done on the IBM-7090, but a complete changeover to the IBM-360 computer has been accomplished.

The process flow [5, 6] for a document and the information that is eventually stored in the computer on that document is as follows:

1. Documents for review are selected by an information specialist, who routes them to appropriate staff scientists or engineers, each a specialist in some aspect of the Center's scope.

2. The technical specialist then scans the reports, journals, and articles, etc., assigns categories (see Table 1) and keywords from a vocabulary of 3000, and prepares a 100-word abstract on office forms called "green sheets." [9] Other information centers at the Laboratory use other colors.

3. The technical specialist sends the green sheets and documents back to the information specialist, who makes certain bibliographic indexing entries on the forms.

4. The green sheets are then sent to the editor, who edits the entries and then sends the green sheets to typists for computer entry.

5. The typists use IBM-2260 Cathode Ray Tube (CRT) scopes to enter the abstracts, keywords, etc., directly into the computer where the references enter a storage "data cell" and are immediately available for retrieval.

NSIC processed over 15,000 items during 1970 and there are, as of May 1, over 57,000 accessions in the NSIC files, 53,000 of which are in the computer. The description for each document in the computer file includes the elements listed in Table 3.

Table 3. Elements Included in NSIC Computer Files for Each Accession

1. Accession number
2. *Type, such as reports, journal articles, etc.
3. *Evaluation of contents (as to pertinency)
4. *Category (such as Accident Analysis)
5. *Journal abbreviation (ASTM's Codes)
6. *Date
7. Availability
8. *Language
9. *Country
10. *Corporate author
11. *Personal author(s)
12. Title
13. Item, such as pages, figures, and tables
14. Abstract (~100 words)
15. *Keywords

*The asterisk indicates that the element is searchable or may be used to restrict a search.

3. Computer Hardware

NSIC's computer activities are conducted on a time-shared basis on an IBM-360 Model 50/65, a third generation digital computer, at the Oak Ridge Computing Technology Center (CTC) which is nine miles from NSIC.[7] NSIC has remote console equipment consisting of two IBM-2740 typewriter-printers and four IBM-2260 Cathode Ray Tube consoles.

Two types of direct access storage devices are in use at CTC, data cells (IBM-2321 which have a theoretical storage capacity of 400 x 10 characters) and disks (IBM-2311 with a capacity of 7.25 x 10 characters).

The system allows NSIC personnel to search the direct access master files by keywords, keywords and categories, or authors from the typewriter console (IBM-2740) or CRT (IBM-2260). In addition, SDI and retrospective search capability are provided from magnetic tapes which are used for backup.

Since the typewriter terminal is a slow-speed device (15 CPS), the terminal user frequently has the results of the query placed in a temporary file to be printed on a high-speed printer at the computer site and delivered by mail.

Other IBM-2740s are located in the offices of the AEC Regulatory Staff in Bethesda, Maryland, the AEC Division of Technical Information Extension at Oak Ridge, and the ORNL Ecological Sciences Division, Oak Ridge. Thus, the NSIC communications network currently consists of nine remote consoles (six at NSIC) each connected to the IBM-360 computer by leased phone lines. The total CTC network consists of nineteen consoles.

Terminal use at NSIC is by its own staff members for the ultimate user. In fact, searches to be printed on the Bethesda console can be structured on an NSIC console. The inquiry is initiated by a phone call to the Center.

4. Computer Programs

The Center's programs (8) with a brief description of the function each performs follows:

1. 2740 and 2260 File Search Programs
   - A. NSICPRG1 - The purpose of this program is to provide retrospective real-time file searches from the remote terminal. Searching on the basis of keywords or keywords and categories is performed.
   - B. NSICPRG2 - This program supplements NSICPRG1 by providing listings for queries on request.
   - C. NSICIRK (9) - Similar to NSICPRG1 except it functions in the conversational mode. This "dialogue" capability enables the user to manipulate his query in order to obtain a greater degree of relevancy.
   - D. BATCHPRT - Follow-up to NSICPRG2 and NSICIRK. The drops are processed by NSICPRG2 on NSICIRK and stored on a direct access device for overnight listing on the local high speed printer by this program.
   - E. ABSTPRNT - Obtains a listing of an abstract by accession number on the typewriter console. Also prints the citation and keywords along with the abstract.
   - F. AUTHFIND - Provides on-line capability of retrieving document information using the author as the search parameter.
2. 2260 Input and File Maintenance Programs
   - A. HEDRFMAT - Provides a scope format for entering the basic information (header) required for building a new document record.

B. PRCSHEDR - Reads the information which was keyed in the format provided by HEDRFMAT and validates the data before placing it in a temporary storage on a direct access device.

C. KWDSAUTS - Reads and validates the keywords against the authority file, accepts the authors, and places the accumulated information in the temporary file with the header data.

D. PRIMUPTD - This program accepts the remainder of the document, combines it with the previously entered pieces of information, and creates the new record in the direct access master file.

E. CHNGPRG1 - Provides display of all information except the abstract and keywords. It accepts changes to every item except the header and keywords, and it processes deletions for all or part of the item.

F. CHNGPRG2 - Processes header revisions similar to PRCSHEDR except information is handled as revisions rather than new entries.

G. CHNGPRG3 - Accepts additions and deletions to the keyword list of the accessions already on file. The authority file is updated in case of new keywords.

3. Alternate Batch Mode Input, File Backup, and Retrieval Batch Programs

In addition to adding, deleting, or revising records directly from a remote terminal, the transactions are placed into another file to provide input to background batch file maintenance programs. This step is required in order to provide backup should there be a direct access device failure.

A. DALYBKUP - Daily transactions entered via the 2260 console and placed in a temporary direct access file (in addition to being used to maintain the direct access master file) are retrieved and stored on magnetic tape for later use as file backup.

B. KWDUPDTS - This is a batch program to add, respell, replace, and delete terms in the keyword authority file.

C. CARDNPUT - Provides for alternate input via cards of new data items. This program provides a backup capability in case the 2260 console becomes inoperative.

D. MERGNPUT - Sorts and edits 2260 backup data and merges with the alternate card input file.

E. NSICSDI - Provides selective dissemination of information to over 1700 participants using as input the most recent document information entered into the system.

F. RETRO - Provides batch retrospective search capability using coded user profiles and the NSIC document master tape file.

G. BIBLIO - Produces bibliographic reports along with author and keyword indexes.

4. File Conversion Programs

These last two programs convert the master files in order to utilize the present 7090 SDI and Bibliography preparation programs. They were used during the conversion period.

A. VARBLTOU - Converts System/360 master for use on the 7090.
B. RETRET - Although this is a 7090 program, it is listed here because it provides a necessary link in order to make use of the 7090 SDI and Bibliography programs.

5. PPIF Programs

In addition to the regular NSIC programs, three programs were developed for the Program and Project Information File (PPIF). A brief description of their functions follows:

A. PPIFUP - Provides document input and file maintenance capability for the PPIF master file.
B. PPIFSRCH - Selective dissemination of information and retrospective search capabilities are provided using the PPIF document master file as input.
C. PPIFREPT - Produces a cumulative bibliography report with keyword and persons-in-charge indexes.

5. Computerized Operations

A number of services offered by NSIC owe considerable credit for their effectiveness to computer application. Those that will be discussed in this section are file searching, bibliography preparation, SDI, PPIF, and KWIC indexes. Equally outstanding center products and services such as state-of-the-art reports, the journal *Nuclear Safety*, and technical consultation are outside the scope of this paper.

The cost to the information center for some of its computerized services is summarized in Table 4. Costs are given for the IBM-7090 system which was used initially and for the IBM-360 system to which NSIC converted in 1969. Other services where technical man/hours are involved probably average approximately $20 per hour.

Table 4. NSIC Cost Comparison

Function	Unit	IBM-7090	IBM-360	Volume
1. Input	/Document	$ 1.40	$ 2.25	1000/month
2. Search	/Query	25.00	6.00	60/month
3. Topical Bibliography	/Issue	50.00	25.00	4/year
4. SDI	/Abstract/user	0.07	0.04	(1900 × 38) biweekly

5.1 File Searching (3,4)

The computer system gives NSIC personnel the capability of searching the direct access master files by keywords, keywords and categories, or authors from a CRT (IBM-2260) or a typewriter-printer console (IBM-2740). Figure 4 is an example of a query which takes three hours and the results. The first line of the query designated a selected information category, and the second line lists the keywords to be used. The third line is the name of the program which was called to process the search. All the other lines shown in Fig. 1 are the results of the search.

The keyword portion of the statement of a search consists of one or more groups of keywords where the start of each group is indicated either by the symbol "KWDS = " or by the symbol "AND" which represents the logical *and* connector for groups of keywords. Within each group, keyword codes and weights are assigned to each keyword. Each keyword code and weight is separated by commas and the assigned weights are always signed (+ or −) numbers. Within a given group the last combination of keyword codes and assigned weight is followed by the target weight for the group (labeled as "TOT-WT = "). A document is selected as satisfying the requirements within a group if enough of the keywords have been assigned as being descriptive of a document such that the sum of their assigned weights is equal to or greater than the specified target weight.

The terminal user has three options based on the number of documents selected by the search. He can:

1. Ignore the indicated output and structure a tighter query. For instance, the single keyword FUEL HANDLING - 0170 found 560 documents. The number of documents was reduced to 9 by requiring the keyword RADIOCHEMICAL PLANT SAFETY - 1064 to occur also. The latter keyword had been used 189 times by itself.
2. Have the output printed on the typewriter terminal at NSIC (15 cps).
3. Have the results of the query placed in a temporary file to be printed on a high-speed printer at the computer site and delivered by mail the following morning.

Figure 2 contains the balance of the information available in the system on the first two accessions selected in the subject search of Fig. 1.

Although all of the discussion of file searches has assumed the use of a console to an on-line computer system, provisions have been made to run any of the output programs without using the consoles. This may at times be more practical, especially when a large amount of output is desired and/or the time delay is not important.

5.2 Bibliographies[1]

NSIC has used the computer to generate at least three types of bibliographies: periodical, topical, and retrospective. Each will be discussed in turn.

Periodical. The first program developed by NSIC when it computerized in 1965 was one that sorted its accessions that were added to the system during the prior three months into the Centers' subject categories. The abstracts were grouped within each category in order by accession number. Keyword and author index preparation plus page makeup, including the numbering of pages, were handled by the computer so that no editorial makeup was required before publication.

```
cats = 13, e=n
kwds = 0170 + 01, 1064 + 01, tot-wt = 02
$nsicprgl

SEARCHING CATS:
 I3
ACCESSION NUMBER  10104     DATED 1956
UNDERWOOD JW
HAZARDOUS MATERIALS - REDOX PLANT
HANFORD ATOMIC PRODUCTS OPERATION
HW-43319 +.  14 PAGES, 32 REFERENCES, MAY 22, 1956, CFSTI, $3.00 COPY, $0.65 MICROFICHE

ACCESSION NUMBER  11330     DATED 1966
PROVISIONAL OPERATING LICENSE - LICENSE CSF-1
NUCLEAR FUEL SERVICES INC.
83 PAGES, APRIL 27, 1966, DOCKET NO. 50-201, PDR

ACCESSION NUMBER  14340     DATED 1966
ABRAHAM GE + FINNEY BC
CALCULATED MAXIMUM TEMPERATURES OF SPENT YANKEE ATOMIC TYPE POWER REACTOR FUEL DURING SHEAR-LEACH PROCESSING
OAK RIDGE NATIONAL LABORATORY, OAK RIDGE, TENNESSEE
ORNL-3948 +.  81 PAGES, 40 FIGURES, 8 REFERENCES, NOVEMBER 1966
AVAILABILITY - CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION, NATIONAL BUREAU OF STANDARDS, U. S. DEPARTMENT
 OF COMMERCE, SPRINGFIELD, VIRGINIA 22151, $3.00 COPY, $0.65 MICROFICHE

ACCESSION NUMBER  25985     DATED 1968
KING LJ
SAFETY ANALYSIS FOR THE TRANSURANIUM PROCESSING PLANT, BUILDING 7920
OAK RIDGE NATIONAL LABORATORY
ORNL-3954 +.  180 PAGES, FIGURES, TABLES, APRIL 1968
AVAILABILITY - CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION, SPRINGFIELD, VIRGINIA  22151, $3.00 COPY, $0.65
 MICROFICHE

    4    DROPS

#NSICPRGI ENDED. TIME: 00:12.58. (XXXXX)
```

Fig. 1. Typewriter Console Query

```
10104akc
11330akc
$prntabst
ACCESSION NUMBER   10104     DATED 03/[illegible]

THIS REPORT PROVIDES DATA ON THE POTENTIALLY HAZARDOUS CHEMICALS (NONRADIOACTIVE) USED IN THE REDOX PROCESS AT HANFORD.
  THE PURPOSE OF THIS REPORT IS TO ASSEMBLE IN A SINGLE, READILY AVAILABLE DOCUMENT ALL THE INFORMATION ON THE NORMAL
 CONCENTRATION LIMITS AND SAFE HANDLING PROCEDURES NEEDED TO ADEQUATELY CONTROL THE USE AND STORAGE OF THESE MATERIALS.
  THIS REPORT SHOULD BE HELPFUL IN PREPARING SAFETY BULLETINS AND PROVIDE GUIDANCE IN CASE OF UNUSUAL INCIDENTS,
 MALFUNCTIONING OF EQUIPMENT, OR CONTEMPLATED CHANGES IN EITHER THE PROCESS OR EQUIPMENT.  REAGENTS INCLUDE HEXONE,
 HNO3, H2O2, HF, H2, INERT GASES, NITROGEN OXIDES, OZONE, PROPANE, ETC.
CATEGORIES = 01,17,13

00130 - - - EXPLOSION
00331 - - - SAFETY REVIEW
00049 - - - *CHEMICAL TOXICITY
00170 - - - FUEL HANDLING
01064 - - - *RADIOCHEMICAL PLANT SAFETY
00048 - - - CHEMICAL REACTION
01280 - - - HANFORD SITE
01501 - - - *FUEL REPROCESSING
00797 - - - HAZARDS ANALYSIS

ACCESSION NUMBER   11330     DATED 04/01/66

APPENDIX A OF TECHNICAL SPECIFICATIONS LICENSE CSF-1 FOR THE PLANT CONTAINS INFORMATION ON - (1) PLANT SITE, LOCATION
 AND FLOW OF MATERIALS, (2) POSSESSION LIMITS AND FORM OF MATERIALS, (3) SAFETY LIMITS FOR RADIOACTIVE MATERIALS, (4)
 MINIMUM OPERATING REQUIREMENTS, (5) SURVEILLANCE REQUIREMENTS, (6) ADMINISTRATIVE REQUIREMENTS.  SPECIFICATIONS COVER
  IN DETAIL THE HANDLING AND STORAGE OF LIQUID, GASEOUS, AND SOLID RADIOACTIVE MATERIALS.
CATEGORIES = 13

01064 - - - RADIOCHEMICAL PLANT SAFETY
01323 - - - *NFS
00227 - - - *LICENSE STATUS
00438 - - - WASTE DISPOSAL, ATMOSPHERIC
00170 - - - FUEL HANDLING
00440 - - - WASTE TREATMENT, GENERAL
00815 - - - WASTE MANAGEMENT
01371 - - - SOLVENT EXTRACTION PROCESS
00383 - - - FUEL STORAGE
00091 - - - CRITICALITY SAFETY
01501 - - - *FUEL REPROCESSING

#PRNTABST ENDED. TIME: 00:19.62. (XXXXX)
```

Fig. 2. Typewriter Console Abstract and Keywords

The indexed periodic bibliography was issued quarterly for the first three years but then the frequency was increased to bimonthly as the volume of input increased. After two more years of publications, it was discontinued in 1970 as part of a review and reallocation of resources within the Center.

Topical. The same program that produced the periodic bibliographies is used to prepare indexed bibliographies on specialized subjects of interest such as an NSIC category or subset of a category as defined by selected keywords. Bibliographies containing the usual keyword and author indexes have been issued on the following topics:

Transportation and Handling of Radioactive Materials (10)

Effects of Thermal Modifications of Ecological Systems (11)

Seismic Considerations in the Siting of Nuclear Facilities (12)

They usually contain 600-800 references and are reissued when that many new references on the topic have been collected. However, if the subject is a "hot one" the bibliography will be published with fewer references. New topics are selected for publication each year depending on the interest and timeliness of the particular subject.

Retrospective. Special searches for bibliographies to meet a particular need are made of NSIC's master computer reference file at a current rate of about 60 per month. Since each document that is added to our computer file is described by keywords, we are able to retrieve all documents in which a particular keyword or a combination of keywords is used. The searches are usually made on the basis of combinations of keywords, authors, or corporate authors with category or date used as a limiting parameter (delimiter).

Of course, not all inquiries are answered by generating a bibliography. Answers to the inquiries take different forms depending on the type of question asked. Sometimes the reply will be a written discussion of the problem, while at other times it will be a bibliography or a combination of discussion and bibliography. Questions vary from very simple requests that can be answered "off the top of the head" to involved requests that could take days or weeks of technical work. However, since the number of staff members available to answer questions is fixed and since they also perform other duties for the Center (such as the preparation of state-of-the-art reviews and indexing and abstracting), the amount of technical time allotted to any one question cannot be allowed to exceed four hours except in extreme cases. In any event, NSIC does not attempt to solve the user's problems, but to provide information and guidance that will help him go about defining and solving his particular nuclear safety problem.

5.3 Selective Dissemination of Information

In a fashion similar to that used to make retrospective searches, a user's area of interest may be described by keywords to develop a "profile" that is kept in the computer system. (3) Biweekly, in our Selective Dissemination of Information (SDI) program, this profile is compared to the most recent entries to the computer and the abstracts that satisfy the profile requirements are automatically selected and printed on continuous-form 5 x 8-in. cards for the user. Initiated in 1965, there are now over 1900 members of the nuclear community receiving the cards selected according to the particular needs of each. A growth curve is shown in Fig. 3, the company affiliations of the users in Table 5. SDI application forms, based upon keywords, categories, or pre-programmed specialized profiles, are available upon request from NSIC.

Table 5. SDI User Affiliations - December 1970

User Affiliations	Percentage of Total Users
Private Industry	65
U. S. Atomic Energy Commission Staff	10
Government Contractors (other than ORNL)	3
Oak Ridge National Laboratory	4
Universities	8
Federal Government (other than the AEC)	4
State and Local Government	3
Canada	<1
Other	2

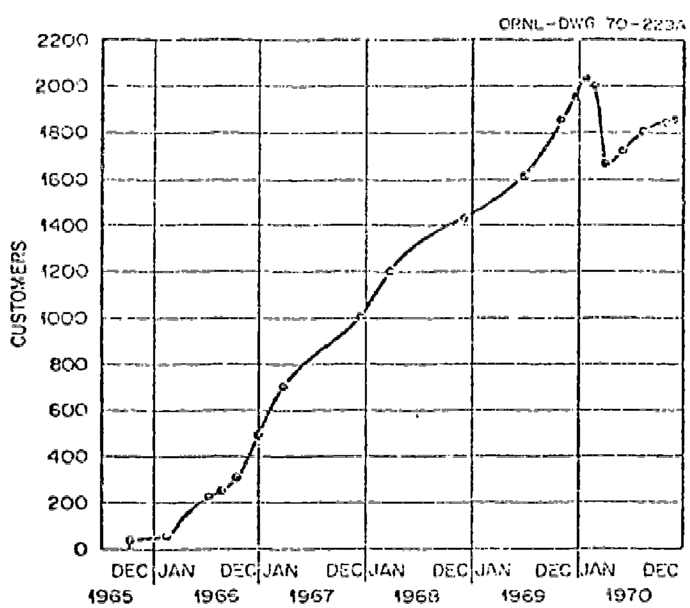


Fig. 3. Growth Curve of SDI Program

5.4 Program and Project Information File

The Program and Project Information File (PPIF) was developed during 1968 for computer storage and retrieval of technical and administrative information on nuclear safety R&D projects.(13) It was initiated at the request of the AEC Division of Reactor Development and Technology to provide RDT and the AEC Regulatory Staff with a means for following current accomplishments on the safety contracts sponsored by RDT. Coverage has now been expanded to include applicable environmental safety programs sponsored by the Division of Biology and Medicine and others.

The system now contains information on 345 contracts which is disseminated on SDI bases to 240 users. The information being stored by the file includes: (a) support group and contractor information, (b) fund and manpower levels, (c) statements of scope and state-of-the-technology, (d) abstracts of the last three progress reports (the oldest one being dropped each time that a new one is added), (e) projection of expected progress for the next reporting period, (f) reports issued, and (g) keyword indexing terms.

Application forms are available on request from NSIC.

5.5 KWIC Indexes

Key-Word-in-Context (KWIC) indexes have come into extensive use by libraries and information centers. NSIC uses them to (1) index annual compilations of national (14) and international nuclear standards, (15) (2) prepare a yearly cumulative index to all articles that have appeared in *Nuclear Safety*, (16) and (3) to supplement keyword indexes compiled with NSIC indexed bibliographies. Their strength lies in the fact that they are relatively inexpensive to generate. Their principal shortcoming is that as an index they are only as good as the titles. If the titles are poor, then the index will be poor—unless the titles have additional indexing terms added.

6. Prognosis

As with any technology, and particularly a new technology which is developing rapidly as is information science, one can project from experience to problems and developments which lie ahead. Inasmuch as our experience has provided us with the insight to do so in several areas related to the use of computers in information centers, we have included these thoughts here for whatever value they may have. The topics considered briefly under each of the following subheadings include hardware developments, software developments, tape exchanges, data (as opposed to bibliographic information) processing, and charges for computerized services.

6.1 Hardware

As an extension to prevailing means for accessing computer stored files of the Nuclear Safety Information Center, now of limited availability via dedicated lines, arrangements will be introduced for direct access from remote terminals coupled to any dial-up public telephone in the United States. Primary user access will be by means of standard Model ASR 33 Teletype devices transmitting in ASCII code at a rate of 10 characters per second. Most subscribers to the network will dial a prescribed Oak Ridge number from teletype terminals already on hand for commercial timesharing services and hence will need to make no capital investment for equipment. Where such equipment is not already on hand, cost of a terminal will run about $700 (plus $30 per month for lease of a Data Set coupler from the AT&T). By using acoustic couplers, terminals can be moved about to any location where a dial telephone is installed. Users affiliated with government-supported agencies will incur no telephone toll charges when calling NSIC through FTS.

It is proposed to provide computer to telephone system coupler-multiplexer equipment in Oak Ridge sufficient to accommodate a family of as many as 50 subscribers. Available to these subscribers will be dialogue selective search routines specifying combinations of keywords and other modifiers equivalent to or superior to those presently in use locally at Oak Ridge. Essentially real time response will be generated by the computer to queries so that users may progressively narrow choices to yield printout of publication abstracts most appropriate to their needs.

To accommodate user preferences, remote access to NSIC can be provided for IBM-2741 typewriter terminals (or equivalents) where transmission will be in BCD code at a rate of 15 characters per second. Similarly, it should be feasible to provide service to a limited number of video-type terminals (CRT's) at transmission rates of up to 150 characters per second still not exceeding the capabilities of inexpensive public dial-up telephone service.

6.2 Software

For the purpose of this discussion, we will consider two types of computer software: (1) the programmer language and (2) the user language. In each case, significant advances are expected in the next few years that will benefit computerized information systems and, in particular, those with direct access capability. The following is foreseen:

1. *Programmer Language* - These programs will become more flexible and generalized. They will be more flexible with regard to the full utilization [illegible] 'ware and in the ability to address direct access storage devices. This will [illegible] that [illegible] only is the equipment being most efficiently used for speed, but also for maximum utilization of storage space.

2. *User Language* - The so-called "conversational languages" have been developed to lead users by the hand and make them feel at ease with computer consoles. However, there are a diversity of languages and approaches that can be quite discouraging to users of several systems. As a result, there exists a great need for standardization of language.

For both types of computer software, but especially the latter, IAC's can anticipate significant advances in technology as well as from standardization. IAC's must exploit their commonness while at the same time maintaining their uniqueness.

6.3 Tape Exchange

The scientific literature is being indexed and stored in computer systems by many large clearinghouses such as *Chemical Abstracts, Nuclear Science Abstracts, Metals Abstracts,* etc., and in very specialized fields by many more specialized information analysis centers. IACs frequently find that information from many of these computer banks falls within its own specialized scope. In order to prevent much wasted manpower in abstracting and indexing, it is very important that there be means of buying or exchanging computer tapes. Already many groups market tapes. One organization, ASIDIC (Association of Scientific Information Dissemination Centers), is made up on companies providing a computer based information service using two or more tape systems prepared by an outside source. The University of Georgia computing center has converted several large data bases to a standard format and has a text search program for searching that format. Through a cooperative agreement arranged by the National Science Foundation, these programs are being provided to ORNL.

Over the years, NSIC has furnished its own computer file to other groups such as Lawrence Radiation Laboratories (LRL) Technical Information Division, Battelle Northwest Laboratory, and NASA's Lewis Research Center. Recently NSIC intensified its coverage in areas related to the environmental effects of nuclear power. As part of its expanded coverage, we obtained a file of references on radionuclide behavior in the environment from LRL's Biomedical Division.

Several steps were required to manipulate the LRL information so that it could be incorporated into the NSIC computer bank. First it was necessary to convert the files created by their CDC-6600 into a format recognized by the IBM-360. Their tape consisted of three files that were of interest to NSIC. They were: (1) the journal authority made up of journal titles and codes, (2) a title-keyword file made up of all document titles plus keywords that did not appear in the titles, and (3) a collective file consisting of report numbers, page numbers, authors, journal codes, etc. File conversion consisted of bringing the information together on to one file in the NSIC format. Keywords were correlated with the NSIC thesaurus and references were assigned to the NSIC categories. Altogether, 5000 references were added to the NSIC system with those references being screened out that were duplicates or that were not relevant. After conversion, the data items were processed into a file structure identical to the NSIC file except that there were no abstracts with these references.

The converted LRL file is now searched as an integral part of the NSIC file and as such provides information that would have required many man-years of work to duplicate. In a similar manner, any tape can be incorporated into another system. Although the necessity of such conversion operations raises the question of standardized tape formats, once a conversion program has been developed, it can be used on all tapes from the same source. This is the pragmatic solution to the alternative of standardization which may some day be achieved.

6.4 Data Processing

It may seem ironic that data processing on a computer is regarded here as a special case of information processing, since computers are essentially numerical devices. However, the data we have reference to here are not solutions to equations but raw data as compiled from numerous observations. Such data not only saturates science and technology today but is the keystone upon which scientific deduction and technological advances thrive. The difficulty with the computerization of such data lies not in its exploitation in this manner, but primarily that so much data exists that reasonable care must be exercised in the selection of such data for storage, both to avoid overloading operating computational systems as well as to avoid duplication of storage files which may exist elsewhere. Let me hasten to add that whether or not this is a problem, and if so, its magnitude varies widely from center to center. Highly specialized, narrow, discipline oriented data centers will have the greatest need for such data storage as indeed that stored data and the information derived therefrom may be the very basis for some centers existence. On the other extreme, broad-based, multi-disciplinary, mission-oriented information centers can generally ill afford to dilute their activities by computerizing all the data in which they have an interest. This position will seem heretical to those whose experience is limited to the former type of center, but it is a reality for the latter.

There are, of course, exceptions to the above generalization. The principal exception would occur when a specific need arises for selected data and where this need is sufficient to justify the effort. Such was the case in our experience when the AEC directed us to establish and maintain a computerized file of reactor design data, operating characteristics, safety parameters, etc. This we have done, together with the specific retrieval and utilization program to provide the information in a form as desired by the AEC. This project, designated CHORD-S, utilizes the same hardware as NSIC, but has its own data files and software. It is fully described in a report published last year.([17])

6.5 Service Charges

The idea of an information (or data) center charging its users for services rendered has been widely discussed. ([18], [19]) The idea has its proponents, as well as its opponents, and each - depending upon their particular situation - can speak with equal justification. The proponents argue that at least those users outside the immediate program of the funding agency should pay their share of the cost and the opponents argue that any charges will curtail useful information flow which government agencies can well afford to sustain for its long term benefits. I believe that both of these positions (and there are more points than those outlined above to be made on both sides) have merit, and that each situation needs to be evaluated on its own. However, one point seems clear from the inherent nature of information centers themselves; namely, that the amount of money that one could hope to recover from users outside the program of the funding agency is generally a relatively small fraction of the total cost. I say that this is inherently true, since the obvious motivation for the establishment of an information center arises from the information needs of a particular organization, society, department, or agency. Furthermore, the sponsoring group - for government information centers at least - generally hold some "charter" for technology in a given area. The government groups therefore have the responsibility for serving the needs of most of the users - be it in support of research as DOD or AEC or for use by the public as well as NOAA or NIH. Hence the market outside this behooven clientele is generally rather limited - but in many instances such a market does exist.

Aside from the added income that might be derived by charging "outsiders" for services, the principal advantages of such an arrangement is that (1) charges automatically weed out the idly curious (who are not always easy to identify), (2) it forces those who would use the information for their own commercial exploitation to be more considerate in the use of the services available, and (3) it spares the center management from having to make seemingly arbitrary decisions on limited information as to how much time and effort, if any, should be allocated in response to a particular request.

The advantages of charges, noted above, are not without certain disadvantages which, in addition to division of clientele into "free" and "charge" classifications, the most troublesome problem seems to be that of bookkeeping. Here, however, the utilization of the computer for the "charged" services - as in the generation of bibliographies or routine SDI - provides a convenient cost basis around which to establish a service charge.

7. Conclusions

Despite the many differences between Information Analysis Centers (IACs), it is difficult to imagine an effective center of any size which has not automated its information store. In specialized applications this automation may not be computerized, although in general the computer offers the greatest flexibility. NSIC storage files have been computerized since shortly after its inception. Since that time numerous retrieval as well as file management programs have been prepared and are routinely employed. The largest single expense associated with a center's operation is directly attributable to the acquisition evaluation and formatting of relevant information. With this investment as an unavoidable minimum, it behooves the center to manage its operation so as to derive the greatest value from this investment. The computerized information store is readily amenable to such exploitation and NSIC's information store is routinely searched for SDI and bibliographies, not to mention the special searches which are presently being conducted at the rate of over 60 per month. In fact, it is conservatively estimated that some reports which were accessed in NSIC five years ago have since been in the subject field of over 1000 searches and included in over 100 printouts. No other practical searching mechanism exists for reviewing a large collection (NSIC has over 50,000 documents) for specific information and obtaining it so quickly at such a low cost.

A relatively recent development in computerized information technology has been the use of remote consoles connected to a central computer via telephone lines. The computer center serving NSIC now supports 19 remote console on a general purpose IBM-360 Model 50/75 computer. Present limitations regarding the type and numbers of such consoles which can be supported, as well as the type of service (dedicated vs dial-up phone lines), are believed to be temporary as the technology in this area is developing rapidly.

NSIC's experience in providing several - but by no means all - services directly from its computerized information store has been very gratifying. We anticipate that future improvements in hardware and software will further enhance this part of our function. Such services supplement - but can never replace - the technical assessments classically associated with an IAC and as reflected in state-of-the-art reports, journal articles, consultations, etc.

We anticipate that exchanges of tapes between centers will become more common, although differences in format will continue to be a problem for some time to come. Charges to some users for some services will become more common. Although in most instances it will not be a decisive matter in the center's overall operation, it will be a useful device for responsible center management. Here again, the computerized information products will be a principal ingredient in the product as well as a convenient basis upon which to hang charges.

In conclusion, NSIC's experience has shown the computer to be an essential element in the operation of an effective information program and it would appear that this dependence would become even greater in the future.

References

1. J. R. Buchanan and Wm. B. Cottrell, **A Summary of NSIC Activities, 1963-1967,** USAEC Report ORNL-NSIC-46, Oak RidgeNational Laboratory, September 1968.
2. J. R. Buchanan and Wm. B. Cottrell, **Summary of Environmental Information Activities of the Nuclear Safety Information Center,** USAEC, Report ORNL-NSIC-79 (also ORNL-TM-3009), Oak Ridge National Laboratory, June 1970.
3. J. R. Buchanan and F. C. Hutton, "Analysis and Automated Handling of Technical Information at the Nuclear Safety Information Center," **American Documentation Institute,** Vol. 18, No. 4 (October 1967)
4. J. R. Buchanan and E. M. Kidd, "Development of a Computer System with Console Capability for the Nuclear Safety Information Center," **Proceedings of the American Society for Information Science,** Vol. 6 (1969).
5. H. B. Whetsel, **Guidelines for Reviewers and the Editor at the Nuclear Safety Information Center,** USAEC Report ORNL-NSIC-47, Oak Ridge National Laboratory, January 1970.
6. Celia Murphy and Carol Julian, **Information Scanning and Processing at the Nuclear Safety Information Center,** USAEC Report ORNL-NSIC-48, Oak Ridge National Laboratory (to be published).
7. Wm. B. Cottrell, J. R. Buchanan, and D. W. Cardwell, **The Operation of a Specialized Scientific Information and Data Analysis Center with Computer Base and Associated Communications Network,** USAEC Report ORNL-TM-3078, Oak Ridge National Laboratory, August 1970.
8. B. H. Stoutt and S. L. Yount, **Console-Oriented Information Storage and Retireval System for the Nuclear Safety Information Center,** USAEC Report CTC-37, Computing Technology Center, Oak Ridge Gaseous Diffusion Plant, Sept. 4, 1970.
9. J. H. Owings and C. E. Price, **CIRK - CTC's Information Retrieval from Keywords,** USAEC Report CTC-36, Computing Technology Center, Oak Ridge Gaseous Diffusion Plant, Oct. 1, 1970.
10. R. D. Seagren and L. B. Shappert, **Indexed Bibliography on Transportation and Handling of Radioactive Materials-2,** USAEC Report ORNL-NSIC-84, Oak Ridge National Laboratory, January 1971.
11. J. G. Morgan and J. K. Franzreb, **Indexed Bibliography of Thermal Effects Literature-1** USAEC Report ORNL-NSIC-81, Oak Ridge National Laboratory, December 1970.
12. W. H. McClain and O. H. Myers, **Indexed Bibliography of Literature on Seismic Considerations in the Siting of Nuclear Facilities-1,** USAEC Report ORNL-NSIC-88, Oak Ridge National Laboratory, June 1971.
13. B. H. Stout and S. L. Yount, **Program and Project Information File - User Documentation,** USAEC Report CTC-INF-1005, Computing Technology Center, Oak Ridge Gaseous Diffusion Plant, 1970.
14. J. Paul Blakely, **Compilation of U. S. Nuclear Standards, 7th Edition, 1970,** USAEC Report ORNL-NSIC-85, Oak Ridge National Laboratory (to be published).
15. J. Paul Blakely, **Compilation of National and International Nuclear Standards (Excluding U.S. Activities), 6th Edition, 1970,** USAEC Report ORNL-NSIC-78, Oak Ridge National Laboratory, October 1970.

16. Nuclear Safety Staff, **Index to Nuclear Safety, A Technical Progress Review by Chronology, Premuted Title, and Author,** Vol. 1, No. 1 through Vol. 11, No. 6, USAEC Report ORNL-NSIC-86, Oak Ridge National Laboratory (to be published).
17. T. E. Cole and F. A. Heddleson, **Computer Handling of Reactor Data for Safety - CHORD-S,** USAEC Report ORNL-TM-2928, Oak Ridge National Laboratory, August 1970.
18. **Proceedings of the Forum of Federally Supported Information Analysis Centers, Nov. 7-8, 1967,** sponsored by Panel No. 6, Information Analysis and Data Centers, Committee on Scientific and Technical Information, Federal Council for Science and Technology, PB 177-051.
19. **Proceedings of the First Ad Hoc Forum of Scientific and Technical Information Center Managers, Directors, and Professional Analysts, Nov. 9-11, 1965,** USAEC Report CONF-651131.

TRENDS IN ABSTRACTING AND INDEXING SERVICES

Burton W. Adkinson
American Geographical Society

Introduction

Someone has said that those who do not know history must repeat its mistakes. I suspect the person who asked me to talk today believed this statement and hoped that I had been engaged long enough in the information business that I would have some perspective on the development of abstracting and indexing services.

In the next few minutes, I wish to briefly discuss the following trends:

1) The growth in the quantity of material abstracted and indexed.
2) The expansion in topical coverage by abstracting and indexing services.
3) The increasing involvement of the Federal Government.
4) The introduction of new technology.
5) The increase in diversification and in specialization.
6) The growing awareness of the need for standards and conventions.
7) The increased realization that cooperation and coordination are a necessity for survival.
8) The increased demands for information brokers.

There could be other trends considered, or the trends might be organized under other headings. An audience as sophisticated as this one need not have a speaker outline the characteristics of these trends. Rather, this discussant wishes to point out that there are many interrelations among these various trends. Most of the changes have been in response to the user communities who constantly demand the quick delivery of reliable, pertinent information that is organized so it can be used efficiently and with some degree of confidence.

Growth

The growth of scientific and technical information in its many different forms and formats has been the subject of many papers over the past two centuries. The increasing magnitude of scientific and technical information and the increased number of users has been identified by authors as a principal reason for the development of the printing press, the scientific and technical journals, the abstracting and indexing services, information analysis centers, specialized bibliographies, catalog cards, technical reports, information processing technology and even the "invisible colleges."

The speaker would suggest that growth both in the quantities of scientific and technical information and, in the number and variety of users, has been a major influence over the past 30 years on the character, size, form, techniques and sponsorship of abstracting and indexing services. These growth trends will continue in the future even though their pace may be slowed. Abstracting and indexing services, in fact all primary and secondary services, are plagued with problems of magnitude and growth of information and will continue to be so in the future.

In the discussion of the other trends I outlined at the beginning of this paper, one should be aware that the growth factor is a strong influence.

Coverage

During this century, there has been an increasing trend for some scientists in each field to extend their investigations into topics generally regarded as belonging to a sister discipline. Thus, one hears of biochemistry, medical engineering, polymer chemistry, physical oceanography, genetics, nuclear science, etc. In addition, the mission oriented research and development programs have broadened, as well as becoming more complex. The abstracting and indexing services responded to these trends by expanding topical coverage. During the 1940's to 1964, the abstracting and indexing services increasingly duplicated topical coverage in the overlap areas. It was not until 1968 that the abstracting and indexing services began to realize that growth in literature and expansion in topical coverage were rapidly over-taxing their systems. Realistic solutions to these problems have not yet been achieved, but efforts at the national and international levels are oriented toward achieving rational cooperative and coordinated plans.

Federal Government

One of the significant trends over the past 1/3 of a century has been the increasing involvement of the Federal Government in scientific and technical information and especially with abstracting and indexing. This involvement has been in the following ways:

a) Increase in number of abstracting and indexing services provided by the government on such topics as technical reports, nuclear science, water resources, cold regions, aerospace, and educational research.
b) Subsidy for improvements and expansion of non-profit services in many scientific and technical fields.
c) Fostering cooperation, coordination and standardization among the abstracting and indexing services; Z39 support; NFSAIS; and the Council of Biological Editors.

One can anticipate more and not less involvement of the Federal Government.

New Technology

The abstracting and indexing services, faced by over-taxed manual techniques, were among the first to take advantage of new information processing technologies. The emphasis among these services has been on replacing manual operations with automated procedures. Few new innovations have been introduced but one can begin to identify trends that should markedly change the character of these services in the next few years. In addition, coordination among these services, as well as interrelations with other components of the information industry, are beginning to develop. For example, services, because of the new technology, are now able to produce specialized issuances of interest to smaller user groups, which enables the services to depend less on the publication of massive abstracting and indexing periodicals to reach their users. Many services are developing cooperative services with other organizations. In addition, these services are experimenting with service charges based on use rather than on the sale of complete files.

Increased flexibility, permitted by the new technologies, will in the future markedly change the character of these services.

Diversification and Specialization

A phenomenon of the 1960's has been the rapid diversification of products produced by the large comprehensive services. For example, *Chemical Abstracts* in 1960, had a few information products. Today, *Chemical Abstracts* offers about 30 different services and products. This diversification is also a characteristic of MEDLARS, BIOSIS, *Engineering Index*, and many others.

On the other hand, scores of specialized abstracting and indexing services have appeared over the past 15 years, such as Oral Research, Tobacco, Urban, Photographic Science and Engineering, Information Science and numerous others.

One could question whether this diversification and specialization can continue at the same rate as in the past. If the SATCOM Committee's analysis is correct that services to users' groups of 1,000 to 2,000 persons are needed, then these trends will continue.

Standards and Conventions

Over the years there has been a constant effort to improve technical and substantive standards by abstracting and indexing services. Until the last five years, the major emphasis has been on upgrading quality and technical standards by individual or small groups of abstracting and indexing services.

With the advent of new technology, that required consistency in application of techniques, abstracting and indexing services have become aware of the need for industry wide technical standards or conventions. Little attention has been given to the requirement for compatibility of the intellectual organization patterns among most secondary services and coordination is almost non-existent among primary producers. This is an area where increased attention will be necessary if abstracting and indexing services are to respond to requirements of problem-oriented research and development projects.

It is the speaker's opinion that increased emphasis will be placed upon development of standards and that the Federal Government will be forced to take a leading role in this endeavor. Certainly, some standardization among the major governmental abstracting and indexing services must be achieved before the Federal Government can insist that non-federal abstracting and indexing services adopt common technical standards as a prerequisite of Federal support. If this need is great among abstracting and indexing services, it is even greater among information analysis centers.

Information Brokers

There has been an increasing demand for information from two or more scientific fields to be organized so that it could be used for problem solving. Within the Federal Government, some of the large abstracting and indexing services and many of the information analysis centers were initiated in response to this need. Outside the Federal Government, numerous specialized services have been developed to meet this growing demand. In most instances, each of these specialized services have had to redo the bibliographic, abstracting and other tasks that have already been performed by libraries, and abstracting and indexing services.

The introduction of new technology and the adoption of common technical standards should facilitate initiation of services that operate in the information field somewhat like an investment broker operates in the financial field. The broker knows and evaluates investment possibilities and aids his clients to build a portfolio to meet his clients' needs.

The initiation of demand and continuing bibliographies are responses to this same need by scientists and technologists. The experimental on-line query services are another example. It is the speaker's contention that this type of information service will have an expanding market in the future.

If abstracting and indexing services can interact with information analysis centers to refine this type of activity, the quality, efficiency, and effectiveness of information services would be upgraded both as to currency and relevancy.

Conclusion

In summary, I have tried to make the following points:

1) Growth of scientific and technical literature has been, and will continue to be, the major factor with which abstracting and indexing services will have to contend. In addition, the increased numbers of scientists and engineers with greater variety of interests further complicate the production and marketing problems of these secondary services
2) There has been, and will continue to be, increasing pressure on all information services but particularly on the secondary services to develop information packages that are responsive to the needs of scientists and engineers who are working on inter-discipline or multi-discipline problems.
3) In order to meet the above demands, the abstracting and indexing services must accelerate the adoption of:
   a) More innovative use of new technology to increase flexibility and to deliver more marketable products.
   b) Closer working relationships with other components of the information industry, who are in a position to modify or supplement the products of the abstracting and indexing services; such as, commercial information services, libraries, information analysis centers, as well as information using organizations.
   c) More realistic working arrangements which must be developed to allocate responsibility for topical coverage among abstracting and indexing services.
4) If the above goals are to be achieved, the abstracting and indexing services as well as other information services must rapidly adopt common standards for:

   1) bibliographic information
   2) format specification
   3) systems configurations.

5) The accomplishment of the above can be greatly accelerated if:
   a) The Federal Government services will take the initiative to develop the above relationships and standards among Federal abstracting and indexing services, and with other government information components.
   b) The Federal Government will act as the catalyst to aid in adoption of areas of responsibility, common standards and systems design that will allow for easier systems inter-connections.
6) Finally, one can identify many trends among the abstracting and indexing services that are oriented to achieving the production of more useful information tools.

The question is how can these be accelerated and directed toward a better integrated network of systems that will improve the delivery of information to the scientists and engineers.

INFORMATION ANALYSIS CENTERS - DoD POLICY ON COST RECOVERY

W.C. Christensen
Director of Technical Information
Office of the Director of Defense Research and Engineering

The Department of Defense (DoD) policy regarding service charges for products from selected information analysis centers can be stated in two sentences. Centers will charge for their services where the users can be reasonably expected to pay. Further, those centers which are not producing revenue at least equal to 50 percent of their operating costs by the end of the fiscal year 1972 contract period will be reviewed and terminated if appropriate.

This policy grew out of increasing concern that the benefits from some of the DoD information analysis centers were not commensurate with their cost - a concern which has been sharpened by the increasing necessity to obtain the maximum return on the defense dollar.

The Department of Defense has been a strong supporter of the information analysis center concept and in fact instituted procedures in 1964 to give these centers special recognition and attention. In the intervening years, the centers have continued to operate within their assigned areas which periodically change in line with changing needs of DoD. However, in recent years the tightening Defense budget resulted in reductions in some of the centers' budgets. The major reason for these reductions was a reluctance on the part of the Service who sponsors the information analysis centers to fully fund activities which support other Services and government agencies when they are not given enough funds to completely carry out their specific, Service related missions. This situation came to a head during the preparation of the Fiscal Year 1972 budget.

In addition to the budget, another related problem was developing. In 1968 the Centers were directed to initiate a system of service charges for their products by 1 July 1969. Due to a variety of legal, contractual, and other problems many of the centers were unable to comply with this directive. I might add that the centers were less than enthusiastic about the service charge concept for reasons which I am sure Mr. Veazie will cover in the next paper.

It was in this environment that a decision was made in November of last year to adopt the policy stated in the beginning of my paper. To circumvent the problem of each individual Service budgeting for the total DoD contribution for their assigned information analysis centers and to minimize the previously mentioned problems encountered by some of the centers in instituting service charges, we transferred administration responsibility for nine of the contractor-operated centers from the Navy and the Air Force to the Defense Supply Agency. This Agency is also responsible for the operation of the Defense Documentation Center. However, these centers will not be placed under the Defense Documentation Center. Instead, the technical monitorship of these centers will continue to be provided by the Navy and the Air Force as in the past.

While all the details have not yet been worked out, there are a variety of methods which may be used to collect and return service charges to the centers such as the use of the National Technical Information Service, direct use of commercial publishers, and direct billings by the center. Each center will be permitted to employ the mechanism best suited to its particular situation.

This then, in extremely capsulated form, is the DoD policy, some of its genesis, and the current status. However to amplify slightly and hopefully to increase your understanding of the DoD position, I would like to briefly present my views on the service charge concept. First, we repeatedly claim that technical information is a very valuable resource and that our technical information activities produce great benefits - yet to varying degrees we have failed to convince the people who control our resources that this is indeed true. Their quite logical response is - "OK, if it's so great, then the users certainly should be willing to foot the bill." Frequently our response is - "Ah yes, but unfortunately the user and the people who control his resources do not realize how valuable our services are - and besides, they are not accustomed to paying for information services." To me, the message is clear - if we are to maintain viable information activities, we must do a better job of establishing the benefits of these activities. Service charges are one mechanism of establishing benefits which is clearly understandable to the people who have to make resource decisions for technical information activities. However, there are certain impacts which may, at least for an interim period, reduce the utilization of information activities. We will closely watch the various effects of information analysis centers and take appropriate action where possible to counter any trends which are clearly having an adverse effect on the DoD R&D program.

As a final point on the institution of service charges at selected DoD information analysis centers, the DoD has wanted to open several centers to the general public in the interest of increased technology transfer. Under the existing DoD mission constraints and the Office of Management and Budget policy, we must recover the cost of providing these services to the general public. Unless we are able to establish a workable service charge system, we will be unable to permit public use of these information resources.

DoD POLICY ON COST RECOVERY AS VIEWED FROM AN INFORMATION ANALYSIS CENTER

Walter H. Veazie, Jr.
*Head, Electronic Properties Information Center
Hughes Aircraft Co.

There are presently objections and confusion about how to implement the Department of Defense (DoD) cost recovery policy for Information Analysis Centers (IAC). Basically, these can be categorized as:

1. What do we charge for?
2. How do we recover costs?
3. What happens to the "purity" of the IACs when they are "tainted by commercialism?"

In considering what we charge for, confusion exists because of the lack of a national policy on service charges. Each IAC has independently developed its own system of service charges. . . .with some charging for formal publications; others, for data retrieval; and still others for all products and services.

This basic conflict, "What to charge *for*, " is compounded by the problem of *what* to charge?

If we sell research..."reactive" technical answers and literature searches...., then we charge for hours and, possibly, computer time. If we sell products... state-of-the-art reviews, data compilations, and bibliographies....then we must charge on a different accounting system. The topic of marketing IAC services and products is covered in greater detail in a report published by ERIC* in June of this year.

The confusion attendant on cost recovery relates directly to the first problem. Except that here, even if we assume a national policy that permits us to charge for the same things on the same basis, we must decide *how* to sell our service or product and recover our costs. We have tried several approaches. These have included direct IAC-to-user sale of publications and services, promotion and sale by a commercial publisher or society, and publication by the National Technical Information Service (NTIS). We have collected cash from users, obtained publisher royalties, and adjusted to government industrial fund transfers.

None of our product sales or collection techniques have, however, provided the fifty percent recovery required by DoD! Moreover, our separate approaches to sale of products and services do not appear to provide for such recovery. So we see, our confusions have been not only how to charge and what to charge for; but how to collect enough money to reach the fifty percent goal.

The *objection* of the IAC's to the entire question of service charges is not idle or recalcitrant. We object because service charges impact on our understanding of the IAC mission. My fear is that service charges will reduce the number of users of our information and our research and analysis capabilities. In essence, that service charges will destroy the IACs as a national resource.

Because of my concern, I am first going to discuss more fully why DoD's policy for service charges is objectionable to IAC managers. Then, I will review how we have tried, nonetheless, to implement this policy.

I am then going to turn from information specialist to businessman, and show that there is an effective means for implementing DoD's policy without destroying the IAC's! We can charge without destroying what we know must be done, and products can be developed that will return the required revenue.

IAC CONCERN FOR SERVICE CHARGES

Let's look, first, at our problems and objections.

Each center has built its close relationship with its users on a free exchange principle. Together, we have developed a network which has been invaluable in providing DoD with state-of-the-art information and data. How can EPIC, or any IAC, request "free" input data and then turn around and sell it back to the user? Such a charge may force us to break up the network.

For example, EPIC has sampled its users to determine their reaction to service charges for various outputs. Responses by 56 industrial, 14 government, and 23 university associated users are tabulated in Table 1. These data indicate that our users, at least, will be more selective in their use of various Center outputs. Yet this group of respondents also have reported *time and cost savings of over $50,000* from EPIC's output. They, therefore, of all people, should recognize the value of our publications and technical inquiry answering service, but they will not be as eager for our service when we produce a bill.

The spiraling effect from reduced customer usage upon service charges is part of our fear and objection. It could destroy the IAC concept and its usefulness as a national resource. For example, EPIC currently has approximately 2,500 users. For any one "free" interim report, we would expect to distribute 350 over a one year period. Our survey shows that we are possibly going to lose half of this clientele when we start charging. Thus, whereas we might have been able to recover costs by selling this report at $5.00 per copy, we now have to charge $10.00 per copy to reap the same income. But, not all of our remaining customers will remain at $10.00. So, the next interim report will have to have a higher price to make up for the reduced circulation. Finally, we will have only one customer for our report; the one who is willing to pay $1,750.

TABLE 1

SURVEY OF ANTICIPATED EPIC UTILIZATION
AFTER SERVICE CHARGE INITIATION

Influence Of Service Charges*

EPIC Output	Increased Use	No Effect On Use	Selective Use	Reduced Use	Prevent Use
Formal report	-	24.5%	43.3%	25.5%	7.7%
Interim report	-	18.1	45.1	28.5	8.3
Reactive literature searches	-	21.7	48.3	15.0	15.0
Reactive technical inquiry service	$1.5%	21.8	48.5	15.7	12.5

*Based on 93 responses

As costs go up, our concern is "How do we stay in business when our users (including the government) think they can do analysis of the literature, or even duplicate once-reported studies, cheaper by themselves?" This appears to be an inevitable problem of service charges, particularly with tight user budgets.

I am also concerned with the time lag between compilation of data and its availability to the user. EPIC under a "free" distribution-to-authorized-users system was able to disseminate formal data tables within one month after Air Force release for public dissemination. Via a commercial publisher, that time is now increased to six months! I recognize that the time required for publication release may not always be a major factor in the usefulness of distilled and printed information, but time is critical when answering telephone inquiries. These users want their answers in real-time. They will not take the time or make the effort to process the paper work required to purchase such services. Particularly, when the answer to their request might be, "We don't have any data."

HOW DO WE RECOVER COST?

Our initial efforts at cost recovery were thwarted by legal and administrative problems. These have been resolved one by one, but not without the expenditure of time and other resources. Once resolved, various routes to cost recovery were tried.

EPIC, **TPRC, DMIC and others utilized commercial publishers. DCIC utilized a society as agent for its "Engineering Properties of Ceramics" handbook. AFMDC, MPDC, and RAC sold handbooks, reports, and services directly to their users. PLASTEC used the NTIS as its sales agent.

Because I am most familiar with EPIC/Hughes Aircraft Company, let's look at our cost recovery efforts. We recover royalty on sales of "Electronic Properties of Materials - A Guide to the Literature" and "Handbook of Electronic Materials" — from a commercial publisher. Sales figures from our publisher (as of March 31, 1971) indicate that 1,360 copies of the Guide have been sold at $150.00 per volume.

707 copies of the Handbook have been sold at $10.00 per volume. We are particularly encouraged by the volume of Handbook sales. The Handbook became available in January 1971 and sales have been achieved with only a brief announcement in the "EPIC Bulletin." Soon, our publisher will distribute over a hundred thousand descriptive fliers on this publication and a press release to all book review editors of appropriate journals and magazines. We will see what effects this level and type of promotion has.

Under our current cost recovery effort, we anticipate the recovery of only five percent of our Fiscal 1971 operating budget. If, we utilize NTIS for the dissemination of our interim reports, which we plan to do, we will increase our recovery to ten-fifteen percent in Fiscal 1972 from the sale of publications.

Moreover, earlier this year, it was learned that cost recovery rates of one to twenty percent were anticipated by other IACs.

Thus, even as sales mount, to my knowledge none of our efforts have produced significant revenue from the sale of publications and services; none have yielded the fifty percent required by the DoD directive. We have all tried and will keep trying, but it does not appear that we will reach the goal by Fiscal 1973 without some major change in our marketing methods.

Now I am going to make a 90° turn. I am going to present a plan which I am convinced will implement the DoD policy and save the IACs as a national resource.

**Acronyms identified at the end of the paper.

DoD-IAC COST RECOVERY

Budget limitations, priorities for hardware procurement, and Section 203 of Public Law 91-121, the military procurement authorization for 1970, have necessitated that DoD look at the IACs strictly as a competitive activity within the business of buying the mission elements vital to national defense. To the IAC's, the competition means providing more benefit from the application of DoD funds to technical and scientific information and data, after application by a user, than could otherwise be purchased by DoD on a research or development contract. Thus, for the IAC managers, to make DoD's cost recovery policy work, we must adapt the basic principles of marketing and organization germane to a profit oriented business venture. In terms of marketing: "People buy benefits, not products!"

- No one buys F-15's, nor, for that matter, F-111's or Minutemen missiles. They buy only the capability to satisfy a mission element vital to the nation's defense. Similarly, no one buys information, only the usefulness of that information as it is packaged to satisfy a need vital to a mission element or discipline.

In terms of organization: function dictates organization!

- Every product-generating organization must contain three line operations: A design/engineering group, a manufacturing group, and a marketing/distribution group. For the DoD/IAC organizations this means the engineer-scientist analysis group, a publications and packaging group, and a marketing/distribution group.

At this time, and until we have prepared the ground for germinating this seed as it relates to the organization of the separate and confederate IAC's, let me state that our present charter and temperament provides only for one and one-half of these organizational blocks.

Now, let us look at a simplified version of the IAC objectives:

1. Obtain, codify
2. Store, analyze and publish
3. Disseminate

Publication, as we use this term, includes the retrieval and output functions, as well as actual typing, typesetting, printing, or any other form of duplication. Moreover, the output portion of this function itself divides into:

Reactive	Simple inquiry answering and bibliographic services
Determinant	Data compilations, state-of-the-art and survey reports, handbooks, etc. which are directed by the IAC mission or discipline.

Let us now look at DoD's cost recovery policy, and relate it to these objectives. DoD's cost recovery policy, it appears to us, is:

1. Recovery of *direct* costs to the government
2. Provisoning of added "direct-dollars" for improved or additional IAC efforts.
3. Achievement of self-support (direct plus indirect cost) thru sale of improved services.

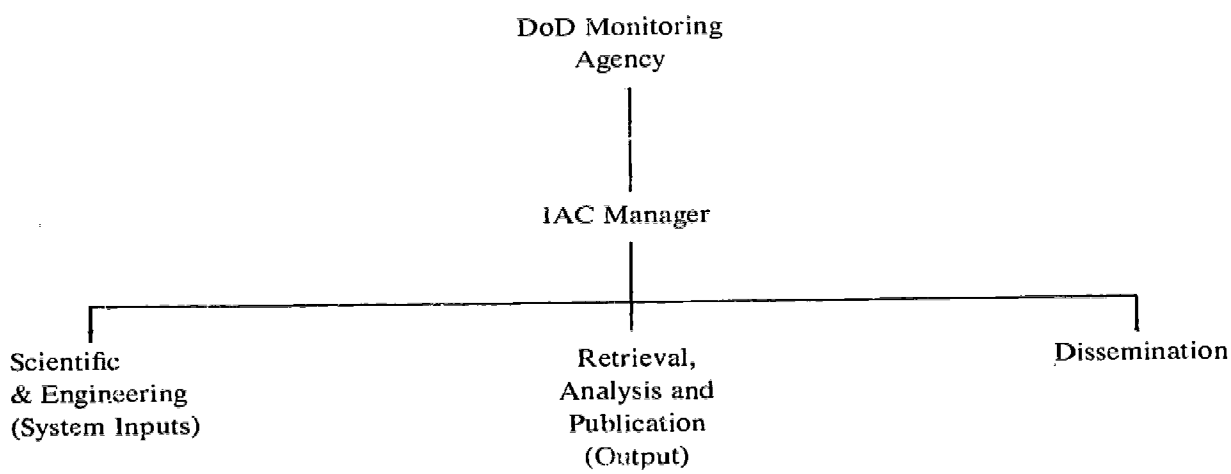
We interpret the first of DoD's policy objective to mean the recovery of the *direct cost* of analysis, compilation, packaging, and dissemination as these relate to the *determinant output*, or "formal" documents. The revenue derived from this output will then be "reinvested" in the IACs to provide for improved "formal" products and services. Because these new products will then become more attractive to the user (e.g. they are offered in format, are packaged for his specific needs or interests, or are more convenient to access) it is anticipated their sale will generate additional revenue.

These new monies would enable the IACs to begin to cover their indirect costs; namely obtaining and codifying information and data, and answering "reactive" publication requirements.

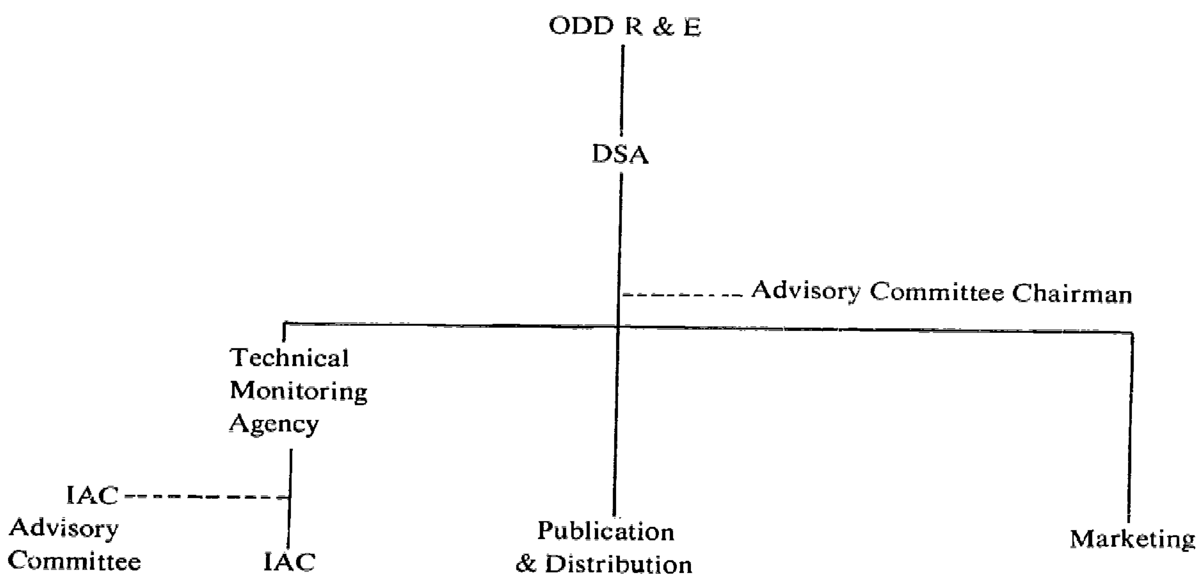
Unfortunately, today's IACs are not chartered, funded, or organized to fully implement the business rationale implicit within the DoD cost recovery program. And, I will show, even though contracts are being modified to correct the "Charter" portion of this problem from a business sense, organizational assistance from DoD is mandatory if full implementation is to be achieved.

ORGANIZATION

The following generalized organization chart indicates the current DoD-IAC relationship and functions:



This organization had no problem differentiating between reactive and determinant publications when no charges were required. We were organized for level-of-effort internal activity and provided level-of-effort service without any need to distribute cost into product accounts. Today, however, level-of-effort accounting must allow cost distribution so that charges can be levied directly to products in order to establish selling prices. This means we must organize functionally—if not on a hierarchal chart, then at least through cost-accumulation codes. Either way, we must organize to take advantage of the functional specializations we would be paying for. Our marketing specialist has suggested the following organization:



Under such a system the IAC's would continue to routinely provide for the functions of acquisition, codifying, storage, and analysis as applied to the provision of quick response inquiries and short bibliographies.

The packaging and distribution of reports would be handled by NTIS or similar agency. The accounting for sales would also be handled by this organization.

Marketing and distribution would be handled by a government agency, society, or commercial organization. This organization would be an *integral* part of the IAC network and have responsibility for market research, determining packaging formats, special reports, advertising, and promotion. The organization would have a key role to play in marketing discipline or mission oriented reports to government agencies and product oriented reports to industry.

The primary difference in this organizational structure from previously proposed decentralized marketing functions is the inclusion of the marketing group as an *integral* part of the DoD-IAC network.

Two major advantages of this organization are that the IACs retain their scientific and technical response services, and have better guidance from ODD R & E and the advisory committees on short and long range DoD goals.

The marketing group, simultaneously, provides a means for making determinant reporting more meaningful. For example, the skills of several IACs could be combined to generate an optimum user oriented product. EPIC could generate an electronic properties section of a report on organic insulation while PLASTEC could provide the chemistry and processing section. If necessary the REIC could be utilized to supply a section on radiation effects. Such a product would be more responsive to market place demands than the bits and pieces products which are now generated.

The centralized marketing organization also would have contacts with DoD, NASA, and other government agencies to:

1. Learn of future mission and discipline plans which require IAC assistance.
2. Determine programs which are now in progress which would benefit from IAC services.
3. Learn of scientific and technical programs in various stages of progress where the IACs could be of assistance to the sponsoring agency in assuring that state-of-the-art technology is utilized.

IMPLEMENTATION

In order to implement the plan we must review the basic principles of marketing and organization as related to the IACs as a business venture:

1. To package for benefits - we must remember that our customers do not buy information, only the usefulness of that information. Therefore, we must establish a method to make our determinant products of greater value to the scientific and technical communities and to our users at large.
2. To organize for function - we must allow the IACs to remain as engineering-scientific analysis groups. Therefore, we must establish supporting publication and marketing organizations.

The latter precedes the former in order of accomplishment. Before we can go further in our implementation efforts, we must:

1. Establish procedures that promote the interface between DoD advanced projects, national science policy, and IAC marketing.
2. Establish procedures for effecting interchange among the IAC research and analysis, marketing, and publication functions.

Once our procedures are established, we can set up the organization to accomplish our goals, retain the IAC network and publish products that sell because they are beneficial to the customer.

SUMMARY

Will the marketing specialist's preliminary concept for recovering fifty percent of the DoD IAC's operating budgets work? We don't know for sure, but I am convinced that it is an approach which warrants the greatest opportunity for success.

It is time for us, all, to stop debating about *should* or *shouldn't* we *charge,* and if we *should,* how *should* we *collect* the required fifty percent of our operating budgets. It is time to proceed with what needs doing *in the real world.* We must *organize* to survive and to maintain the IAC's as a national resource. We must *package* our products and services to sell *benefit,* rather than information.

ACKNOWLEDGEMENT

The author would like to gratefully acknowledge the assistance of Arthur S. Dlott, Hughes Aircraft Company in the preparation of this paper.

ACRONYMS

AFMDS:	Air Force Machinability Data Center, Metcut Research Associates, Inc., Cincinnati, Ohio 45209
DCIC:	Defense Ceramics Information Center, Battelle Memorial Institute, Columbus, Ohio 43201
DMIC:	Defense Metals Information Center, Battelle Memorial Institute, Columbus, Ohio 43201
DoD:	Department of Defense
EPIC	Electronic Properties Information Center, Hughes Aircraft Company, Culver City, California 90230
ERIC/CLIS:	Educational Resources Information Center, Clearinghouse on Library and Information Sciences, Washington, D. C. 20036
IAC	Information Analysis Center
MPDS:	Mechanical Properties Data Center, Belfour Stulen, Inc., Traverse City, Michigan 49684

NTIS:	National Technical Information Service, Springfield, Virginia 22151
PLASTEC:	Plastics Technical Evaluation Center, Picatinny Arsenal, Dover, New Jersey 07801
RAC:	Reliability Analysis Center, IIT Research Institute, Chicago, Illinois 60616
REIC:	Radiation Effects Information Center, Battelle Memorial Institute, Columbus, Ohio 43201
TPRC:	Thermophysical Properties Research Center, Purdue University, West Lafayette, Indiana 47906

THE COPPER DATA CENTER—A TOTAL-ACCESS SYSTEM

William T. Black
Battelle, Columbus Laboratories
and
W. Stuart Lyman
Copper Development Association Inc.

Our subject today is marketing. To the authors this means making the Copper Data Center sufficiently valuable that industry will be willing to support it. We have one big advantage in obtaining support, in that funds are channeled through a single organization, Copper Development Association Inc. (CDA) in New York. CDA is in turn made up of about 75 copper companies, brass mills, wire and cable mills, and foundries.

In addition to engaging in traditional trade-association activities, CDA went one step further. They developed techniques for aggressive market development of copper-based materials primarily through the development of prototypes. Such projects as an electric car, a shrimp boat with a copper-nickel hull instead of the traditional steel (because of resistance to corrosion and biofouling), and a "copper home" recently built and exhibited in Houston, have proven to be effective marketing tools, since they vividly demonstrate the feasibility of new applications. (Of course, the prototype itself can be sold—an excellent example of cost recovery. The copper home was sold the first day of the home show for which it was built.)

It was unlikely from the beginning that such an organization would settle for the "traditional" approach to an information center. When CDA came to Battelle-Columbus in 1964, it brought the requirement that a system be designed that would go to the user and not make the user come to it. Also, the Association wanted to "spoon feed" the user by making it as easy as possible for him to retrieve information. Thus, the requirements really boiled down to the idea of a heavy emphasis on the input side so that the output could be most effortlessly retrieved.

At that time, the Engineers Joint Council system was evolving. This approach stressed the importance of tight vocabulary control with a highly structured thesaurus which was computer based. Along with this was a printed index in "dual dictionary" form which permitted coordinate retrieval of information. Those who favored this system, however, tended to feel that once they gave their user a reference, their job was finished. It was up to him to dig out the actual details.

Another school quite active in operating a number of information centers favored the extract system, whereby the "guts" of a document was extracted from it and multiple copies were filed in a large room under appropriate subject and bibliographic headings. The big disadvantage of this approach was that no one worried too much about indexing (coordinate indexing was not possible), and advocates thought nothing of spending two or three hours browsing through large stacks of cards in order to extract the tidbit or two therein.

It seemed obvious to us that our answer—and one that would satisfy CDA's stringent requirements—was to marry the two systems. In doing so we no longer had to worry about filing multiple copies of extracts—we just needed one copy. These extracts could easily be arranged by broad categories and assembled into books which could be printed and distributed. In order to properly control the indexing vocabulary we created a computer-based thesaurus using the EJC approach. Finally, a computer-based dual index could be used to retrieve appropriate extracts. Thus, the system consisted of multiple volumes of the extracts, and one copy each of the latest issues of thesaurus and index. This formed a *total-access* system—the user had, at his own location, the entire system. To our knowledge there is still no other information system which gives its users a computer-aided means of quickly retrieving information and also gives them the actual information, so that they are essentially independent of the central file.

A logical step in the evolution of the system was to provide the opportunity for users to search out their information with a time-sharing computer using inexpensive terminals at the remote end and ordinary voice-grade telephone lines. Our system is tapped into the Battelle-Columbus computer (Control Data 6400) using a Battelle-developed time-sharing program (BASIS-70). Time-sharing systems can be a dangerous thing if you have inexperienced users. Most of the systems developed thus far seem to require the user to think in basic machine language. When an engineer or scientist is only going to use a time-sharing system once every 2 weeks or so, you need simplicity. A user has a difficult enough time trying to phrase his question using key words without being required to be a computer programmer just to get into the system.

An important component in marketing the CDA system, as well as in making it useful, is the participation of technical specialists in the day-to-day operation. We have about 90 specialists, world-wide, who review the current (and past) literature and decide what is useful in their specific areas of expertise. This is, of course, important to the technical validity of the center, but it also

(1) Gives the experts a current-awareness tool

(2) Makes them a part of the operation and so is psychologically important.

Also, at least one *technical* person, but usually two, index each document which goes into the sytem. The average document is indexed by about 40 technical terms, plus bibliographic terms.

The availability and accessibility of the system to the 75 CDA member companies makes it a tool that they can use in their engineering and research programs and in technical-service activities with their customers. We can back them up in these activities, but one of our ground rules is not to inject ourselves into the supplier-customer relationship.

We have been speaking this afternoon about cost recovery. In general this is taken to refer to sales of special publications of various types and fees for answers to inquiries. The fact that a profit-motivated industry has decided to *give* away these services, believing that it will more than recover its costs in increased sales, certainly has significance. The current drive in government-supported centers for cost recovery seems to us to be headed for the ultimate end point of saving all the money by reducing customer service to zero.

We believe that, if an information center really wants to serve its customers better and yet recover a large percentage of its costs, it must seriously consider providing these customers with total access to the system. In our case this means access to the entire file for each member company of CDA. However, for a segmented user audience, a system could easily be modularized so that a welding shop, for example, would only need to purchase the JOINING module of a metals-oriented information center. The value of total access to a system or a module of a system, as opposed to the traditional method of access to only a piece of information at a time, can be compared with the efficiency of purchasing all your food for 2 weeks at a supermarket, then preparing at home only what you need when you need it, as opposed to going to a restaurant every time you're hungry and having your meals cooked by someone else.

By offering a total-access system, we are now talking about something that's really useful and for which most companies in a given field would be willing to spend a significant amount of money on an annual basis.

Make no mistake about it, however—there are serious dangers to the information center in this. We know of no better way to highlight the inadequacies of a system than by turning it over to inexperienced users in remote locations. For example, if you're consistently carrying 18 months of backlog not yet incorporated into the retrieval system, it can be covered up with the normal central-file approach. With the remote system such a situation can become painfully apparent. Also, frankly we believe that few centers have very good indexing. Without good indexing and the tight vocabulary controls that go along with it, an inexperienced user can quickly become hopelessly frustrated and, as a result, "turned off" by the system. It's much safer to receive your inquiries at a central file and then scramble about inefficiently hand searching both your organized files and your backlog.

In summary, we believe that an information center *cannot* be effectively marketed a piece at a time. On the other hand, a total-access system *can* be marketed, and significant cost recovery can result from relatively few sales to relatively few customers.

PLASTEC Reports
Selling Through National Technical Information Service

The Plastics Technical Evaluation Center (PLASTEC) has utilized the services of the National Technical Information Service (NTIS of the Department of Commerce for almost two years now as the sales agency for PLASTEC reports. Quite encouraging results have been obtained, with $28,000 recouped by PLASTEC towards recovery of the printing costs of the reports. This paper reviews the two-year program.

BACKGROUND

PLASTEC is a DoD Information Analysis Center assigned to the Army and located at Picatinny Arsenal, Dover, New Jersey. The center is a part of the Materials Engineering Laboratory (MEL) which is fortunate, since the capabilities of the PLASTEC staff of 11 are supplemented by the plastics specialists in MEL. In its operation since 1960 PLASTEC has published 75 Reports and Notes. Most of these reports have been available to the public; some have had limitations for foreign distribution.

PLASTEC first considered a sales program for its publications late in 1968, when it became known that Department of Defense policy was moving in the direction of cost recovery for products and services for its information analysis centers. Prior to this, PLASTEC made a complimentary distribution of its technical reports, about six a year, to a mailing list of 1100. This covered defense agencies, their contractors and suppliers, and others with demonstrable defense interests. Copies were also made available to the Defense Documentation Center and the Clearinghouse (now NTIS), who between them dispensed from 300 to 1000 additional copies of each report.

SETTLING OF A SALES AGENCY

In surveying the possibilities of cost recovery, PLASTEC's first action was the institution of charges for certain inquiry services. Money received for this type of work was credited to PLASTEC operating costs through the funding system in effect at Picatinny (known as Army Industrial Funding). The number of jobs processed was small enough so that the paperwork was not a burden. However, handling of the hundreds of orders involved for the sale of a report, each for a few dollars, would have been an unreasonable burden and thus it was decided not to handle report sales directly from Picatinny. Had we done so, we would also have had to contend with strict local regulations on commercial advertising.

Next under consideration were outside sales agencies; and NTIC, commercial publishers and the Government Printing Office were evaluated. With no previous experience or direct DoD authorization for such an action, a contract with a commercial publisher did not seem promising. This avenue was not further explored, although the marketing activities of an aggressive publishing organization were considered very desirable. GPO had no reputation for handling sales on a reimbursable basis and NTIS did, so NTIS was approached.

THE ARRANGEMENT

The NTIS reputation was gained early in PLASTEC's history, in the early 1960's, when limited reimbursement was made by NTIS to PLASTEC for the sale of its reports. When NTIS went to a flat pricing system, this reimbursement ceased. However it was found that NTIS, in order to strengthen its position as an outlet for unrestricted government reports, was willing to establish a flexible pricing system. In fact, they had already made exceptions to the flat price scheme for other special cases. In short order, an agreement was reached with NTIS to handle sales of PLASTEC reports. The essence of the arrangement was that a price would be established for each report which would cover NTIS handling costs (announcement, order fulfillment, inventory, etc.) and PLASTEC printing costs. Therefore the reimbursement is not a "royalty" but only a recovery of the printing cost. PLASTEC would continue to do its own printing and provide the necessary copies to NTIS. NTIS would reimburse PLASTEC for its share of the sales receipts on a semi-annual basis. As with service charges, sales receipts can be used to pay PLASTEC operating costs. We considered briefly printing reports at NTIS, but decided to retain control at PLASTEC to insure the appearance and format were consistent with previous PLASTEC reports and because we could expect faster delivery by doing it ourselves.

A GOOD BEGINNING

There were two more cogent reasons for proceeding with a sales program in 1968. First and foremost was a budget cut, which sharpened the PLASTEC eve in seeking new sources of funding. Second was the imminent publication of a state-of-the-art report on polyurethane foam which was felt would have a strong sales appeal, both in and out of government. The report on polyurethane foams, with their wide range of composition and wide range of applications (thermal insulation, space rigidization and insulation, buoyancy, package and comfort cushioning, vibration damping, energy dissipation, electrical and electronic applications, etc.), turned out to be the largest PLASTEC report ever written, 245 pages. The author, incidentally, had organized his report on the basis of a comprehensive questionnaire to potential users of the material. This particular report publication will be examined in detail to explain pricing, promotion, and related procedures. It makes a splendid example, if not a typical one, because sales were beyond expectation.

SETTING THE PRICE

The pricing was determined as follows. We needed two figures - the number of reports we estimated could be sold and the total printing cost. The cost divided by the number sold would give the cost we needed to recover per copy. Although we had distribution figures from DDC and NTIS for most PLASTEC reports, we did not know 1) how many of the 1100 people receiving complimentary copies would buy copies, 2) how many obtaining reports from NTIS would pay the new higher prices or 3) how many new customers we might find through a wider promotion campaign. It was thus a guesstimate, plus study of our distribution histories, that let to an estimated sale of 800 copies. To that number we added 25 copies for distribution to technical journals and 30 for a VIP complimentary mailing list. A print order for 1000 copies was placed with our publishing contract. The cost was $10,000, which included typing on an IBM Magnetic Tape Selectric Composer (to provide justification, variable fonts, book-like copy and other composition features), extensive art work and the printing. Permission was obtained to use a better quality paper to get the program off to a good start. Printing is done on outside contract because the Arsenal facilities cannot handle the workload. Typing is done in-house if workload permits. Both PLASTEC and the contractor have the IBM MT/SC equipment, but not all reports are prepared on this machine.

Our cost recovery figure per report was therefore $12.50. To this was added $3.00 for NTIS costs (the NTIS cost has since dropped to $2 for most reports) and a selling price of $15.50 resulted. It was with some trepidation that we set this figure, for we were not at all confident that 800 people would pay $15.50 for a government report.

Our fears were unfounded. Sales passed the 800 figure in nine months and a second printing of 500 copies was ordered. The costs of the second printing were recovered with the sale of 170 reports. At the end of April 1971, 22 months after sales began, 1229 copies had been sold. Sales continue at the rate of about 20 copies per month, with a low figure of 11 in January 1971.

To protect the sales potential of our hard copy, it was necessary to restrict DDC involvement to an announcement in their abstract bulletin (which also establishes a bibliographic record), so that the report could not be obtained in microfiche at a token price.

PROMOTING SALES

The promotion was typical of that used on succeeding reports sent to NTIS, with the exception of a special first announcement (appendix 1) and a later 2nd Printing flier (appendix 2). NTIS sends an announcement letter, the draft of which we prepare, to 2000 libraries and to selected journals from a list of 900. The Fast Announcement Service and Government Reports Announcements also carry an item about each report. PLASTEC sends the NTIS letter to its mailing list of 1100 and a separate news release, with a copy of the report, to about 25 technical journals (depending on the subject matter of the report). About half of the journals solicited will carry a release or announcement of any one report. They are a definite help in sales. At the present time we have no feedback on the comparative values of these various approaches, nor do we know how our readers are divided as to organization or fields of interest. We are encouraged to learn that NTIS is now organizing to develop a more comprehensive marketing program.

The early success of the report was attributed to the generally fine reputation enjoyed by PLASTEC reports in the plastics community. The continuing two-year success indicates the report can stand on its own feet. Having had a chance to establish ourselves in the report field in the early 60's certainly helped in the acceptance of a sales policy.

THE OUTCOME

In addition to the polyurethane report, we have sales history on seven other reports in this period. The reports range widely in subject matter, total sales and in price (from $4 to $10). The subject matter has included technical topics - such as the weathering of polystyrenes; compilations - an index to government specifications on plastics, the annual index to technical conference papers on plastics, and trade names; and a directory - knowledgeable government personnel in plastics. Items in the latter two categories are updated periodically and have proven consistently popular. In fact, it should be noted that these more prosaic publications are often in more demand than certain high powered technical treatises. Interestingly, it's the customer who has prodded PLASTEC into making the periodic revision.

The index of plastics specifications, deliberately priced low at $4 to make it widely available to small business, has sold 1000 copies in 16 months, although the demand has now tapered off. Next in demand is the directory, with 685 copies in 18 months, with demand almost dried up. Trade Names is expected to be quite popular, with 300 copies sold in the first 3 months. For the typical technical report the records show 400 - 500 copies will be sold.

PROBLEM CHILDREN

We have made some good and some bad guesstimates on expected demand for reports. Lately we are getting pretty close on the technical reports. There has been one spectacular bad guess, in the case of the personnel directory, where we overestimated almost 100%. The lack of demand is attributed, in some part, to the poor condition of the national economy and the lessening of defense business. The three earlier editions of the directory were very much in demand by industry and it was on this we based our present order. The next revision will have a smaller print order and a corresponding higher price.

A second report which had a limited demand was on ablative composites. This was partly due to its very specialized subject matter and partly due to its foreign distribution restriction. In fact a change in DDC-NTIS policy for limited reports took the report off the market altogether after 200 copies were sold.

We cite the ablation report as an example of publishing on a subject where there is a need (the report was suggested by the Air Force) but where the revenue will not offset the printing cost, unless the selling price is out-of-line with usual book prices. If we had it to do over again, we would either find a sponsor for the report or set a higher selling price.

We always check the selling price per page and so far have never been seriously out of line with commercial book prices. The polyurethane report figures out to 6½ cents per page. The range has been from 3½ to 10 cents per page, the latter for the weathering report employing extensive art work.

SOME CONTRADICTIONS

The observation has been made that PLASTEC - or any center - can sell more cheaply through NTIS than through a commercial publisher. This is attributed to elimination of a profit factor, minimum of contract paperwork, and low marketing costs. This lower cost should in theory put the information in reach of a wider audience. However the strong sale of the polyurethane report, at a comparatively high price, puts question marks on the value of the theory and reminds one of an old adage that the customer has a higher regard for a product that does not have a give-away price. In any event, the total distribution of PLASTEC reports is appreciably less than before free distribution was stopped. Whether we have lost many bona fide readers is another question and one with no answer.

Another observation regarding the selection of a selling agency concerns the risk factor. It has been proposed that a commercial publisher is entitled to a substantial part of the sales proceeds because of the unknown market for IAC reports and consequent high risk factor. The argument has merit. By the same token, the IAC that does its own printing assumes a high risk and should provide for the unknown and unproven market in setting a selling price. The point is made so that any claim of "profit" on an IAC report can be counterbalanced by the risk factor and the losses incurred when an over-estimate of sales is made.

And for a third contradiction, we offer this. PLASTEC is now occasionally torn between its mission of writing reports in areas of need to the government and the temptation of looking for "best-seller" subjects. The two are not mutually exclusive but will probably not coincide very often, either.

FUTURE ACTIONS

The final part of this paper mentions several unresolved problems or areas needing work. One is the time at which each report is either no longer making sales or runs out of copies and does not justify a large reprinting. The mechanism for reproduction, pricing and microfiche availability must be determined.

A second item is reworking of the basis for setting the PLASTEC report cost preparation. An amount for out-of-pocket costs for in-house promotion should be included. And it might be fair to include some editorial salaries and rental costs of the IBM machine. There is no intention, however, to include the author's salary. The latter runs to many months and sometimes tens of thousands of dollars.

A third area is the extent of promotion that is justified by PLASTEC. So far, in addition to the mailings described earlier, PLASTEC has printed a broadside called "From the Looms of PLASTEC" (appendix 3), which has a short item about each of our reports of the past two years. This flier was mailed in April to 3800 plastic fabricators and suppliers and was generally a different set of names than previously solicited. The effect of this mailing is not yet known. The next mailing will be to a list of 1000 foreign names. The desirability of such a mailing is based on a strong interest in several reports abroad and the proven interest in United States technical publications in foreign countries. As part of the increased marketing activity at NTIS we hope to get feedback on the relative values of these special mailings.

For the immediate future we plan to continue our publishing operation as at present, hoping through experience to bring estimated and actual sales into closer alignment and hoping to increase sales overall by more effective marketing and sharper identification and earlier anticipation of user needs. For the longer future we are studying the possibility of subscriptions to PLASTEC reports and inquiry services. Such a policy would involve at least some changes in the handling of individual report sales.

Harry E. Pebly, Jr., Director
Plastics Technical Evaluation Center
Picatinny Arsenal, Dover, New Jersey

MAY 1971

INFORMATION ANALYSIS CENTER LIAISON WITH PROFESSIONAL ORGANIZATIONS

Young Park, *Administrative Coordinator*
ERIC Clearinghouse on Junior Colleges

Educational Resources Information Center (ERIC) is a national information system designed and supported by the U.S. Office of Education for providing ready access to information that can be used in developing more effective educational programs. Through a network of specialized clearinghouses, each of which is responsible for a particular educational area, current significant information relevant to education is monitored, acquired, evaluated, abstracted, indexed, and listed in ERIC reference products. These reference publications provide any educator with easy access to reports of innovative programs, outstanding professional papers, and reports of the most significant efforts in educational research and development.

Of the twenty clearinghouses throughout the country, each focuses its activities on a separate subject-matter area and provides documents for Central ERIC. All the clearinghouses carry on many projects to disseminate ideas and information to the education community. In addition to screening documents for input to the Central ERIC system, each clearinghouse is charged with providing information analysis products.

The Clearinghouse for Junior Colleges emphasizes information analysis and has undertaken to prepare and issue a variety of publications designed to provide analysis of pertinent research findings. Included in its regular series of publications are the ERIC *Junior College Research Review*, a Topical Paper series, and a Monograph series.

All these products deal with topics of immediate concern to the junior college practitioner, such as Teacher Training for the Junior College, Position Papers of Black Student Activists, Laws Relating to Higher Education in the Fifty States, Measuring Faculty Performance, and Personality Studies of students and staff, to mention only a few.

While an abundance of material is produced, the Clearinghouse is faced with a problem common to most information analysis centers—namely delivering its products to its users. It is financially impossible to create a nationwide network to publish and distribute the publications—not to mention the physical impossibility of such an undertaking. Although the Clearinghouse does attempt to publicize some of its products through the use of flyers, the major task of distribution has been assumed by the American Association of Junior Colleges.

Because of the limitations outlined above, the Clearinghouse for Junior Colleges has used the services of AAJC, the national professional organization, for publication and distribution. It also has international affiliations representing almost all two-year institutions. Located in Washington, D. C., the AAJC has been in existence for fifty years. Its major purpose is to promote and encourage the development of the junior college. Institutional membership in the AAJC is diverse as well as universal; the institutions vary in size, programs offered, people served, and financial support. As AAJC seeks a variety of ways to serve this diverse membership, it therefore was quite natural for it and the Clearinghouse to seek mutual assistance. In collaboration with the American Association of

Junior Colleges, the Clearinghouse has prepared a number of publications and, at the same time, expanded its dissemination capabilities.

Two major publications have resulted from this alliance—the *Junior College Research Review* and the ERIC Monographs. The Clearinghouse staff prepares the manuscripts (some written by staff members, others by authors from outside the Clearinghouse), provides the copy, and makes all editorial decisions. The AAJC prints and distributes the publications to all AAJC members and individual subscribers, and provides copies for sale throughout the country. Subscription fees to the *Junior College Research Review* and payment for the ERIC Monographs are retained by AAJC to help defray its printing costs. (It might be added that, as there is no contractual agreement with AAJC, they include the cost of publications in their budget).

During the past four years, the Clearinghouse and the AAJC have produced 35 issues of the *Junior College Research Review* and 12 Monographs. The JCRR is published 10 times a year and ERIC Monographs are published whenever suitable manuscripts are available. There are over 1300 individual subscribers to *Junior College Research Review*, in addition to the 1000 or more institutional members who receive copies. There are over 1300 individual subscribers to *Junior College Research Review*, in addition to the 1000 or more institutional members who receive copies. On the average, 2500 copies of each monograph are sold throughout the country. Since the AAJC distributes its publications to all its members, writers are assured of exposure of their material to almost every two-year institution in the country. The *Junior College Research Review* is generally based on ERIC documents; Monographs also rely heavily on the ERIC collection, thus exposing readers to related information on whatever topic is in the publication.

In this manner, the two agencies, one an information analysis and document resource center and the other a professional organization, are able to provide the practitioner and researcher with relevant information on many timely problems. The contents of the publications are directly related to research designs and evaluative models intended for institutional use.

A second major undertaking of the ERIC Clearinghouse for Junior Colleges involves direct use of the ERIC collection by a professional organization. By 1970, some 350 organizations had purchased the entire collection of ERIC microfiche. Recently, the AAJC has been added to the list of microfiche owners. The advantages of maintaining an active file of ERIC documents are obvious. One way the information in the ERIC collection can be used is in response to specific queries by practitioners, of which by virtue of its position as the most visible agency in the junior college field, AAJC gets some 300 per month. As the major source of information is the ERIC collection, it is most expedient for AAJC to maintain a microfiche file to answer these queries.

Since the retrieval system of the ERIC collection is unique, a basic orientation is necessary for those intending to make direct use of documents. The ERIC Clearinghouse for Junior Colleges recognizes this problem and has taken steps to instruct AAJC staff members in the use of the microfiche collection. In essence, the Clearinghouse has expanded its functions to include that of consultant to AAJC in the use of ERIC documents.

A close working relationship with a professional organization has many benefits for both agencies. First, the Clearinghouse for Junior Colleges is better able to accomplish its major purpose —distribution of information. Second, the professional organization is able to provide its members with relevant research findings as well as suggestions for designs and models that could be adopted by the practitioner. Third, the Clearinghouse is better able to educate agencies and organizations in the availability and use of materials in the ERIC collection. Finally, since the information analysis product is directed toward a specific group, dissemination of information is most easily accomplished by working with the group's professional organization—in this case, the AAJC. Association with a professional organization also means direct involvement with the practitioners who will become users of ERIC documents as well as potential contributors of materials.

Cooperation with a professional organization is only one means of expanding the services of ERIC. Other agencies and organizations can also become involved in much the same manner. For example, a Laboratory for the Study of the Community College is being planned at UCLA and its plans include the use of the Clearinghouse services. The Clearinghouse serves as a central communication and document source for the Special Interest Group for Junior Colleges, a branch of the AERA and is also a participating member of regional research groups throughout the country.

It is the intention of the Clearinghouse for Junior Colleges to continue this association with professional organizations—indeed, to seek ways to expand its association with a variety of organizations for its publications and its information dissemination services.

SUCCESSFUL MARKETING VENTURES, LIAISON WITH A COMMERCIAL FIRM

William E. Burgess
CCM Information Corporation
New York, N. Y.

CCM Information Corporation has drawn from several types of data banks as source materials for publications and, in one case, an educational package. These efforts represent cooperation with a government agency as well as the development and refinement of material beyond the original intention of the agency involved. I will expand upon this topic by discussing three types of products:

(1) Products developed in cooperation with a government agency.
(2) A product designed and created from a data base purchased from a government agency.
(3) A product created from a data base developed and maintained by CCM Information Corporation but drawing upon a service of a government agency.

All of these products rely upon technology (data manipulation and photocomposition) for the printed page.

PRODUCTS DEVELOPED IN COOPERATION WITH THE U. S. OFFICE OF EDUCATION

A. *Current Index to Journals in Education*

Since June 1964, the U. S. Office of Education has maintained the Educational Resources Information Center (ERIC), a national information system which disseminates educational research results, research-related materials, and other resource information. Through a network of specialized centers, or clearinghouses, each of which is responsible for a particular educational area, information is acquired, evaluated, abstracted, indexed, and listed in *Research in Education* (RIE). This reference publication provides access to report literature in the field of education. RIE has been unable to incorporate a proper awareness of the vast amount of literature published in periodicals and journals. This inadequate coverage has indicated the need for a second publication devoted exclusively to the periodical literature, drawing upon the subject expertise of the ERIC clearinghouses and vocabulary of descriptor headings developed for the indexing of educational literature. *Current Index to Journals in Education* was thus created to serve the information needs of the practicing educator, reference librarian, and educational researcher. The monthly publication has been given a unique organization to meet this multiple requirement.

CIJE currently covers 530 publications. The majority of these publications represent the core periodical literature in the field of education. The other publications indexed in CIJE represent coverage devoted to peripheral literature relating to the field of education. This unique feature assures access to important articles published in those periodicals which fall outside the scope of education-oriented literature.

All articles listed in CIJE are indexed by one of the 20 ERIC Clearinghouses or the ERIC Facility. Citations to journals in a particular issue of CIJE represent the titles received by the various processing centers during the month previous to publication. The *Thesaurus of ERIC Descriptors* is used for assigning descriptive terms listed in the Subject Index.

CIJE is compiled by means of computer manipulation of the data received from the ERIC Clearinghouses. Typesetting is accomplished by photocomposition. All entries are preserved on magnetic tape and forwarded to the U. S. Office of Education for merging with the ERIC computer file.

B. *Thesaurus of ERIC Descriptors*

The Thesaurus of ERIC Descriptors is a vocabulary developed by subject specialists at the ERIC clearinghouses. It is used for indexing the various documents, projects, and journal articles which are entered into the ERIC information system. All descriptors in the *Thesaurus* are based upon documents or journal articles previously indexed and currently included in the ERIC system. If a needed term does not exist in the *Thesaurus* during the indexing process, the required descriptor is introduced on the basis of the subject matter covered by the journal article. The candidate descriptor then undergoes lexicographic review prior to permanent assignment in the *Thesaurus*. The ERIC *Thesaurus* will be useful for a comprehensive search in the Subject Indexes of CIJE.

C. *ERIC Educational Documents Index 1966-1969*

This index brings together, for the first time, references to all research documents in the ERIC (Educational Resources Information Center) collection. These include *Research in Education, 1966 through 1969, Office of Education Research Reports, 1956 through 1965*, and *The ERIC Catalog of Selected Documents on the Disadvantaged.*

Includes documents ED 001001 through ED 031604. There is a Subject Index with complete titles and ERIC accession numbers (ED number). Complete titles and ED numbers are also listed with each entry in the Author Index. The ED numbers refer the user to abstracts in the publications covered, to microfiche of the documents, and to copies of the original document obtainable from the ERIC Document Reproduction Service.

D. *The Reading Micro-Library*

In an attractive and functional case, *The Reading Micro-Library* offers more than 1,000 microfiche of documents covering the significant developments in reading during the recent past. The collection includes documents on reading as well as articles from pertinent educational journals which were indexed, abstracted and announced in *Research in Education* from 1966 through 1969.

It's easy to locate the document required. The collection of microfiche is accompanied, in the same container, by a complete bibliography, *Recent Research in Reading, 1966-1969*. For every document in the microfiche file, the printed bibliography gives: (1) Main entry section, with citations and abstracts for ERIC documents. (2) Subject Index and Author Index. (3) An entry number which matches the microfiche containing the document cited. This entry number is included in the main section and in the Author and Subject Indexes.

E. *Recent Research in Reading A Bibliography, 1966-1969*

This selective bibliography provides the reading specialist with convenient access to the report and journal literature in reading which was indexed, abstracted and announced in *Research in Education* from 1966 to 1969. In addition, *Recent Research in Reading* includes citations of articles from more than 500 educational journals covered by *Current Index to Journals in Education* Volume 1 (1969). Items cited are arranged as follows:

1. Complete citations, including abstracts, to ERIC (Educational Resources Information Center) documents.
2. Citations with annotations from educational journal articles covered by ERIC.
3. A Subject Index to the above sections.
4. An Author Index to the above sections.

F. *C.L.A.S.S.-Current Literature Awareness Service Series: Reading*

For teachers of reading, including remedial reading programs, developmental reading, beginning reading, reading improvement, high interest-low vocabulary reading, and reading programs at elementary and higher levels.

CLASS:Reading provides current access for every classroom to what's new in reading, from articles appearing in more than 500 educational journals and from the research report literature which is indexed, abstracted and announced in *Research in Education.*

Advanced computer processing provides the most rapid service ever offered.

CCM Information Corporation developed the C.L.A.S.S. service in cooperation with the Office of Education and the ERIC Clearinghouse on Reading at the University of Indiana.

CLASS:Reading is published eight times a year, in issues dated September, October, November, January, February, March, April, and May.

Each issue contains approximately 32 pages in five sections:

1. Complete citations, including abstracts, to ERIC documents cited in *Research in Education.*
2. Citations with annotations of every article on reading published in over 500current educational journals.
3. Complete descriptions of all new reading research projects funded by the Office of Education.
4. A Subject Index to all of the above sections.
5. An Author Index.

A PRODUCT DESIGNED AND CREATED FROM A DATA BASE PURCHASED FROM A GOVERNMENT AGENCY

Bibliography of Agriculture

CCM Information Corporation began publication of the B of A in 1970. Data on magnetic tape is purchased from the Department of Agriculture on a monthly basis.

The *Bibliography of Agriculture* is a monthly index to the literature of agriculture and the allied sciences received in the National Agricultural Library. Publications from any country are indexed, provided that an entry has a summary, abstract, or translated title in one of the following languages:

Chinese	Hungarian	Russian
Czech	Italian	Serbo-Croatian
Dutch	Japanese	Spanish
French	Korean	Turkish
German	Polish	Ukrainian
Greek	Portuguese	

Literature received more than one year after publication is generally not indexed. Exceptions are made for important scientific publications.

Indexing of articles on the processing of agricultural products is limited to those on primary processing.

Unsigned articles and those signed with pseudonyms or initials, editorials, letters to the editor, and columns appearing regularly are omitted.

The bibliography is divided into five sections; a main entry section, a checklist of new government publications, a list of books recently acquired by the library, a subject index, and an author index.

A PRODUCT CREATED FROM A DATA BASE DEVELOPED AND MAINTAINED BY CCM BUT DRAWING UPON A SERVICE OF A GOVERNMENT AGENCY

Bibliography and Index to the U. S. Joint Publications Research Service (JPRS) Translations

The Transdex Publishing Program is both a bibliographic service that lists and indexes all the translations of the United States Joint Publications Research Service (JPRS) and a microform service that makes the JPRS translations available on film and fiche.

About 30,000 articles or books are translated by JPRS annually, and these materials are published (by JPRS) in approximately 3,000 documents. The 30,000 items were originally published in more than 145 countries other than the United States. Some of these countries have Communist or Socialist governments; some are developing nations; some have active Communist movements; and some are countries, or areas, in the midst of political conflict. All of the publications translated were done so at the request of at least one U. S. Government Agency. The publications consist of books, newspaper and journal articles; science abstracts; medical and technical journals; conference proceedings; economic and industrial reports; and military documents. JPRS issues all documents either in continuation by serial title, or as Ad Hocs.

CONCLUSION

The products just described were all developed, manufactured, distributed, and marketed at the expense of CCM Information Corporation. Plans are underway for additional publications and specialized educational products, many of which will be accomplished in cooperation with federal agencies and professional societies. CCM is aware of the mass of data available for the market place, and looks forward to profitable ventures.

IAC'S AND THE PRIVATE SECTOR

Jeffrey Norton, *President*, IIA
Publisher, Holt, Rinehart and Winston

In my "charge" from Harvey Marron, he asked me for an "overview talk with a healthy mix of philosophy and/or practical comments." That's broad enough, and wide enough, to permit talking about almost anything, for if a talk isn't "philosophical" it must, almost by definition, be "practical." So here, th. . is some philosophy and, I hope, some practicality.

As some of you may know, this year I am the President of the Information Industry Association—a group of private-sector firms active in the new forms and technologies of information products and services. Its purpose is to promote the development of private enterprise in the information field and to provide its members with a voice in determining the course of that development.

The Association now has approximately 50 member firms, ranging from some of the very largest in the country to very small single data base companies. A representative list of members include IBM, Xerox, Kodak, McGraw-Hill, Wiley, my own firm, Holt, Rinehart and Winston, Bell & Howell, Information Handling Services, Herner & Co., Institute for Scientific Information and Congressional Information Services, Inc.

The private publishing industry has traditionally been the principal agent supplying the informational needs of society. As new information needs developed at about the time of World War II, the private sector's response was slow and almost undetectable. At that time the markets for abstracting, indexing, micropublishing, and the forerunners of today's information systems were too small, and the risk too great, to make risk investment attractive. Besides, the publishing industry had a lot of other important matters to attend to, such as responding to the tidal wave of demand for new textbooks and the reissuance of many books whose reprinting had been forestalled by the wartime paper shortage.

So, while the information needs were expanding and becoming more and more sophisticated, the private sector's attention was focused elsewhere. Thus it was most appropriate that the Federal government stepped in to support and create information programs.

Well so much for history. In the past 5 to 10 years the private industry—both publishers and merchants of the new media—have made a large step forward to the point where we now are a significant part of the information age. The private sector now has the capability, experience, know-how, and willingness to take reasonable risks to attain reasonable profit objectives.

However, the background—almost tradition—of government funding of either its own operations or the not-for-profit organizations has been a difficult one for the private sector to break into. Happily, there are encouraging signs that various government agencies and departments are aware of the real and potential contribution to be made by the private sector.

Over the years this evolved into a complex specialized information enterprise requiring ever-increasing funding from public sources. Yet many agencies in the face of mounting demands for greater support are being faced with congressional cutbacks in funding which threatens existing services as well as the start-up of desirable new information services. Professional societies have had to cut back and agencies have had to eliminate or curtail support of various services. Yet the need for professionally developed and managed information systems is increasing logarithmically each decade.

I propose that the only effective long-range solution is a merger of public and private sector interests. The private sector—though slow to respond—has developed the capability to publish complex information services and to make them pay their own way. And at the same time, about 30% to 40% of the gross sales income goes back into support of the information service through direct "author" or "information center" support and through payment of taxes.

There are a number of private sector companies that have been in the business for many years and have demonstrated both professional competence and the ability to produce services that enough people will want to pay for to permit the information service not only to survive but to prosper. For many private-sector sector services *no* outside funding was required, or at least is not required now. Income from the users supports the service. The Institute for Scientific Information in Philadelphia is one such firm. I am sure most of you are familiar with its widely accepted publications ranging from print to computer tape and SDI services.

Another IIA firm with an even longer involvement with special information services may be less familiar. This is the Plenum Publishing Company which was formed almost 25 years ago in anticipation of the need of scientists and engineers in this country for English translations of the Russian scientific periodical literature. Over the years this company's translation program has grown as a sufficient number of customers have paid the freight to make expansion possible. Last year this firm published approximately 75 different Russian periodicals in translation, with all but the newest showing a respectable profit. I hate to think what the development and continuing operating costs of the program would have been under a public-support program. The Russian periodicals now produce for Plenum several million dollars of revenue annually, and it was all done without public support. And at the same time, a unique translation resource was developed, so that in turn it now provides several hundred thousand dollars per year of professional translation services on a contract basis to one of the leading professional societies. Plenum now also provides a computer-based patent search service and is developing its information handling and publishing capacities in several other innovative areas. In short, Plenum, The Institute for Scientific Information, and many other private firms have the capabilities to work hand-in-hand with IAC's to develop, publish, market, and help to sustain existing or new information services.

Well, enough philosophy and point of view. Now I'd like to comment on some of the practical issues that the private sector has to face in deciding on the feasibility of investing in an IAC-type information service, whether with or without government support. A typical private-sector firm would have to take into consideration and fully evaluate most or all of the following seven key questions:

Question 1. *Exactly how big is the market?* Though a given subject field may have many professionals or semi-professionals working in it, the number that feel they need and will use an information service is almost always severely limited. No matter whether they *should* use it or not, most won't. To embark on a publishing program assuming otherwise is to court disaster.

This leads to:

Question 2. *What is the probable cost of customer education?* If most people in a field are at best passive prospects for a new service, how much will it cost and how long will it take to get enough customers to make the service self-supporting? It's far from enough to rely on publicity announcements, a mailing piece, and a few ads to induce the far-from-breathless customer to subscribe. Most new information services *are* different and the publisher must allow for the cost of explaining it and educating the marketplace.

These customer-education expenses can range a wide gamut and usually include some or all of the following:

1. Arranging for extensive reviews and articles in the professional literature
2. Elaborate demonstration brochures
3. Wide distribution of sample copies
4. Preparation and distribution of sample computer tapes with support program adapted to an individual prospective customer's requirements
5. National and international sales force
6. Customer services
7. Participation in conventions and professional meetings
8. And, inevitably, troubleshooting

Which provides a lead-in to:

Question 3. *What are the probable total start-up expenses and can I realistically expect to recover them?* Start-up expenses include *everything* from the original development of a data base until the point at which enough income is coming in to support the on-going costs of the service. In the private sector, hardly any services reach this self-maintenance point in less than two years, and many take three or four years or more. And once that point is reached, the publisher then must look to subsequent years and to an increased customer list to recoup his original start-up investment! And even when the service involves a "supported" publication, the road is far from easy. CCM Information Corporation found this to be the case with its ERIC-sponsored monthly publication *Current Index to Journals in Education*. There was another education index which had been on the market and in use for many years. However, *Current Index to Journals in Education* had a lot going for it:

- It was sponsored by the Office of Education
- It indexed 50% more journals
- Its subject headings and descriptions were more in line with current educational research and practice
- Each entry was annotated
- It was more attractive
- It was cheaper

Despite all this, the first 18 month sales were disappointing. CCM had anticipated obtaining by that time approximately half the number of subscriptions held by the other index. In fact sales by that time were less than half what had been anticipated. Meanwhile sales of the competing publication had increased by almost 15%!

All of which proves something about the difficulties of launching a new information service as well as about the buying practices of libraries. Frequently, then, it's necessary to hedge one's bet by developing other uses of parts of the data base, and through being alert to opportunities to make the file more useful to more people.

To do this profitably requires evaluation of:

Question 4. *How can I obtain effective customer feedback?* With the start-up costs grinding on week after week and month after month, it's vital to include an efficient feed-back loop to permit quick modification or improvement in a system that isn't meeting what the customer needs. Often this will mean extending or changing the coverage, altering the depth of treatment, or speed or method of processing. To embark on a new information enterprise on the assumption that it is possible to conceive *de novo* what the customer needs is the height of foolhardiness.

Customer feedback is also important in permitting the publisher to evaluate:

Question 5. *Are there any by-products to be produced from the data base that can help make the overall enterprise profitable?* Customer feedback helps here as well as the ability of the publisher to identify related or peripheral markets. If the data base is maintained on computer it is often possible to produce worthwhile income-producing subservices at only marginal extra cost. The previously mentioned ERIC data base provides a good example of how this can be done:

Product	Estimated 1971 Income
Current Index to Journals in Education	$93,000
Semi-annual cumulation	11,000
Annual cumulation	40,000
All-ERIC Index	28,000
ERIC Thesaurus	12,000
Reading Bibliography	14,000
Early Childhood Learning Bibliography	6,000
Reading Microlibrary	28,000
Early Childhood Learning Microlibrary	12,000
Class: Reading	15,000
Total	$259,000

Prospective:

Computer tape service	?
Computer-on-line service	?
Microfilm journal articles	?

In essence, then, the ability to innovate is a key element in the proper management of a data base. It opens the way to maximum utilization of the information resource, and vastly enhances the likelihood of long-run success.

And success may, in the last analysis, depend on:

Question 6. *How can I cut costs and expenses and still keep the service going?* If the market doesn't respond broadly enough, or quickly enough, in the private sector, the problem very quickly becomes: *adapt or die*. The expectation of owners or stockholders to make a profit provides an inexorable force toward finding solutions to seemingly impossible problems. Believe me, I don't think there's any of us in the private sector who haven't had to find out the hard way how to do more for less.

And finally:

Question 7. *Is there reasonable expectation of exclusivity for the marketing of the data base to permit risking the substantial start-up investment for programming, packaging, marketing, and customer education?* Most information service markets are limited. They may be profitable for a single entrepeneur, but become unattractive if two or more parties are free to publish the same output from the same data base.

In most cases what this requires then is some sort of license plus a copyright in the output. It's not too much to hope that in time various government agency regulations and even the basic copyright legislation will be modified to permit exclusive assignment of copyright—at least for a period of years—to the private-sector publisher who through an innovative approach or responsiveness to an RFP demonstrates his capability and willingness to invest his share in bringing a new information service to the marketplace.

It's apparent by now, I suspect, that I believe the private-sector information industry is qualified to work with IAC's and other publicly-supported information centers in the development of cost-effective and widely useful information services. The for-profit industry is probably in a unique position to:

1. Evaluate market size objectively and realistically
2. Evaluate the need for, and then perform, customer education
3. Accurately identify all probable start-up and on-going expenses and costs and use them as a measure for developing a realistic relationship between cost and price to the user
4. Obtain and quickly apply feedback from users and prospective users
5. Identify and make risk investments in the development and publication of useful by-products
6. Apply cost-reduction techniques
7. Develop maximum penetration of the market provided he has protection by license or copyright from unfair competition

And finally, and hopefully, if he makes a go at it, to be profitable enough to pay taxes that may in part go toward filling new information needs not yet dreamed of.