DOCUMENT RESUME

ED 043 524 SE 009 925

AUTHOR Phelps, C. Russell, Ed.
TITLE Proceedings of a Conference on a National Information System in the Mathematical Sciences (Harrison House, Glen Cove, New York, January 18-20, 1970).
INSTITUTION Conference Board of the Mathematical Sciences, Washington, D.C.
SPONS AGENCY Alfred P. Sloan Foundation, New York, N.Y.
PUB DATE Jul 70
NOTE 58p.
AVAILABLE FROM Conference Board of the Mathematical Sciences, 834 Joseph Henry Bldg., 2100 Pennsylvania Ave., N.W., Washington, D.C. 20037 ($1.00)

EDRS PRICE EDRS Price MF-$0.50 HC Not Available from EDRS.
DESCRIPTORS Chemistry, *Conference Reports, Engineering, *Information Retrieval, Information Science, *Information Systems, *Mathematics, *Natural Sciences, Physics
IDENTIFIERS Conference Board of the Mathematicsl Sciences

ABSTRACT

The purpose of this conference was to consider ways of developing a system of improved information services for the mathematical sciences and for the interfaces with related scientific fields. Conference memebrs investigated the achievements, coverage, and technology of existing information services and systems in the fields of engineering, physics, chemistry, computing and control, and statistics, as well as mathematics and mathematics education. Descriptions of the current status and developmental plans of various information systems are provided by scientists involved in these systems. Possible compatibilities of the classification schemes and data bases of each system with other systems were explored, as were the economic and management concerns of the several systems. These discussions were related with respect to their implications for a national information system for the mathematics sciences. (Author/FL)

ED0 43524

Conference Board of the Mathematical Sciences

Proceedings of a Conference on a National Information System in the Mathematical Sciences

Edited by

C. Russell Phelps

Harrison House
Glen Cove, New York
January 18-20, 1970

Supported by the Alfred P. Sloan Foundation

CONFERENCE BOARD OF THE MATHEMATICAL SCIENCES

Proceedings of a Conference

on a

NATIONAL INFORMATION SYSTEM IN THE MATHEMATICAL SCIENCES

Edited by

C. Russell Phelps

Harrison House
Glen Cove, New York
January 18-20, 1970

Supported by the Alfred P. Sloan Foundation

July 1970

Additional copies available at $1 from
Conference Board of the Mathematical Sciences
834 Joseph Henry Building
2100 Pennsylvania Avenue, N.W.
Washington, D.C. 20037

CONTENTS

Conference on a
National Information System in the Mathematical Sciences
January 18-20, 1970

Conference Participants

Conference Chairman
*Donald L. Thomsen, Jr.
IBM Corporation

NISIMS Group Chairman
*Robert M. Thrall
Rice University

Franz L. Alt
American Institute of Physics

Richard D. Anderson
Louisiana State University

*Garrett Birkhoff
Harvard University

I. Edward Block
AUERBACH Info, Inc.

*Truman A. Botts
Conference Board of the Mathematical Sciences

*R. Creighton Buck
University of Wisconsin

Eugene Garfield
Institute for Scientific Information

*J. Wallace Givens
Argonne National Laboratory

Arthur Herschman
American Institute of Physics

Eugene B. Jackson
IBM Corporation

Richard Kenyon
American Chemical Society

Myer Kessler
Massachusetts Institute of Technology

William J. LeVeque
University of Michigan

Morton Malin
Institute for Scientific Information

Hugh J. Miser
University of Massachusetts

C. Russell Phelps
Conference Board of the Mathematical Sciences

*Robert W. Ritchie
University of Washington

John A. Scopino
National Science Foundation

Fred A. Tate
Chemical Abstracts Service

Daniel Teichroew
University of Michigan

Howard E. Tompkins
Institute of Electrical and Electronic Engineers

Joseph F. Traub
Bell Telephone Laboratories

Herbert Trotter
McGraw Hill Publishing Company

Geoffrey S. Watson
Johns Hopkins University

*Member of NISIMS Group; Jack E. Forbes, Purdue University, is also a member.

INTRODUCTION

The Conference on a National Information System in the Mathematical Sciences was sponsored by the Conference Board of the Mathematical Sciences at Harrison House, Glen Cove, New York on January 18-20, 1970. The purpose of the conference was to study ways of developing a system of improved information services of all kinds for the mathematical sciences and for the interfaces with related scientific fields. The conference was organized by the CBMS working group on a national information system, under the chairmanship of Robert M. Thrall, and was under the direction of Donald L. Thomsen, Jr. Support for the conference was graciously provided by the Alfred P. Sloan Foundation.

In summary, the conferees explored the achievements, coverage, and technology of existing information services and systems in the fields of engineering, physics, chemistry, computing and control, and statistics, as well as mathematics and mathematics education. Scientists centrally involved in these information systems described the current status and developmental plans of their reviewing and abstracting systems, and their potential for interrelationships with the mathematical sciences. Possible compatibilities of the classification schemes and data bases of each system with other systems were explored, as were the economic and management concerns of the several systems. All of these discussions had as a common thread their implications with respect to a national information system for the mathematical sciences.

The 26 invited participants--listed heretofore--included the members of the CBMS working group and representatives of information systems and services in the mathematical sciences and related fields.

Opening Session

In opening the conference the chairman, Donald L. Thomsen, Jr., conveyed greetings from Dr. Larkin H. Farinholt, Vice President of the Alfred P. Sloan Foundation, who was unable to be present because of illness.

Keynote remarks were given by Garrett Birkhoff, chairman of CBMS. He commented as follows on the purposes and objectives of the conference:

> We have come together to discuss, all too briefly, the problems of developing an improved National Information System in Mathematics (NISIMS). Most of the problems which we face stem basically from the vastly increased number of people using mathematics in depth, and the vastly increased quantity (if not quality!) of the mathematical information which they produce. The solutions to our problems will come, presumably, from the economic resources available to those people needing information, and to improved techniques and understanding of "information processing."

In mathematics, as in other sciences, the basic problem is to filter the vast quantity of information available, so as to select the most relevant and high quality information for the use of individual human beings. Loosely speaking, it must be filtered from the scale of a library containing millions of volumes to that of a human individual who has time to read carefully only one or two volumes a week. Traditionally, this filtering has been done on a smaller scale by human beings, reviewers and referees, and editors of journals and writers of books among others. Indeed, good books provide among the best information systems.

Our meeting is not the first of its kind; the American Mathematical Society organized a Conference on the same subject in Providence on December 5-7, 1967, attended by 172 people [3]. In the two years since the Providence meeting, there have been intensive work and thought devoted to these problems. In particular, large-scale experiments have continued in the areas of computer-aided translation (which now looks unpromising), of computer-aided photo-composition, and of various ways of supplementing and expediting the normal flow of mathematical information through oral communication and the channels of journal and book publication and Mathematical Reviews. Perhaps the most promising experiments are the Mathematics Offprint Service (MOS) and the biweekly title listings distributed as Contents of Contemporary Mathematical Journals; but it remains to make these services financially self-supporting,* and to perfect them.

It is my personal impression that these supplementary services, if fully developed to include the selective dissemination of titles and abstracts, would make the normal flow of mathematical information reasonably adequate for communication between specialists in the same field of pure (or "core") mathematical research.

However, mathematics has important interdisciplinary uses; moreover our information problems are shared by other sciences, who are developing their own systems for disseminating title lists, abstracts or reviews, and reprints or preprints.

Chemical Abstracts has been notably successful in incorporating modern computer techniques into their information system, and we are fortunate in hearing about what they do. In addition, the Information Division of the American Institute of Physics has been awarded a grant of over $3 million by NSF-OSIS to support a three-year implementation of its clearly stated and carefully thought out plans [1]. This system includes as a key element the review journal Physics Abstracts, which is published by the Institution of Electrical Engineers in London.

In cooperation with this Institution, the IEEE in the United States is spearheading a related Technical Society Information Program. Both the IEEE system and that of AIP include mathematics as an element; it will be interesting to compare their classification schemes for mathematical information with those used by Mathematical Reviews.

* Currently, MOS costs about $200,000 a year and gets about $80,000 in subscriptions.

Though we hope to learn much from organizers of information systems of other sciences, of course our problems and users are different from theirs. For example, there are many more chemists than mathematicians. Also, large-scale efforts play as yet a relatively small role in the mathematical sciences. Again, mathematicians tend to read and write more than physicists or chemists. In view of these and other differences, it will be very interesting to see how much the systems considerations for the mathematical sciences resemble those utilized by chemists, physicists, and engineers.

In my opinion, one of the deepest problems consists in adapting general ideas about "systems" and "information-processing" to the structure of information in particular fields.

In pure mathematics, information consists in statements of implication of the form "H implies C," where H is some set of hypotheses $H_1,\ldots,H_m$ and C is some set of conclusions $C_1,\ldots,C_n$. Moreover, $\{H_1,\ldots,H_m\} \to \{C_1,\ldots,C_n\}$ is equivalent to $\{H_1,\ldots,H_m\} \to C_k$ for $k = 1,\ldots,n$. This permits one to classify much of pure mathematics into areas involving different subsets of hypotheses (G is a group, E is a Banach space, etc.). However, even in pure mathematics, only a superficial idea of the structure of mathematical information is obtained in this way.

Thus, consider the statements:

(a) $x^n + y^n = z^n$, $(x,y,z \in \underline{P})$ imply $n = 1$ or 2;

(b) If $f(x)$ is continuous, then $f(x)$ is (Riemann) integrable.

Their significance is apparent only to educated mathematicians. This is because their information content for the consumer depends on the response evoked through chains of association; it conforms only superficially to the classic Shannon theory.

The same is true of natural languages in general, and computer scientists are very busy these days trying to develop programming languages having "associative" features for retrieving data having "tree" or "ring" structure.

The structure of information in physics and chemistry is presumably quite different from that of the information covered in _Mathematical Reviews_. It may be very hard to devise a classification scheme for mathematics which is compatible with the polyfaceted system of classification being developed currently by AIP. Yet we must try to make it so by developing compatible interfaces across which information can flow between _Mathematical Reviews_, _Computing Reviews_, _Applied Mechanics Reviews_, _Physics Abstracts_. Otherwise, scientists will soon find themselves communicating in a veritable Tower of Babel! Even mathematicians and computer scientists dealing with closely related problems will find cooperation increasingly difficult.

Indeed, information monitoring is already becoming more and more difficult for those making new applications of mathematics. Mathematical Reviews has been forced to abandon much of its coverage of applications; whereas 57% of the contents of Volume 22 (1961) were on probability, statistics, and applications, in Volume 37 (1969) this coverage has been reduced to 23%. On the other hand, the mathematical significance of papers reviewed in Physics Abstracts, Applied Mechanics Reviews, and even Computing Reviews is seldom clear from the abstracts in these journals. Moreover, the sheer volume of relevant material makes browsing in the journals and books reviewed very inefficient for those engaged in multidisciplinary research.

Therefore, special efforts are needed to improve the processing of information relevant to multidisciplinary research which uses mathematics in depth. The New Areas Committee of CBMS has been working for two years, on a voluntary basis and with negligible financial support, on what can be done to improve this deteriorating situation.

I have spoken at some length about the aspects of our problem which are closest to my heart, I will be very brief about the others which will be discussed during our meeting.

The problems of education are obviously of the greatest importance; well over 50% of the individual members of CBMS-affiliated societies are professional mathematics teachers! We will then turn our attention to the related problems of libraries and journal publication, with emphasis on the former. Many of us have heard or read fascinating futuristic and broad-brush descriptions of library-based information systems of the future. It will be interesting to learn something about the current status of these futuristic ideas.

Next to last, we will hear some thoughts about the all-important question of the economics of improved information services (including miniaturization and magnetic tape storage). We hope to be helped materially in our assessment of the costs of and demand for such services by experts on operations research and the management sciences, disciplines so ably represented as members of CBMS.

Finally, we will consider critically the present NISIMS plans which are almost exclusively concerned with systems aspects, and discuss how they might be implemented in the future. This reconsideration of our basic mission is, indeed, what this meeting is primarily directed towards, and I am confident that it will be very fruitful.

I. SCIENTIFIC COMPUTATION AND MATHEMATICAL FUNCTIONS

J. Wallace Givens discussed the information needs of scientific computing. He pointed out that useful and accurate numerical results for major scientific problems require the interdisciplinary combination of (1) a knowledge of the experimental scientific facts; (2) skill in mathematical analysis; and (3) skillful computer programming. Because of its interdisciplinary nature, the subject is not well covered in existing reviews, of which the best collections are in Mathematical Reviews and Computing Reviews. The former reviews about 600 papers annually under the headings "Numerical Methods" and "Computing Machines", while the latter reviews about 250 papers annually under "Numerical Methods".

In the area of mathematical functions Givens noted the development at the National Bureau of Standards of the Abramowitz-Stegun work on tables of mathematical functions which has now sold over 100,000 copies. This is solid evidence of the usefulness of an organized account of properties and values of mathematical functions useful for scientific computation. The computer has, however, radically altered the needs and possibilities. On computers the values for most functions are now generated more efficiently by special computer algorithms rather than by direct use of tables. However, the algorithms and computing techniques now in use are far from uniformly good. A study of possible informational improvements should take into consideration:

a) which functions might best be computed by algorithms;
b) the greatly increased capability of producing graphical output under direct computer control;
c) properties and interrelations of functions, e.g., asymptotic behavior, upper bounds, identities connecting functions, etc.;
d) the critical need for accurate cataloging of sets of functions such as integrals;
e) the maintenance of the system by keeping pace with up-to-date hardware, correcting errors and sending corrections to users, and replacing software by improved versions;
f) the need to consider not only real functions of one variable but many essentially more complicated areas as well, such as eigensystems of real or complex matrices, solutions of certain classes of ordinary or partial differential equations, etc.; and
g) systematically related computer codes for users.

He characterized this body of information on scientific computation as follows:

1) It is of service to outside users.
2) It has an inherent organization.
3) It can be handled only with the aid of computer hardware.
4) It will require the development of methods of maintenance, correction, and frequent repackaging.
5) It is difficult to produce through industry financing, because it can't be copyrighted and involves many university personnel.
6) It raises many mathematically interesting questions, often at a research level.

He noted that the design of computers would probably be dominated by the interests of business use rather than by those of scientific computing. Thus, the solutions to problems in scientific computing must be built around computer capabilities.

Miser commented that the state of the computer art is changing more rapidly than the mathematics art and, therefore, one would have to consider probable computer changes even while the project is in progress.

Birkhoff noted the pitfalls in the elaboration of function formulas ad infinitum as in the case of elliptic functions in the 19th century. He also noted the problem of eliminating obsolete information, which might otherwise overload the system.

Buck observed that at present there is considerable unnecessary duplication in programming which might be avoided through more effective communication.

II. INFORMATION SYSTEM DEVELOPMENTS IN SELECTED DISCIPLINES

Eugene B. Jackson, the chairman of this session, commented on the information explosion as measured by the growth in numbers of abstracts. Biological Abstracts is now in its third million, having taken 8 years for the second million compared to 35 years for the first. Medlars, a mechanized system for medical abstracts centered on the National Library of Medicine covered its first million abstracts in 5.8 years, while the Engineering Index has covered 1.7 million articles since 1885. He emphasized the importance of the SATCOM [6] recommendations concerning abstracting and indexing services (of the 55 SATCOM recommendations, those covering "Classical Services" are especially relevant ones) and recommended that the entire summary report be read for background. He also noted that at Boston last December, Burton W. Atkinson, Head of the NSF Office of Science Information Service, had reminded abstracting and indexing services of the growing proportion of world science literature originating outside the U.S. and of the need for the several services to be aware of the "repackaging" of scientific information that goes on for use in interface areas.

These points were strongly reinforced by the other speakers at this session. Their presentations stressed the international scope of operations, and even more strongly, the need for detailed knowledge of each other's scope, procedures, and plans. Jackson offered a diagram (see page 7) to clarify some relationships of various international organizations.

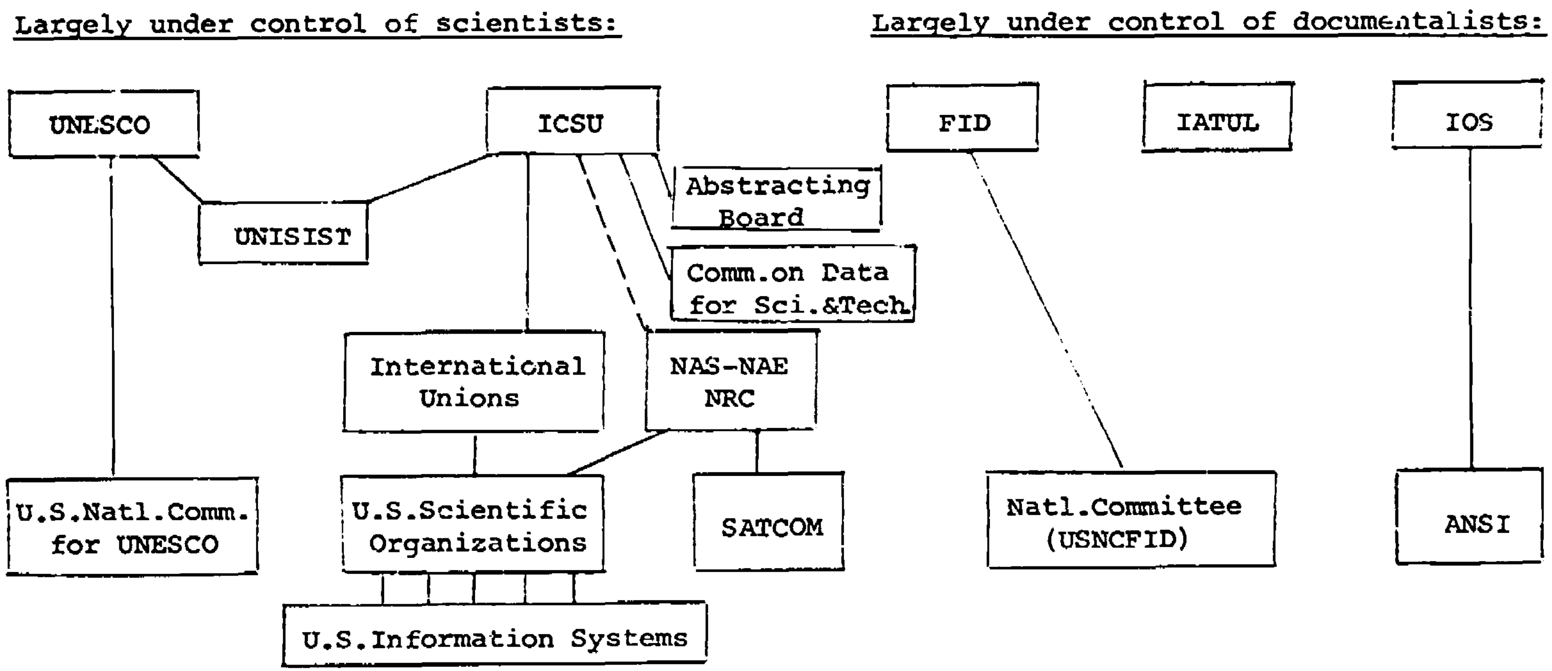
Chemistry

Fred A. Tate, Assistant Director of the Chemical Abstracts Service (CAS), discussed some of the current activities and concerns of the CAS operations. With respect to an Automated Processing System for Chemical Abstracts, he noted that through automation the 35 manual operations (14 professional and 21 clerical) used in 1966 now had been reduced to 22 operations (10 professional and 12 clerical), and that by 1974 to 1978, when the computer system is fully operational, they hope to reduce this to 14 separate manual operations (6 professional and 8 clerical). Their principal methodology in automation involves avoiding human transcription and the resultant necessity for human verification.

He stressed the necessity of evolutionary development--it is impossible to stop the present services while making changes in the system.

It is necessary to share input responsibility. In the international area, this has taken the form of an agreement with the U.K. Chemical Consortium on Chemical Information to supply abstracts and to assume marketing

International Science Information Interrelations



UNESCO: United Nations Educational, Cultural, and Scientific Organization
UNISIST: a world science information system
ICSU: International Council of Scientific Unions
FID: International Federation for Documentation
IATUL: International Association of Technological Libraries
IOS: International Organization for Standardization
ANSI: American National Standards Institute
SATCOM: NAS-NAE Committee on Scientific and Technical Communication

responsibility. An agreement has been reached with the German chemical community under which the Chemische Zentralblatt is ceasing publication. Its staff will instead abstract and index the German chemical literature for CAS, and the German Chemical Society will market CAS publications. Similar amalgamated activities involving the Russian and Japanese chemical communities are under discussion, but such discussions take years to consummate*.

Another facet of sharing input responsibility involves persuading publishers of primary journals to provide computer-readable materials. Also, similar efforts should be made with the secondary (abstracting and indexing) services of related disciplines. He expressed a hope that compatibility of input could be negotiated with other disciplines.

Sharing output responsibility is needed also--with subscribers, with foreign information centers, and with other services which purchase information for repackaging.

Sharing marketing responsibility is important, not only with cooperating foreign agencies who provide input services, but also with other services provided by the same organization (ACS). There is a close relationship with input.

Tate discussed the difficulties in computer representation of symbols used in writing about science, since the standard computer keyboard (typewriter) has only 88 symbols. This is an area in which standardized representations must be agreed on.

He presented the following table showing the evolutionary development by CAS of its experimental computer-readable services project, together with the number of characters (letters and symbols) for each individual entry in the computer file:

	Year Automated	No. of Characters Required
Chemical Abstracts--Semiannual Volumes		
Formula Index	1967	200
Author Index	1968	400
Subject Index	1969	700
Patent Concordance	1967	48
ACCESS	1969	200
Chemical Abstracts--Individual Issues	1972	1,500

* For fuller details on Automated Processing and on international agreements, see the CAS Report on the Twelfth CAS Open Forum [5].

Physics

Arthur Herschman, director of the Information Division of the American Institute of Physics (AIP) presented a philosophical paper, "Toward an Integrated Information System," in which he discussed the nature of scientific information, its producers and users, the functions of an information system, and criteria for integrating an information system. This paper is reproduced in Appendix B.

The AIP activities in information were described by Dr. Franz L. Alt*. The AIP "Information Store" is a computerized collection of "files", each relating to an article in a primary journal. The Store has 70,000 files, presently being added at the rate of 2,000 per month (1968, 17,000 added; 1969, 24,000 added). The average file contains information consisting of about 2,500 characters, containing the following ten items of information:

1. Journal Article Identifier
    a) Journal Coden
    b) Volume and issue numbers
    c) Beginning page of articles (inclusive pagination under consideration for 1970 input)
    d) Year of publication
2. Title of Journal Article
3. Personal Authors
4. Author(s) Affiliation(s)
5. AIP Classification Number
6. Free Language Descriptors
7. Journal Articles Cited in Article
    a) Author(s)
    b) Journal Coden
    c) Volume/issue number
    d) Page
    e) Date
8. Other Literature Cited in Article
    a) Author(s)
    b) Title
    c) Publisher's name
    d) Publisher's location
    e) Publication date
    f) Page(s) referenced
9. Abstract
10. Other Bibliographic Elements (Language of publication, type of paper, etc.)

Alt indicated the relative weight of the various items in the file as follows: items 1-4, a total of 300 characters; items 5-6, 100 characters; items 7-8, 1,500 characters; and item 9, 600 characters. He noted that the AIP abstract (item 9) used only 600 characters, whereas Chemical Abstracts

---

* For a general description, taken from Physics Today, December 1969, see Appendix C. A thorough analysis of the physics information system is the subject of a special AIP report [1].

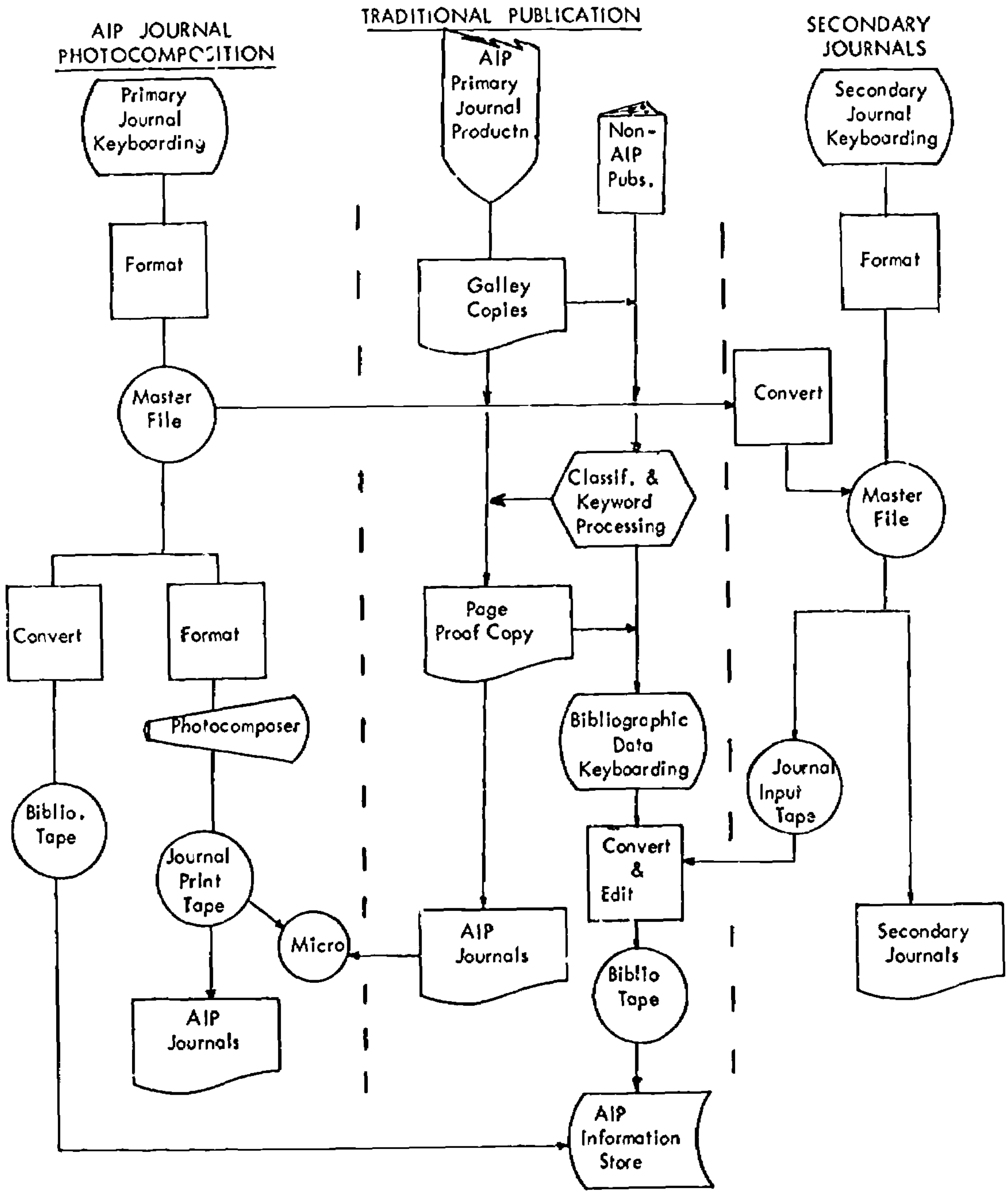


Fig. 1 - AIP INFORMATION STORE INPUT - PUBLISHED LITERATURE

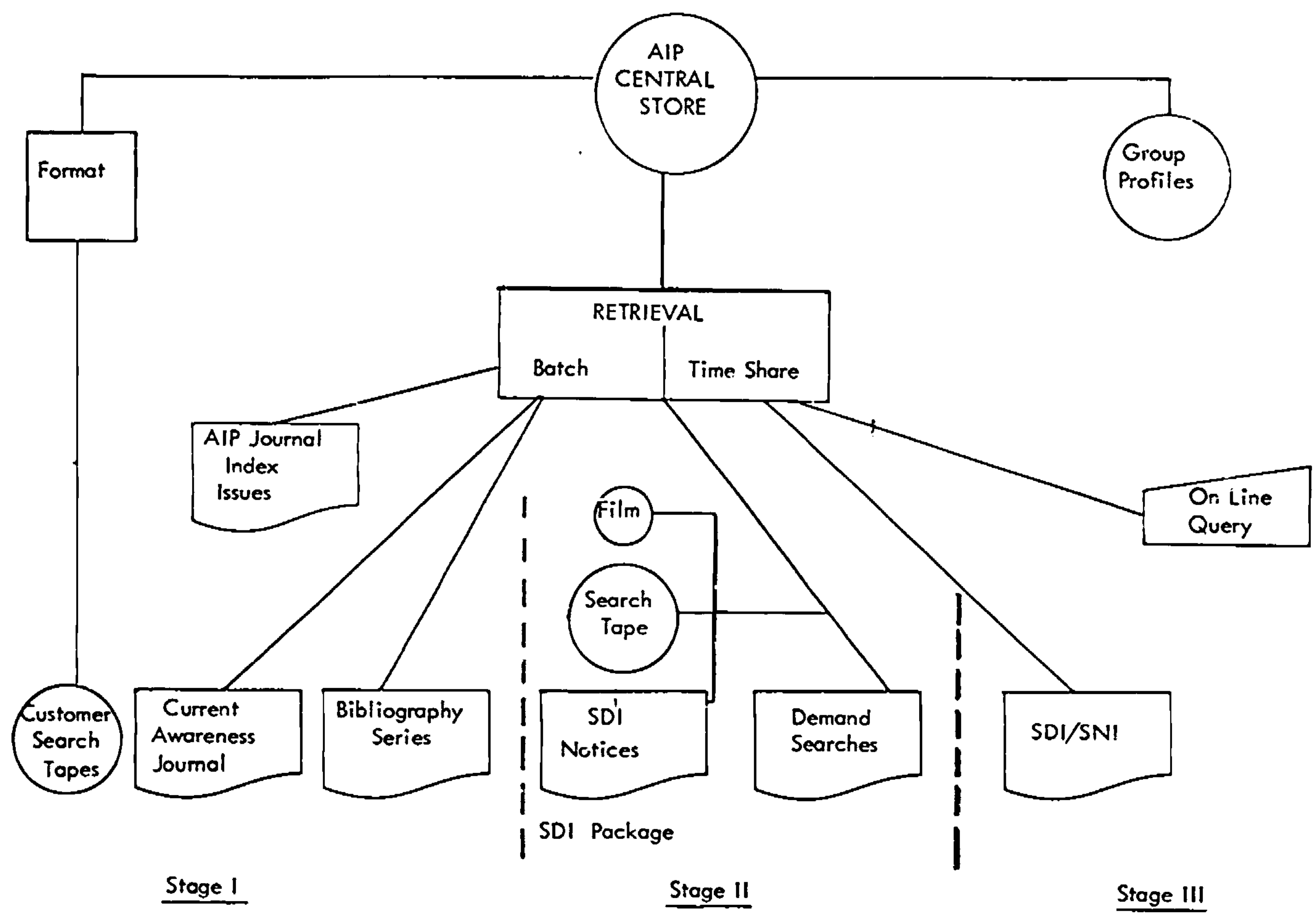


FIG. 2 - AIP INFORMATION STORE OUTPUT

allowed 900 for this item. The AIP's extensive citation listings (items 7 and 8) do not occur in the CA system, but are felt to be of special importance in physics. Alt emphasized the close ties between the primary journals and the Information Store. Because many of the primary journals are also published by AIP, there is a close connection throughout the process of production shown in Figures 1 and 2, which illustrate the flow of input and output, respectively.

Alt emphasized the role of computer composition, both for the AIP primary journals and the Information Store (secondary), and the multi-faceted classification system mentioned earlier. With respect to output, it is expected that computer tapes, for individual search, will be available for customer purchase in 1970. "On-line query" of the Central Store computer is presently available only through the MIT TIP system (to be described later). The AIP approach to increasing the usefulness of magnetic tape services will feature differing programs for user computers of different sizes.

Engineering

Howard E. Tompkins, director of Information Services for The Institute of Electrical and Electronics Engineers (IEEE) offered as background a paper presented at the National Engineering Information Conference in June 1969 on the IEEE Information Program. This paper, describing both the primary and the secondary publishing of the IEEE, is reproduced in Appendix D.

Tompkins discussed some "political" problems of information systems in the general engineering field. He characterized engineering as a loosely interrelated collection of activities with many roots in the physical and mathematical sciences. Because of an apparent need for better coordination and improvement of information resources in engineering, the Tripartite Committee, consisting of senior officers of the United Engineering Trustees, the Engineers Joint Council, and the Engineering Index was formed in 1965 to study the problem. Many task studies were conducted under the aegis of the Tripartite Committee, culminating in a study by the Battelle Memorial Institute, out of which an Action Plan for the Establishment of a United Engineering Information Service (UEIS) was published in 1969. This report asserted that "the most important unfulfilled need in the engineering information community is for a central unifying force to coordinate and enhance the efforts of already established institutions." The plan provided for a budget of $1.2 million a year, to be financed primarily through (industrial) membership fees on a sliding scale of $1,000 - $5,000. If funding commitments were secured during 1970 the operations would commence in 1971. Tompkins stated that it was not, however, at all clear that the plan would be backed sufficiently to get started*.

* Subsequent to the CBMS Conference, the Tripartite Action Plan was rejected by each of the three participating bodies, after several public discussion meetings and many private soundings of opinion indicated that the plan would not be supported, and would probably fail to achieve meaningful operational goals. The Tripartite Committee was dissolved March 1, 1970. The history of the rise and fall of the Tripartite Committee should be studied by all who attempt to get coordinated action on information systems within a diverse professional community--H.E.T.

At the moment, the Engineering Index has the broadest coverage in the general engineering area but has not rallied the entire field around itself as the central unit of an all-engineering system.

In the electrical-electronics area, including many aspects of computer science, IEEE has 160,000 members and has a natural interest in the application of computers to information processing. It has associated itself with the Institution of Electrical Engineers (IEE) of London, England, as a participant in part of INSPEC (Information Service in Physics, Electrotechnology, and Computers and Control). INSPEC publishes three "Science Abstracts" journals: series A: Physics Abstracts (PA), series B: Electrical and Electronics Abstracts (EEA), and series C: Computer and Control Abstracts (CCA). These journals include fortnightly (PA) and monthly (EEA and CCA) Abstracts issues, and semiannual Index issues. Over 100,000 items are covered annually. Cumulative Indexes are separately published as are three companion "Current Papers" journals which publish the full bibliographical references, without the abstracts, in the corresponding abstract journals.

IEEE copublishes EEA and CCA, and CPE and CPC, and handles their sales in the U.S. "Copublishing" means participation in policy decisions, financing, and the development and adaption of indexing and classification schemes. At the present time, 1970, all acquisitions and editorial work, computer processing for composition, and printing, are done in England. The U.S. distribution of PA is handled by AIP. The IEE publishes and distributes all cumulative indexes directly from England.

The three "Science Abstracts" journals are produced by a process of computer-controlled photo-typesetting, and cover the world's periodical, conference, and book literature in their respective fields; they are being extended into the fields of reports and patents. Microfiche editions are also available. The computer data base for Science Abstracts is being made available in magnetic tape form by IEE and in part by IEEE. The IEEE plans to add additional information to the data base in accord with its own indexing practices.

IEEE does the annual indexing of its own journals, because of the time delays at IEE, and because of a desire for deeper indexing. Authors of papers in IEEE journals are asked to contribute self-indexing consisting of (1) classification (in accordance with the printed IEEE-IEE classification scheme); (2) subject index headings and modifier lines (context lines); (3) significant citations for retrieval of related items; and (4) background references needed for reading of the article by a non-specialist. About four citations appear, as suggested by authors, out of an average of a dozen appearing in the articles indexed. (About 15 citations appear in a typical AIP file entry.)

---

* Present annual subscription rates in the U.S. for these journals are: PA, $192; EEA, $156; CCA, $84; combined subscription to EEA/CCA, $192; "Current Papers" in Physics, $28.80; in Electrical and Electronics Engineering, $28.80; in Computers and Control, $24. Microfiche editions are available at the same rates. Double subscriptions, printed and microfiche, are 1.5 times the single rate.

Tompkins commented that the economics of the IEEE information activity were not in balance at the moment, but that there were hopes for improvement over the next few years. Some assistance from NSF is received, and EEA and CCA do pay their way. He noted that international cooperation has its difficulties, but seems to be a valuable activity.

Mathematics

William LeVeque of the University of Michigan, former chairman of the AMS Committee to Monitor Problems in Communication (Com.-Com.), and first chairman of the Commission on a National Information System, described historically the development of information system plans in the mathematical community, as outlined in Appendix A. He then discussed several publication problems the mathematical community has been studying. (1) A revision of the subject classification system for Mathematical Reviews (MR) has been completed. (2) The problem raised by the unwillingness of some institutions to pay page charges for primary journal publications because of restrictions on expenditures under NSF grants is being studied by the AMS Committee on Support of Primary Journals, which will report in the spring of 1970. This Committee is also studying new methods of printing and the costs thereof. (3) The AMS Mathematical Offprint Service (MOS) is in a serious financial position, primarily because of a lack of a sufficient number of subscribers. The AMS will support it with a limited subsidy if NSF will continue to provide back-up support. (4) The next 5-year index of MR will be produced from computer tape.

Commercial Services

Morton Malin, vice president of the Institute for Scientific Information, Inc., described the services provided by his company in the areas of (a) current awareness, (b) selective dissemination, and (c) retrospective retrieval:

Current Contents (CC) is issued weekly in separate issues for each of the seven fields of life sciences; physical sciences (including mathematics); chemical sciences; education; agricultural, food, and veterinary sciences; behavioral, social, and management sciences; and engineering and technology. These pocket-size pamphlets provide photo-facsimile tables of contents in their original format for the current primary journals in the field. The physical sciences series, for example, covers annually some 129,000 articles from 700 domestic and foreign research journals (priced at \$100 per year, but \$67.50 to educational organizations and affiliated persons). A copy of any article from any journal covered in CC can be obtained immediately by first-class mail through ISI's Original Article Tear Sheets (OATS) service (at \$2 for each 10 pages or fraction).

Another ISI publication is the Science Citation Index, which is issued quarterly and cumulated annually. Beginning with a known work, one can find all articles published during the period indexed which cite that work, and from those, other cited articles.

Another service is ASCA, the Automatic Subject Citation Alert, which provides a weekly personal computer printout of bibliographic data on a pre-specified combination of topics--words, phrases, questions--in a personal profile of interest. These can be keyed to titles and/or citations; the charge varies with the average frequency of the required topics within the system. The full articles may thereupon be ordered from OATS as desired.

Malin expressed as his general philosophy that science is interdisciplinary, that there is increasing need for cross-references between disciplines, and that new entrants to the field of science information should make every effort to profit from the experience of others. ISI is now processing some 375,000 source articles per year, retrieving four million articles. Of these, perhaps 20,000 are in mathematics.

Discussion

As leader of the resulting discussion, Richard Kenyon of the American Chemical Society first commented that the NSF aim, as funded by Congress, was to support the creation of a network of systems which would cover all scientific fields and provide adequately for their coordination and for overlapping areas of interest. He indicated that a primary criterion of creation of a system should be its potential as a facilitator of scientific progress, and that another criterion was its economic feasibility. He noted that the ACS had found by experience that the "pure" scientists could not go their own way satisfactorily.

Tate noted the heavy emphasis on citations in the physics literature, and to a lesser extent in the electrical engineering literature. At the present time, Chemical Abstracts does not publish citation indexes, but citations are used for retrieval from the computer tapes.

Tompkins noted that IEEE has its citations stored as separate file items in its data base, and so does not have to store repeats of the citations separately, whereas AIP (following the practice of Project TIP at MIT) has the citation as part of the main storage item. In the IEEE system, file pointers exist in each direction, so reference can be made from the cited to citing item, or vice versa.

Tate noted that indexing costs are much higher for papers which are central to the discipline because they require more cross references than peripheral subjects.

On the question of back issues, Herschman indicated that the half-life of papers in physics was about 3 years (Jackson indicated 5 years in engineering). As a result, older historical papers have not been incorporated into the AIP information file, except in the citations--Alt suggested that these identify the important historical elements. This is also the planned approach in the IEEE system.

Miser emphasized that the user is a most important part of the system, and that its eventual success depends on his response to the system.

In summarizing this session, Jackson noted that the strong emphasis on interdependence, intercoverage, and input economic sensitivity of the existing abstracts services indicated clearly that any mathematics information system that evolved would necessarily have to show detailed interrelationship with those in other disciplines.

III. SYSTEMS CONSIDERATIONS

Thomsen, as chairman of this session, stressed the need for looking at the systems aspects of information handling. He characterized the present "system" as a loosely-connected federation of information services which had grown to meet the internal needs of the mathematical community, with inadequate attention being paid either to the full range of ultimate users or to the interconnections of the various components. He reported deep concerns on the part of young research mathematicians, especially those in applied areas, with their inability to get the information they need from the "system".

R. Creighton Buck discussed specific needs which appear inadequately satisfied by the present information network. He gave as his problem list:

(1) MR is now difficult for even senior mathematicians to keep up with, because of its voluminous contents, and yet these no longer cover the field, especially in applied areas.

(2) The informal networks consisting of a small number of specialists in an area providing each other with preprint information on new developments tend to leave out the younger researchers, as well as the older ones who are in the process of changing fields.

(3) These people are in special need of expository works for those fields not directly within their highly specialized area of training; without these they have difficulty in profiting from MR.

(4) The classification of mathematical ideas is continually changing; is it possible to devise a classification scheme which is (more) time-independent?

Buck noted that concern with the information explosion extends as far back as Warren Weaver's farsighted paper in *Harper's* (1945). He stressed the responsibility of established research mathematicians to the rest of the mathematical and scientific community--particularly the younger scientists and mathematicians who are not members of an older group where information does pass back and forth on a basis of personal acquaintance with those working in the field. Whatever "system" does evolve, the method used to produce that system is most important. He suggested making use of the current "overproduction" of mathematicians to help. He reported a communication from John Kemeny which emphasizes the central role of computers and the evolving technology of miniaturized storage for any forward-looking information system today.

Buck's preliminary model of information flow appears in Appendix E.

He expressed the hope that the mathematical science community could to some extent put aside its various vested interests, recognize that the present publication system is in serious need of change, and solve the general communication problem before it becomes overwhelming.

He called attention to Steenrod's painstaking bibliographical survey* of the field of topology. He suggested as one project, if financing can be

* "Reviews of Papers in Algebraic and Differential Topology, Topological Groups, and Homological Algebra", American Mathematical Society, 1460 pp., 1968. Reviewed in *American Mathematical Monthly*, January 1970, p. 107.

obtained, a computerized approach to a literature survey of a selected specialized field, comparable to this work or the early work of Dickson in number theory. Givens suggested in this connection that maintenance of such a bibliography could be done by individuals with moderate professional competence, even though the original survey probably requires an authority in the field.

I. Edward Block, chairman of the SIAM Publications Committee, discussed the concept of <u>system</u>. He pointed to the Buck model as an example of a system--it has inputs, outputs to users, and feedback from the users; and it also has various functions to perform, such as processing and storage. He continued:

> System design is pretty conventional in concept these days, although in practice it can still be considered an art. In connection with the end objective of this meeting, however, it is interesting to observe that one of the great training needs in industry today is systems analysis and design. Generally it is not done very well--implementation gets behind schedule, design is over-engineered, systems don't serve the purpose for which they were designed, the needs change before the systems are completed. The problems the systems are to solve are seldom apparent--this results from the fact that we as people have trouble defining our needs.
>
> Good design depends very much on a good definition o. the objectives of the system and on recognition of the relation of the system to the environment in which the system must operate. Generally the definition will not be static--people change, needs change, and environment changes. Furthermore, Buck suggested people like to reinvent mathematics rather than refer to a retrieval system. A good systems design must recognize all of this.
>
> I propose we design our system in the same way as a manufacturer designs products. The products must serve a large enough market, and they must be accepted in the market place at a price high enough to pay for operating costs and long enough to pay for development costs.
>
> We tend to focus on the technical aspects of the system when we begin design. I suggest there are market problems of at least equal concern:
>
> -- Do the products of the system have a utility? What are the products? Who wants them and how many "who's" are there? How will the products be used, and is the use lasting? What is the user value and how is it measured?
>
> -- Consider the products in relation to their information content: does the way the information is packaged match its end use from the point of view of the user?
>
> -- What is the effect of the form and format of the product? Consider the type of package in relation to its end use, e.g., the medium, the graphic design, and other physical characteristics?

The technology is here. We can build almost anything we can design, and we can design almost anything. We may have implementing problems, but we eventually get the job done.

The thing we have to be sure about is that when the implementation is complete, there are enough people who want, and are willing to buy, the information products we produce--or should we simply say that the explosion of technical information is inevitable, we must build systems to deal with it, and the Federal government must subsidize what we build.

Discussion

As discussion leader, Treub noted a very useful service provided at the Bell Telephone Laboratories--a biweekly listing of published papers, classified by fields. This service is provided by their library, which would be a very willing customer of an information system which provided such materials. Malin noted that the typical university library cost is about 3% of the university's total budget, and that a well designed information service can and will be purchased by libraries.

Thomsen and Anderson pointed out that a peculiarity of mathematics--as compared with the physical sciences--is that the number of industrial users of research in pure mathematics is very small. Thus the market for information in pure mathematics does not yet have a strong financial base.

Tompkins expressed the opinion that the information problems in the mathematical sciences are reasonably comparable to those being handled by IEEE and AIP, and that perhaps the AIP system might be a good initial model. Others, however, felt that the citation methodology which is a strong component of the AIP system does not play as significant a role in mathematics. Perhaps this is an area worthy of experimentation. Alt indicated that the AIP system could be modified so as to de-emphasize the citations.

IV. INFORMATION PROBLEMS IN APPLIED AND RELATED AREAS

Robert W. Ritchie, chairman of the CBMS New Areas Committee, discussed the areas of the applied mathematical sciences and interfaces with other fields. In these areas the sources and users of relevant information are especially diverse and hard to organize systematically. Therefore, the organization of a useful National Information System will require careful preliminary study to assess relevant sources and user-needs in at least a few representative areas. The CBMS New Areas Committee, after some investigation, has selected a number of areas as being especially suitable for such preliminary study, having in mind the desirability of selecting areas in which: (i) current activity is great, (ii) our member organizations have a direct special interest, and (iii) an expert with a broad professional competence and a deep and lively concern with information problems is available.

Various applied mathematical disciplines are developed to different extents, and some are sufficiently well established to have review journals and bibliography projects which can be utilized to supply much information. In these areas the preliminary study would be aimed at refining existing classifications and designing automated procedures to implement search. In some other areas, such as mathematical biology and mathematical economics, the need seems to be greatest for surveying (evaluating) and even generating written information to be made available to both the mathematicians and the prospective appliers in order to serve as a coherent basis for such a system.

The CBMS New Areas Committee has explicitly identified the areas listed below as being especially suitable for these studies. Initial outlines for projects in some of these areas already exist, and in the majority of cases contacts have been made with leaders in the field and with highly qualified persons who would be seriously interested in working on these projects.

1) Scientific Computing. This has already been discussed by Givens in session I. It is an area in which potential economies are unique: better computing methods can save hundreds of dollars per hour. This is an opportunity to be jointly explored with SIAM and the Association for Computing Machinery.

2) Mathematical Biology. On behalf of William H. Bossert of the Biology Department at Harvard, Birkhoff recommended an exploratory evaluative study in this area. Thrall noted the possible significance of bio-engineering for biomathematics, and said that bio-engineering is five years ahead in information exchange. He called attention to the existence of journals in medical electronics and medical physics.

3) Optimization and Control. This is an area of direct interest to SIAM, which publishes a journal in the field. The present information system needs to be examined from the point of view of the users. It is based on MR, which gives some coverage, and the IEE-IEEE Computer and Control Abstracts, which has developed from Control Abstracts, first published in 1966.

4) Mathematical Physics. This is a vast and historic interface. Herschman noted that this, as well as the applied mathematics of continuum mechanics, heat transfer, celestial mechanics, and astronomy are all covered in major degree by physics information systems. He suggested a joint study with AIP on possible gaps. Jackson noted the need for careful monitoring of interfaces to assure their continued coverage. Birkhoff noted coverage by Applied Mechanics Reviews. Givens noted the developing field of geophysical mathematics.

5) Mathematical Economics. The Econometric Society is developing an information system under an NSF grant. This system indexes biographical and bibliographical material selected by key workers in the field, who are self-identified. It was pointed out that Mathematical Reviews has only minimal coverage in this field (classifications 90A and 62P20*).

6) Information and Communication Theory. It appears that the American Society for Information Science has submitted a proposal to NSF in this area, but cooperation by mathematicians is needed to provide adequately for the mathematical interface.

7) Mathematical Chemistry. This is a well-developed area which was discussed by Tate in session II. Givens noted interdisciplinary problems in computational problems for chemistry and in computerized process control for industrial chemistry.

---

* For the AMS (MOS) Subject Classification Scheme, see the Appendix to MR, Index for Volume 39 (1970).

8) Mathematical Statistics. Geoffrey Watson, representing the Institute for Mathematics Statistics, commented that he finds very useful the IMS 5-year and 10-year indexes of journals. He doesn't particularly like the Mathematical Reviews subject-classification system. He finds an alphabetical classification more useful and suggests the type used in Biometrika. The Journal of Applied Probability has recently begun a new expository section on the subject of "New Advances". He noted that the government Armed Services Technical Information Agency system uses a rather crude sort of indexing. He stressed that statisticians have to work with the literature of many other fields.

9) Logic and Linguistics. Both the Association for Symbolic Logic and the Association for Computational Linguistics publish journals. These are among the 800 journals abstracted in Language and Language Behavior Abstracts, a quarterly published for the Center for Research on Language and Language Behavior by University Microfilms. The publisher provides xerographic reprints of the original articles on order. The interests of these associations should be coordinated with information system planning in mathematics.

Discussion

Givens suggested in these interdisciplinary areas a system of cross-referral by the editors of the review journals with a cross-reference listing. This possibility should be investigated. Tate felt this would require pre-agreement as to who covers what. He suggested joint indexing with, as the most practical method, the assignment for indexing individual journals being made to one group for an entire journal rather than having different topics selected by secondary journals on an ad hoc basis.

Miser commented that each of these fields is generically different from mathematics, in that some of its literature is descriptive of phenomena and is of very limited interest to mathematicians. Its primary interest often lies in its phenomena rather than in its theoretical--including mathematical--aspects.

Ritchie noted that Mathematical Reviews has withdrawn in large part from coverage of the various fields of applications. There are a number of these areas of mathematical applications, several of them relatively new, in which professional mathematicians do not usually work. One reason is that there has been little professional reward for working outside of the central core of mathematics.

V. INFORMATION NEEDS IN MATHEMATICAL EDUCATION

R. D. Anderson read the following remarks prepared by Ralph Boas, chairman of the MAA Committee on the Undergraduate Program in Mathematics (CUPM), who could not attend:

> The information needs of the college teacher, outside of whatever research he or she may do, are different from the needs of the research mathematician only because the college teacher is less well served by existing information services. The college teacher needs

information about books, new and old; journal articles, new and old; and other things, such as what users of mathematics are doing, or what educational experiments are being tried elsewhere.

Thirty-five years ago a competent teacher of mathematics expected to know all the books in mathematics. He or she might not have read them all, but an analyst would know what the best books in topology were, and vice versa. Even as recently as 1947 it was possible for Parke to prepare an annotated bibliography of less than 100 pages covering all of mathematics and physics. (It listed, for example, 19 books on functions of a complex variable, 30 on topology, 44 on quantum mechanics.) I suppose it would take 100 pages to list a single year's production nowadays.

How can one now learn about all these books at every level from elementary textbook to advanced monograph? Book reviews come out late; older books one never finds except by accident. The college teacher has no easy source of information about books.

The same goes for journal articles other than research articles. Expository articles and surveys are dealt with very briefly by Mathematical Reviews (and are scarce anyway). "Classroom notes" and articles on mathematics education are not reviewed at all. No one can watch all the journals that contain articles of educational interest; and the older educational literature doesn't go out of date as fast the research literature--if anybody knew what was in it.

We need more survey and expository articles. We need articles on new ideas in mathematics that help the presentation of standard (as well as of new) topics. (Think of the impact of a good presentation of the vector approach to solid analytic geometry if it had appeared in 1925! Dreadful thought: maybe it did, and had no impact.) We need an abstracting service, preferably with critical comment. It would be useful to have for the college mathematics teacher something like the Scientific American's "Science for the Citizen" column that would call attention to particularly striking publications. (Why, for example, didn't any publication that reaches the college teacher have a short article when Carleson solved the convergence problem for Fourier series, explaining what the problem was and what the new result meant?) A column of this kind would be an appropriate feature of the CBMS newsletter. We need more collections like the new MAA volume of selected papers on teaching calculus. We need a flow of information to the colleges on the changing mathematical demands of graduate schools, of other disciplines, of industry (such as in the CBMS Survey volumes). We need a forum for the exchange of new ideas about courses or parts of courses, without formal articles having to be written. For example, CUPM's newsletter on computer calculus was very effective in getting interested teachers in touch with people who had tried computer calculus.

Leading research mathematicians claim not to need any information services because they have their own (someone comes from Paris to tell them the hot news). College teachers have no such resources; they are pressed for time, and their work could be improved by a better flow of the right kind of information.

Anderson then added thoughts emanating from his earlier experiences as chairman of CUPM:

1. CUPM has served as a primary source of information on curriculum and related questions for the national collegiate mathematics faculties. Its publications have been very widely requested and very widely distributed--in the tens of thousands of copies--especially its General Curriculum in Mathematics for Colleges, its teacher training recommendations, and its newsletters. Its basic library list (now being revised and now also being alternatively modified as a two-year college list) has had wide effect on small college libraries.

2. The changing pattern of employment which is bringing more Ph.D.'s into the smaller colleges will require a modification of the past patterns of graduate education so as to include broader areas of mathematical preparation. These may well be dependent upon better bibliographical information. At the same time, these new Ph.D.'s in small colleges will be quite dependent upon better secondary journals for their access to the literature of all the mathematical sciences and their applications.

3. The rapid and significant changes in research mathematics are forcing more rapid changes in mathematical education. This increases the need for more effective communication.

4. Many potentially useful ideas and suggestions in mathematics education, even though published, are effectively lost because of the current inadequate retrieval of information. This is an area in which the communications system can be vastly improved--for example, through computerized bibliographies.

5. There is a real need for greater exchange of information about curricular innovations which have been instituted in individual colleges--most such activity is known only locally.

6. Much effective communication, especially back and forth communication, takes place at conferences and meetings of MAA and AMS, and it is in the nature of mathematics that active communication is a vital factor in generating concern for innovation in mathematical education. For example, CUPM has had considerable success with conferences on various educational problems; adequate reporting on these is an essential component of an information system.

Mathematics Education in the Schools

In substituting for Jack E. Forbes, on behalf of the National Council of Teachers of Mathematics, C. Russell Phelps noted that there are two information systems now providing services with respect to mathematics education for the schools:

(1) The Education Resources Information Center (ERIC) for mathematics education is now in operation as a part of the Science Education Information Analysis Center at Ohio State University. The Center's Associate Director for Mathematics Education is F. Joe Crosswhite.

The specialized ERIC clearinghouses form part of a national network system collecting, evaluating, abstracting, indexing and disseminating information and resource documents in education research in all fields. The clearinghouses in particular subject areas perform the initial steps of collecting, evaluating, and abstracting the primary and secondary journals and newsletters, and forward abstracts of relevant documents, together with the documents themselves, to a central facility in Washington. The central facility processes these abstracts by computer and prints a monthly acquisition listing entitled Research in Education. The corresponding original documents (if released by the authors) are filed with a contract facility, the ERIC Document Reproduction Service in Bethesda, which will furnish either microfilm or hard copy prints at prices specified in the monthly listings.

(2) The International Clearinghouse on Science and Mathematics Curricular Developments is a joint project of the American Association for the Advancement of Science and the Science Teaching Center of the University of Maryland. It publishes an annual report which indexes and describes, in one or two pages each, the activities in curriculum development, at both school and college levels, in the United States and foreign countries, and including all fields of science, engineering, and social science. The sixth report, for 1968, extends to 480 pages, of which about one quarter is devoted to mathematics.

Communication among mathematics teachers is facilitated by meetings; NCTM sponsors a number of large, regional meetings annually which feature nationally-known mathematics educators. NCTM publishes the monthly Mathematics Teacher (secondary) and Arithmetic Teacher, and the new quarterly Journal for Research in Mathematics Education. The Central Association of Science and Mathematics Teachers publishes School Science and Mathematics. NCTM provides a large number of other publications for teachers. There is, however, no central indexing system of current articles or of expository materials suitable for school use. Another need exists for collections of applications of mathematics and/or indexes relating applications to the school curriculum.

VI. LIBRARY AND PUBLICATION CONSIDERATIONS

Myer Kessler, associate director of libraries at MIT, described the project TIP at MIT, which was designed as an experiment concerned with the information needs of a particular academic community. It is essentially a text management, editing, and retrieval system conducted in conjunction with the MIT computer project MAC, an early multiple access computer. He discussed its chronological development as an internal information system:

1) The program began with the physics journal literature. The first steps involved on-line interaction by physicists with the computer storage involving author search, subject search, citation and bibliographic coupling. This early stage was useful for personal development of the techniques of asking questions.

2) Physicists prepared standing orders for specified information categories--"Self-controlled Selective Dissemination of Information" (SDI).

3) It was found that several groups had the same general objectives and hence they merged their requests for retrieval information. This thus created a "computer-based journal" on a particular subject.

4) Each user had further private sources and comments on those sources which he now has introduced into the system, thus sharing his own informational background.

5) Individual professors seem to use the output for maintenance of their private library collections. The individual user with a specific allotment of computer time in his budget has to decide whether to use his allotment for information gathering or for direct computing on his research problems. This tends to be a controlling factor on how much information he requests. Because the MAC computer is a large installation it is able to handle in a reasonably economical fashion both kinds of activity as a general public utility.

6) The techniques used by professional physicists have been extended to other disciplines. This additional usage adds to the economic feasibility of the system.

7) To facilitate post-retrieval operations the computer can now take the previously selected retrieved information, re-sort it, index it over time or other variables, and file additional comments by the user and his reports on the retrieved material. A printout can then produce automatically additional publications; three such have been produced so far.

8) This is the present stage of the operation. The next steps at MIT involve extension to fields other than physics. It will be necessary to ensure that the same instructions can work on all disciplines. It may develop that a user who has broad disciplinary interests may require a survey of so many disciplines that it becomes non-economical. A question is, can we cut down such a broad survey so that it only searches corners of fields?

9) Another potential usage is the extraction of data from a large file, such as a census, to make a specialized file for further computation, or to publish an abbreviated table.

Kessler discussed the question of cost. The development costs for the computerized retrieval system were supported principally under an NSF grant. Individual users at MIT were paying terminal charges for their own usage. There does seem to be a large differential in costs between deep requests and shallow requests. Traub asked what should be on the computer tapes or discs which are produced so that they are most useful. Kessler suggested that tapes such as those produced by Computer Abstracts, which are very rich in information, are very useful. Traub suggested that an economic factor in developing such a system might be the market for the sale of tapes even though only a small number of people wish them.

Alt suggested for long-range consideration a multidimensional orthogonal classification scheme for bibliographical items in the physical sciences, perhaps with the following dimensions: (1) object studied; (2) method used; (3) application; (4) language of publication; (5) date published. He noted that such a classification precludes linear arrangement in the computer, but that this is all right if adequate programing is available. However, it would be quite wasteful of computer memory because of incomplete usage of many areas of the multidimensional space.

Trotter commented that a new IBM machine allows corrections to be made in punched typescript so as to achieve clean copy automatically, and a similar arrangement should be available for computer manuscripts. As the volume of material grows there is less need for hard copy. Microforms will provide 98 pages of text at $5 for the first copy and 10¢ for each additional copy. Thus microforms would be feasible for small colleges to purchase. Where it is needed, hard copy costs about 4¢ or 5¢ a page. It is also possible for any individual to dial a computer center and get an in-depth computer indexing from which he can order microforms at the nearest library depository. He characterized microforms as providing cheap storage but expensive search, whereas computers provide cheap search but expensive storage. For detailed working use, hard copy would be appropriate.

Givens was unenthusiastic about the present quality of retrieval of mathematical information with respect to such quantities as subscripts on subscripts. Others thought that it might be all right if notations were carefully pre-planned. Tate saw problems, first, concerning copyright if the origin of the material is commercial, and second, in filing and maintenance of a large volume of microfiche. Block questioned whether authors would accept a form of publication which is merely computer storage. This raised the general question of motivation of research. Several people thought that it might be difficult to satisfy authors in this way, though the audience and the university administrators would probably care less. Birkhoff noted that if a researcher is paid $25,000 per year for doing research with an output of (say) 50 pages, this comes to $500 per page for cost of preparation and that 4¢ per page for hard copy seems very reasonable.

Malin called attention to the experiences at NIH with respect to prepublication of research information within the laboratory, where they had established Information Exchange Groups (IEG) which circulated "preprints" of research results within the groups. For the first few years this was very successful. It has been terminated, however, because the various research authors want to have printed publications, whereas the professional journals don't want articles which appear to be secondhand. Implications of this idea for the mathematical sciences are discussed in the final report of the AMS Committee on Information Exchange and Publication in Mathematics [2].

VII. THE ECONOMICS AND MANAGEMENT OF INFORMATION SYSTEMS

Daniel Teichroew, representing The Institute of Management Sciences, described a continuum of subjects which impinge upon mathematics, extending through physics, chemistry, engineering, operations research, management science, quantitative finance, and behavioral science. He stressed that very definite needs exist for information services in the mathematical aspects of the areas from management science through quantitative finance to the behavioral sciences.

Theory of Information Systems

He described the present state of the theoretical basis of information systems, and the process of building management systems. The totality of an information system consists of producers, users, and a processing system. At the processing stage these systems look very similar no matter what the subject. They have the following common characteristics:

a) They all operate with respect to economics over a span of time, that is, money is invested for later recovery.
b) There are many different programs to be run and they have important interrelationships.
c) There are problems with large masses of data.
d) Many people are involved.
e) The capabilities of the hardware are important, because the hardware interacts with the system.
f) The problems are continually changing, and up to one-third of the personnel may be involved in the process of change at any time.
g) The computer is not thoroughly involved yet except in such specialized activities as language compiling.

A new direction would consist of trying to bring computers into a process of continual re-design; a current research project entitled ISDOS (Information Systems Design and Optimization System) has this as its objective. A related research problem is to develop a problem-statement technique.

Jackson commented that there seem to be about 50 different information systems, and IBM is among those working on the problem of developing an art of information systems.

Design of Information Systems

Hugh J. Miser of the University of Massachusetts described the design of information systems. He noted that many systems have been attempted--and many mistakes made--and that the art is still very much in a developmental stage. At the (n)th stage of development a system involves users and if they are reasonably satisfied it is in continued use. At the (n + 1)st stage, we cannot expect a perfect design, but we can have developmental increments over the (n)th stage, with a series of pilot studies of potential design improvements. We might raise the question whether such systems can then be final at the (n + 2)nd stage; the answer appears to be no: the system has to continue to be flexible, so as to respond to new stimuli. New technical developments must be anticipated continually, with cycles of complete change every 8 to 10 years. In short, because the theory is not yet well understood, cyclical experimentation is needed and is going to continue to be needed.

Economic Aspects

Eugene Garfield, president of the Institute for Scientific Information (ISI), discussed some of the economic aspects of information systems. He suggested that the political status of the support of the scientific enterprise was very difficult to separate from the economic aspects. When basic research is being supported to the hilt, much experimental aid is available even though the information explosion is severe. However, when support is not quite so easy to get, as at the present time, there are questions as to how much the producer community and the user community can pay for experimentation in information systems. He strongly suggested making a study of the user market. He observed that all societies think their problems are different, but he doesn't think that they are quite that different.

He raised the general question, if individual mathematicians (perhaps a thousand in number) were subsidized at $100 to $1,000 each, presumably by the government, in order to buy the information services they decide they need, what would happen? He suggested that under such circumstances the suppliers of information services (especially the non-commercial ones) would then develop systems which are user-oriented, perhaps more so than at present.

Garfield noted that the literature on mathematics is not at all equivalent to the literature of interest to mathematicians. His organization therefore feels that an information system has to be completely multidisciplinary. He noted, too, that many significant contributions to information science had been made by non-automated systems. Some questions which should be considered are: Is an abstract necessary? Is a title plus available reprint better? Each of these methods (abstract or title plus copy) has its supporters, and some people like to have both. He also called attention to the usefulness in some fields of citation indexing and noted the utility of this method of citation networks in the study of the history of science.

Discussion

In the general discussion it was observed that lawyers expect to spend up to $1,000 a year on legal information, books, etc., and that the "free service" concept exists primarily in the academic world. Industrial employers normally expend large amounts for informational services, including their own library services. It was suggested that it might be appropriate to make a specific study of who is to be asked to pay for the usage of an information system.

It was suggested that it might be possible to prepare a journal of abstracts in such a way it could be split into parts centered around certain broad classifications. It was pointed out, though, that the income to the journal would go down while the abstracting expenses would be the same. Since the abstracting costs are much larger than the printing and paper costs, this does not appear to be a desirable approach. Attention was called to the card abstract service provided by the Engineering Joint Council as another solution to this problem.

LeVeque described the present cost problems of the Mathematical Offprint Service (MOS) operated by the AMS. It presently serves some 1,200 people who average less than 50 requests per year. The AMS is willing to provide an annual subsidy of about $20,000 provided that the remainder of the deficit in operational costs (presently over $10,000 per month) can be covered by a grant from NSF. Suggestions from others which might be considered to help balance the cost of operations were (a) that the offprints be furnished directly by the societies which publish the primary journals, (b) that standardized user-packages be offered in various categories rather than individual-profile service, or (c) that the user charges be set higher for individual-profile service.

The conferees agreed that studies of review and abstracting journals and dissemination services should take into consideration the experience of commercial enterprises as well as the experience of other societies.

VIII. PROPOSED *NISIMS* ACTIVITIES

Birkhoff reviewed CBMS' initial planning for NISIMS, as proposed to NSF.

The immediate purpose of the CBMS proposal is to formulate an information-system concept appropriate for the mathematical sciences and acceptable to the mathematical community as represented by its professional societies. In general, CBMS would view agreement on a well-defined system concept as the first stage in a projected multi-stage effort, looking toward subsequent program definition, acquisition, development, and implementation. The present information system in the mathematical sciences is, however, somewhat unevenly developed, and CBMS feels that in two ways this fact justifies more special mathematical-information efforts, concurrently with the overall-planning activities here proposed.

On the one hand, the community of core mathematics has already evolved, over a period of years, information services of very considerable sophistication and effectiveness, and CBMS certainly supports as worthwhile such clearly valuable and well-conceived experiments to improve these services as the Mathematical Offprint Service. On the other hand, under the present system the information needs of applied and interface areas are much less well served, or even understood, than are those of core mathematics. Thus, while the present proposal is confined to formulating the conceptual development of an appropriate overall information system for the mathematical sciences, CBMS views it as important for development of the overall system to proceed with studies in various applied and interface areas as soon as funds for these are available. CBMS and several of its member-societies have already devoted considerable thought and planning to such studies.

Development of a plan for a national information system in the mathematical sciences should start initially with an analysis of the present flow of mathematical information with respect to volume, producers, users, existing procedures for transmittal, storage and retrieval, secondary services, etc. The initial analysis should entail an examination of current journals, meetings, so-called "invisible colleges", reviews, abstracts, special user-oriented services--in short all the recognized means of transfer of mathematical information. Naturally, advantage would be taken of the valuable studies and analyses already made, such as the AMS report on its preliminary studies [4]. It is anticipated that additional work will be required to identify, characterize, and gauge the full range of producers and users of mathematics, including professional engineers and scientists outside of the mathematical sciences.

Some specific (overlapping) questions that need to be answered include the following. What are the information services that constitute the existing system? What parts of the broad mathematical community, including interface areas, are served by the existing system and in what ways? How is this system sustained financially and by whom? In particular, how much do the various parts of the mathematical community presently pay for the information services they receive? How much of these services is supported by subsidies, and to what extent and length of time are they dependent upon the subsidies? What estimates can be made as to what mathematicians and users of mathematics might be willing to pay for information services improved in clearly defined ways?

Once the present "state" is known with reasonable accuracy, some assessment should be made as to its effectiveness, particularly with respect to costs and the possibilities for improvement. Here again there are a number of (overlapping) questions of importance. What are the possibilities and prospects technologically for better designed systems providing broadened and improved services, taking into account new data storage and transmission devices and all the rest of the manifold capabilities of modern computers? What changes, if any, would permit the present system to serve the same clients more effectively and perhaps more cheaply? How could the system be broadened to serve clients it should serve but at present does not? What additional funds would such new clients bring into the picture that might sustain, or at least help sustain, these broadened services which would serve their interests?

One study will of necessity address coverage in the reviewing and abstracting journals with respect to particular branches of the mathematical sciences. Initial attention would be concentrated on analyzing and correlating the existing media such as Mathematical Reviews, Computing Reviews, the review section of the Journal of Symbolic Logic, International Abstracts in Operations Research, and other reviewing and abstracting journals. For all but core mathematics, this would involve information not provided by professional societies in the mathematical sciences *per se*. Concern has already been expressed in several quarters with regard to the fragmentary reviewing and abstracting coverage of applied mathematics which exists at present. In addition, little serious effort has been made to improve the interdisciplinary aspects of mathematical information services over the past five years. On the contrary, the coverage by Mathematical Reviews of mathematical physics, in particular, has diminished by a factor of five during this time period. The facts here should be carefully looked into before any detailed steps are taken to plan for an overall information system in the mathematical sciences.

Another study would call for further investigation in the effectiveness of the current system, studying the probable outcomes of current trends, with a view toward substantial beneficial modifications, or conceivably toward a complete change in the mode of operation (in response, for example, to technological advances). If new concepts are envisaged, they could relate to the actual existing means of information transfer (new devices to do the same thing), or to the introduction of some new and more efficient total system which could perform the existing information transfer and, in addition, do new things to serve the mathematical community--in other words, make the mathematical sciences more readily available, to both the mathematician and the non-mathematician. In this connection, modern computing systems with their on-line rapid input/output capability and their interactive features might very well eventually assume a central role.

This broad plan would envision close working arrangements with the other sciences and with engineering, taking full advantage of techniques developed for processing and making available information in these other fields which might also be useful in the mathematical sciences. Overlapping areas between the mathematical sciences and physics, chemistry, engineering, etc., should be closely studied with a view to achieving desirable compatibilities. Also, any overall system developed under the auspices of CBMS would clearly have to be coordinated with any developments considered by the American Mathematical Society for core mathematics.

Birkhoff raised the question as to whether these objectives were reasonable and attainable, and asked for additional suggestions not already covered in the preceding deliberations.

Tompkins commented that it might be possible to approach the various special activities in many applied and related interface fields in a more unified fashion; as a non-mathematician he felt that the differences in methodology might be of second-order, or that studies in these various fields might be amalgamated into a relatively small number of coherent areas.

Block suggested that, while systems design is important, the study of concrete issues might be even more important; he cited the problems inherent in marketing the output as a major issue. Miser agreed, suggesting that it is important in any such study to get close to the user point of view with some experiments of limited size and expense.

The conferees expressed their general agreement on the objectives of the proposal, as well as on the increasingly urgent need for an improved system of communication for the full range of producers and users of information in the mathematical sciences.

BIBLIOGRAPHY

[1] American Institute of Physics, A Program for a National Information System for Physics 1970-1972, 55+107pp. 1969.

[2] American Mathematical Society, Committee on Information Exchange and Publication in Mathematics, Final Report: see The Mathematical Sciences: A Report, National Academy of Sciences, 1968, Appendix B, pp. 233-242.

[3] _____, Final Report: Conference on Communication Problems in the Mathematical Sciences, December 5-7, 1967, 22pp. 1969.

[4] _____, Report on a Preliminary Study Leading to the Design of a National Information System in Mathematics, Volume I: Results of the Study, and Volume II: Guidelines for a Design Program, 55+34pp. 1969.

[5] Chemical Abstracts Service, Report on the Twelfth Chemical Abstracts Service Open Forum, New York, September 7, 1969, 22pp. 1970.

[6] National Academy of Sciences, Scientific and Technical Communication. A Synopsis of the Report of the Committee on Scientific and Technical Communication of the National Academy of Sciences-National Academy of Engineering, 30pp. 1969.

Appendix A

HISTORICAL DEVELOPMENT OF INFORMATION SYSTEM PLANS IN THE MATHEMATICAL COMMUNITY

Prior to 1965 the mathematical community had devoted little systematic consideration to its problems of communication. In that year, responding to a growing concern over the increasing volume, expense, and delay in mathematical publication, the American Mathematical Society (AMS) appointed an ad hoc Committee on Information Exchange and Publication in Mathematics. This Committee endorsed wider use of prepublication, suggested several experiments in abstracting, indexing and information retrieval, and recommended the formation of a continuing committee to monitor problems in communication. It also recommended greater attention to the needs of applied mathematics and interface areas through such devices as conferences and expository writings, and it made a number of suggestions for improving the effectiveness of mathematical meetings generally. Its final report was published as an appendix to the COSRIMS report [2].

In mid-1967 the AMS set up the recommended standing Committee to Monitor Problems in Communication ("Com.-Com."). This new Committee recommended and monitored several experiments and innovations in information services, and AMS has been engaged in a series of experiments on photocomposition, automatic translation from oriental languages, alternatives to conventional manuscript-journal forms of research publication, and other innovations in the processing and dissemination of mathematical information.

With NSF support Com.-Com. sponsored, in December 1967, a Conference on Communication Problems in the Mathematical Sciences. This Conference, whose 22-page Final Report was issued as a separate pamphlet by AMS [3], was especially valuable in focusing attention on a number of specific problems, such as the need for expository journals, the possibilities in author-prepared abstracts, better communication between mathematicians and the users of mathematics, and international cooperation in reviewing and abstracting. The discussions brought to bear the ideas and experiences of thoughtful spokesmen, not only from both pure and applied mathematics and statistics, but also from various physical sciences, the National Science Foundation's Office of Science Information Service, the National Research Council's Committee on Scientific and Technical Information and several foreign and international mathematical organizations concerned with problems in information exchange. In mid-1968 AMS received from NSF a $64,150 grant for preliminary work leading to the design of a national information system for mathematics, the study to be carried out by a Commission broadly representative of mathematical professional organizations, and work began with the formation of the Commission.

Consisting initially of the AMS members of Com.-Com. plus representatives from five other societies, the Commission had, by the fall of 1968, grown to a body of 23 individuals representing eleven professional societies in the mathematical sciences. Guided by a five-man Steering Committee, and with staff assistance from AMS headquarters in much of the actual writing, the Commission produced, around May 1969, a two-volume Report [4]. Volume II contained descriptions and detailed budgetary estimates (totalling $1.2 million) for a proposed two-year program of studies and experiments beginning around January 1970. Because of mixed reactions to this Report on the part of several information-systems experts who read it, however, this time schedule was abandoned; and in June 1969 the Commission drafted a $100,000 proposal, to have been submitted by AMS to NSF, for interim support for calendar 1970, during which a revised proposal for a two-year program of studies and experiments could be prepared.

These plans too were changed when the AMS, through action of its Council in August 1969, decided to withdraw its representation from the Commission at the end of the current period of funding, and to confine its further proposals to NSF to the sponsorship of information activities in its own immediate area of interest and responsibility rather than for the design of a broad national information system in the mathematical sciences.

Following this AMS action, CBMS addressed to the Commission an offer to sponsor, with Commission advice and counsel, continuing work toward a National Information System in the Mathematical Sciences (NISIMS). This offer was accepted at the Commission's August 1969 meeting, and it is the ad hoc NISIMS Group appointed by new Commission Chairman Robert Thrall which planned the Harrison House Conference and which prepared the CBMS proposal. That group consisted of Thrall, chairman, R. Creighton Buck of the University of Wisconsin, Jack E. Forbes of Purdue University, J. Wallace Givens of Argonne National Laboratory, Robert W. Ritchie of the University of Washington, Donald L. Thomsen, Jr., of IBM Corporation, and Garrett Birkhoff and Truman A. Botts, ex officio, CBMS. The stated objective of this proposal was to formulate an information-system concept appropriate for the mathematical sciences and acceptable to the mathematical sciences community as represented by its professional societies*.

* The NSF Office of Science Information Service awarded CBMS in May 1970 a one-year grant of $49,900 for Initial Planning toward a National Information System in the Mathematical Sciences (NISIMS). Under this proposal, a small group of consultants--information systems experts and mathematical scientists--will make a preliminary delineation of the existing information system of the mathematical community, and make a critical evaluation of this baseline system, identifying gaps in needed data concerning both the system and its community of users. This group will then develop a plan, which will be reviewed by an advisory committee, the CBMS NISIMS Committee, of which Thrall is chairman; other members are Forbes, Givens, Ritchie, and Thomsen from the ad hoc group; William LeVeque of the University of Michigan, Alex Rosenberg of Cornell University, Joseph F. Traub of Bell Telephone Laboratories, and Eric Weiss of the Sun Oil Company; and Birkhoff and Botts, ex officio.

Appendix B

TOWARD AN INTEGRATED INFORMATION SYSTEM

Arthur Herschman
Information Division, American Institute of Physics

I. The Nature of Scientific Information.

Scientific Informatio is the set of messages, transmitted or received, oral or written, in a particular science. It should not be confused with scientific knowledge which is the result of a scholarly analysis and synthesis of various messages and the ensuing "public consensus" (as emphasized by John Ziman) among the community of knowledgeable scholars as to its acceptability.

Information goes through various stages beginning with its initial formulation through the traditional trial and error processes of established (or extensions of established) methodologies (each set of methodologies being the hallmark of some small subdiscipline of science). The information enters the system usually through some informal channel such as a conference, meeting, or preprint generally directed to an audience working in the same subdiscipline in which it was found. It then goes through the process of becoming more formal (as elucidated by William Garvey and collaborators) and entering into the "public archive" in the form of publication in a primary journal. This is the first stage in obtaining a "consensus" since it involves the agreement of editors and referees.

The next stage is one of scholarship in which the results of various messages are distilled into the form of reviews, compilations, etc., resulting in a great compaction of the bulk of the primary literature (as emphasized by Conyers Herring). The final stage in the transition from raw information to scientific knowledge is achieved by the embodiment of the results into the textbooks and handbooks of the discipline.

This process of the evolution of information is shown schematically in Fig. 1 (with the approximate time scale as determined by Garvey). Also shown on this figure is the development of information about the messages, what is called secondary information, procedures developed to help control the information flood.

Fig. 2 lists the various forms of information messages including primary (original information), secondary (information about primary information), and tertiary (distillations of the primary information) according to whether it is published (publicly available), quasi-published (available if one knows where), or unpublished (mostly unwritten).

II. The Producers and Users of Scientific Information.

A scientist generally has a double allegiance; on the one hand, to the organization which employs him and thereby to the mission of his organization or organizational unit and on the other hand, to the discipline in which he claims expertise. Although a scientific fact may be discovered while working in a mission-oriented context, information about it is generally reported in

Fig. 1. The Evolution of Information

INFORMATION	Time(yr)	INFORMATION ABOUT INFORMATION
Start Research	0	
	1	
Local Colloquia Special Group Meeting		
	2	
Convention		Convention Bulletin
	3	
Preprint Distribution Submission to Journal		
	4	
Journal Publication		Current Awareness Bulletin
	5	
Reprint Distribution		Abstract Journal
	6	
Review Article		
	7	
Textbook		

Fig. 2. The Kinds of Information Messages and Information Channels

	Published	Quasi-Published	Unpublished
Primary	Journals Monographs	Reports Preprints Patents	Meetings Seminars Letters Conversations
Secondary	Abstract Journals Current Awareness Journals Bibliographies	Internal Alerting Services Tape Services	Letters Conversations
Tertiary	Reviews Compilations Monographs & Books	Reports Information & Data Centers	Lectures Conferences Symposia

the context of a discipline, if this is possible. The major reason for this is that the disciplinary group involved plays the role of the custodian of the "public archive" in that area and it is only when his work is entered into this archive that appropriate credit for the work done can be claimed. This priority-recognition (as Robert Merton has emphasized) is largely the "currency" with which his activities in his second allegiance are compensated.

Information is entered into the system according to a discipline orientation, largely at the choice of the producer and it is generally processed by the system along these same lines. Traditionally, information systems have been producer oriented. The contents of journals are largely determined by the authors who submit manuscripts to them and not by the subscribers. This is especially true for reports and preprints, since they do not undergo the leavening effect of editors and referees. The contents of scientific meetings are mostly determined by the speakers, etc.

Information is taken from the system still largely in the same discipline in which it was entered; however, it is also required by neighboring disciplines as well as various missions. The scientist as user requires information for both of his allegiances and it is here that the producer-orientation of the information system thwarts him. Although customs developed at an earlier time allowed users to cope with this producer-orientation of the input by hit-or-miss scanning of what was available, the much publicized exponential growth of the literature has made it increasingly difficult to find relevant information, let alone digest it. The problem does not seem to be due to an increased productivity of individual scientists but rather to an increase in the number of scientists working in any one field, no matter how small, as well as to the increasing need, as specialties become finer, of being aware of information in peripheral areas.

The information problem which has motivated to many of the new designs for information systems, is largely that of the scientist as user rather than that of the scientist as producer of information.

III. The Functions of an Information System.

The Information System which we have been discussing can, of course, be regarded as the collection of the information messages, the channels through which they are transmitted, the scientists who produce and utilize them, the institutions which process them, and, above all, the set of behavior patterns, customs, and traditions by which all of these interact. The part of the system which we wish to focus on here is the system dealing with the transmission and processing of messages. And as we have noted earlier this system has been, by and large, a passive one, tending to preserve the producer-discipline orientation of the input throughout its stages. It has also tended to be a fragmented one, lacking integration along all of the axes of Fig. 2, i.e., between primary, secondary, and tertiary; between formal and informal; and among the various disciplines and even among the subdisciplines within a given discipline. This lack of integration, due, in part, to the passive nature of its processing, has not served its users in achieving what may be called the larger goal of an information system, viz., the facilitation of the transition from information to knowledge.

The basis for the information system which we want to discuss has been the institutions which maintain the channels for transmission. Foremost among these, since the inception of modern science in the 17th century, has been responsible for the "public archive" of primary information (the primary journals) as well as convening the "congresses of scholars" in the forms of scientific meetings. Second only to this has been the great libraries which have been the repositories of the archives of scientific information as well as scientific knowledge. Traditionally, the connections between these institutions and the channels which they provide, could be found only in the habits and customs of their users.

In the last hundred years we have seen the development of the secondary information service, in part, as a hope of supplying some of the needed integration. This service, which is primarily based on an abstract journal, adds indexing information to the abstract of what is included and makes copies of this information available to the potential information users. At an earlier time, when users were still able to scan the contents of such a product and thereby supply the missing links, this was reasonably successful; of late it has not been. Recently the general trend, to cope with the sheer size of the information store involved, has been to computerize the secondary services, both in order to produce the original products more efficiently and also to provide the capability of other types of searching of the data base.

In more recent years there have been several other trends which are worth noting. These include the development of mission-oriented information systems, generally under the auspices of various mission-oriented agencies. Also, the tendency of commercial processors of information to eschew their traditional roles of passive producers of the archive, in favor of the more active role of exploiting these parts of the archive which have the greatest financial returns, generally those having the greatest impact on the activities of some well-financed mission, and naturally leave the rest to the scientific society.

Although many of the newer developments in information systems have been designed to bridge some of the gaps which we have discussed and many of the experimental programs of some of these systems appear quite promising, all of the systems are still quite far from being integrated. What I would like to discuss now is how an information system could be integrated.

IV. Integrating an Information System.

The present information transmission system has been based on the three independent operations of the: secondary service, the primary publisher, and the library, with the secondary service attempting the role of a control point. The system has not been successful; secondary services have been traditionally based on Abstract Journals, i.e., abstracts of primary information published in groups according to disciplinary categories and supplemented by annual indexes using more extensive terminology. At best, the abstract supplies a very tenuous connection with the original document and categorization and indexing suitable for printed lists are rarely optimal for computer searching, particularly for interdisciplinary or even inter-subdisciplinary purposes. As the bulk of the literature has increased, the size of abstract journals has also increased until the point has been reached where individual users can no longer purchase them and can barely use them, since it is no longer possible for users to scan the great bulk and supply the relevance links themselves.

What is needed first is an entity which takes a more active view toward the role of secondary information processing than does the traditional secondary service. I have termed such an entity the "Integrated Information Control Center" (which is at least acronymically a higher powered Information Center). The major function of such a center would be to monitor the interests of user groups in the discipline involved and to devise and operate means for manipulating the secondary file so as to tag appropriate documents for the appropriate groups. This active manipulation might necessitate devising new types of indexing systems or new ways of using other "handles" in the document, e.g., the citations or references which the document makes to others.

The Center would be linked to the primary-journal editor and perform its analysis on the manuscripts before publication. The editor would have on-line access to the central file, augmented by appropriate private files, in order to expedite referee selection and manuscript control, thus speeding up the publication process, through this new primary-secondary link. Speed is important in information handling, since the commodity, by its very nature, is highly perishable.

A second primary-secondary link would be effected by the tagging of manuscripts according to potential user interest by making it possible for publishers to repackage the primary-archive journals into smaller journals designed for specific user groups. A given archival article may appear in several such "user" journals. It is my belief that the operation of these links and other means for speeding up the publication process, would remove most of the pressure for preprint distribution and similar phenomena.

An important link to the unpublished communications could be supplied by such a Center by having its professional staff produce "topical status reviews" and bibliographies in place of or in addition to the proceedings of conferences. In a sense, staff members would act as professional reporters at conferences.

An important function of the Center would be to tag manuscripts for use outside of the discipline involved and to effect transfer of secondary information between the Centers of different disciplines and missions, i.e., to be part of a "network" of Centers, thus effecting interdisciplinary integration. This network might include the local, institutional, centers who would have on-line access to the central file.

The Center itself, with appropriate library backup, or the local centers would then be ideal places to house itinerant scholars who would prepare review articles or data compilations and thus provide a secondary-tertiary link.

Fig. 3 shows some of these ideas in schematic form (see also Koch and Herschman, <u>A Network for Physics Information</u>, AIP Report ID 68-13).

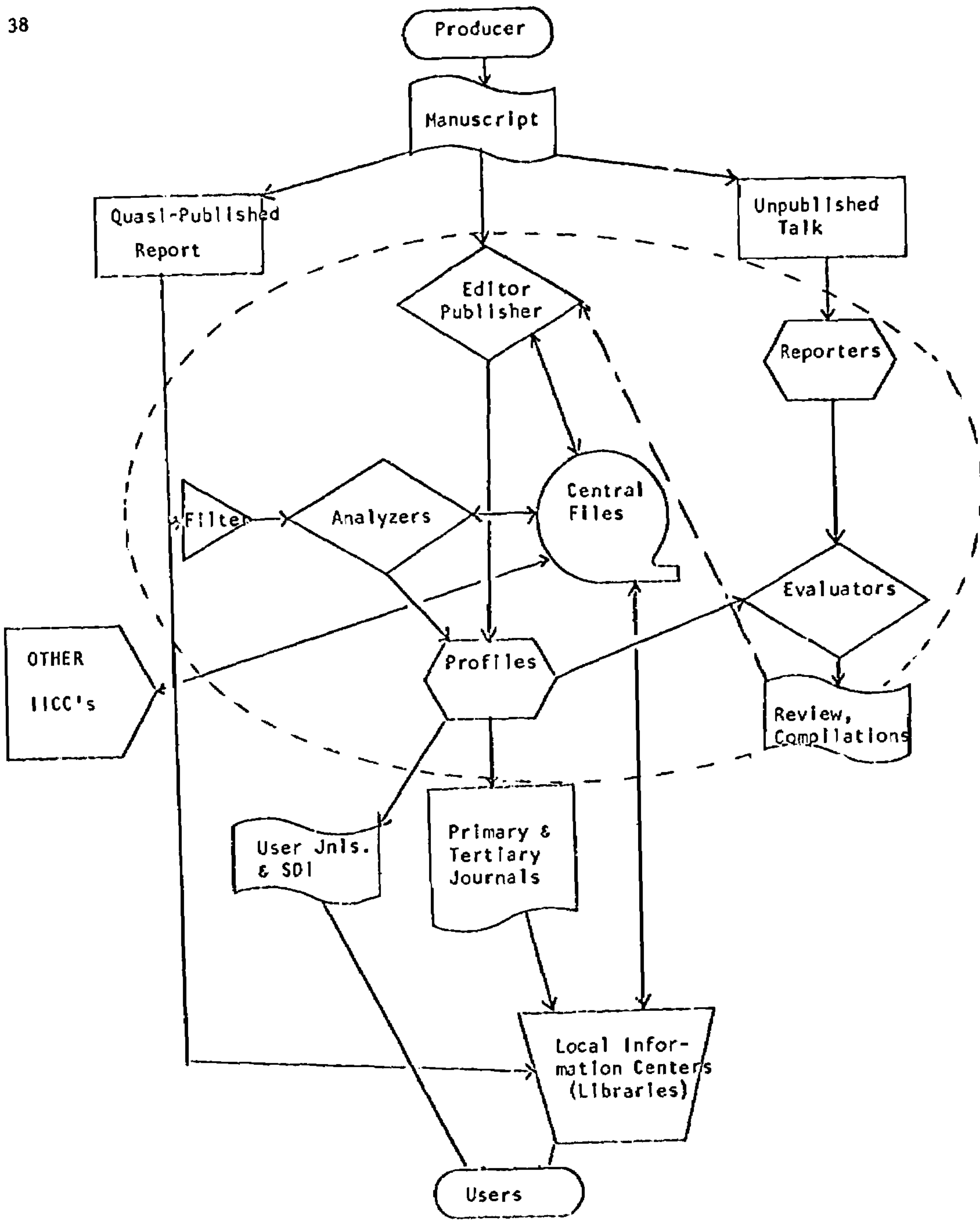


Fig. 3. The Role of the Integrated Information Control Center

Appendix C

NEW INFORMATION PROGRAM FOR AIP

How do you cope with the ever-increasing flood of literature? A new computer-assisted system will offer new and better ways of obtaining physics information. We seek your opinions.

ARTHUR HERSCHMAN, FRANZ L. ALT and H. WILLIAM KOCH

THE AMERICAN INSTITUTE OF PHYSICS, with support from the National Science Foundation, is currently engaged in a major effort to develop and implement a computer-assisted "National Information System for Physics." Designed by physicists, for physicists, the new system is scheduled to begin pilot operations early next year. We at AIP believe the system to be urgently needed, but how do you, the physicists, feel about it?

This article presents a description of the main features of the new system so that you can form an opinion on its merits and its potential usefulness. After you have read the article, we hope to hear from you on this important question: Do you feel there is need for this program and is it aimed in the right direction?

AIP responsibility

Because we believe that there is a need for a physics-information system and that AIP is the logical place for its development, we have assumed the responsibility and undertaken the formulation and development of a new system.

AIP was founded in 1931 as a federation of leading societies in physics to serve those needs of the physics community that could best be fulfilled by the societies jointly. It presently has seven member societies, whose 47 000 members are also institute members, plus 19 affiliated societies with an interest in physics, 150 corporate associates and a Society of Physics Students. The institute services for this sizable community run the gamut from publicizing physics and physicists, strengthening educational programs, documenting the history and development of physics and representing physics nationally and internationally—to the largest single publishing effort for physics in the world. AIP publishes 16 archival journals, comprising 25% of the world's articles in physics and translates 13 Russian journals, for an additional 10%.

As a natural extension of its responsibility and in accord with its mandate to engage in activities "for the advancement and diffusion of the knowledge of the science of physics," AIP, with support from NSF, has been actively planning further information services since mid 1966.[1] These plans led to our program for the design and development of a national information system for physics. At the end of 1967, a new division was organized within AIP to handle the project.[2]

The division has the assistance of a 15-member advisory committee, which was appointed by AIP member societies, of about 100 physicist-respondents selected by the advisory committee and of liaison members from other interested groups, both from related scientific societies (chemistry, mathematics and engineering) and from interested government agencies. The results of this effort, a national physics-information system, will be ready for implementation during 1970. (A document describing the proposed system was recently presented to NSF in support of a request for funding the pilot operations.)[3]

Why a new system?

A new system is needed to cope with the exponential growth of physics literature, which has been doubling about every seven and a half years. It

Arthur Herschman (left) has been director of the AIP information division since its inception in 1967. A theoretical physicist, who received his PhD from Yale University in 1954, Herschman was formerly coeditor of *The Physical Review*.

Before becoming AIP director in 1966, H. William Koch (center) was chief of the radiation-physics division at the National Bureau of Standards. Koch joined NBS in 1949, after receiving his PhD from the University of Illinois, and worked in the high-energy-radiation section until becoming division chief.

Franz L. Alt (right), who took his PhD in mathematics at the University of Vienna, became deputy director of the information division after 19 years with the National Bureau of Standards. At NBS he was assistant chief of the applied-mathematics division and, later, area manager for information systems, design and research.

is not that physicists are writing more, but that more physicists are writing—more physicists in every speciality. In 1968 alone, over 50 000 research papers were published in more than 500 journals. Finding information in the traditional way, by scanning journals and through formal and informal talks at meetings, is no longer practical. Although one may still keep up with new developments of immediate interest, it is almost impossible for any one physicist to keep abreast of bordering areas and related specialities with which he should be familiar.

The tendency of the present procedures to be designed for authors' convenience has also aggravated the problem. Authors, not readers, determine when, where and how the information is presented. As a result, papers on any given subject are dispersed over many journals, and a single journal may contain, side by side, papers on widely different subjects. The reader is left to cope with the flood as best he can, which all too often results in information coming to his attention too late for his need.

What is clearly required is a better way to organize and manage the information so it can be routed more accurately and efficiently from author to reader.

Computerized file

The only feasible way to organize and manage a collection this large is by computer—to have a computer record of each new paper that allows a file to be organized and searched on a current basis according to physicists' interests.

As presently conceived, the file would initially contain records for about one half of the world's physics-journal articles, but would be expanded to cover almost all journal literature, as well as nonjournal material, in the not-too-distant future.

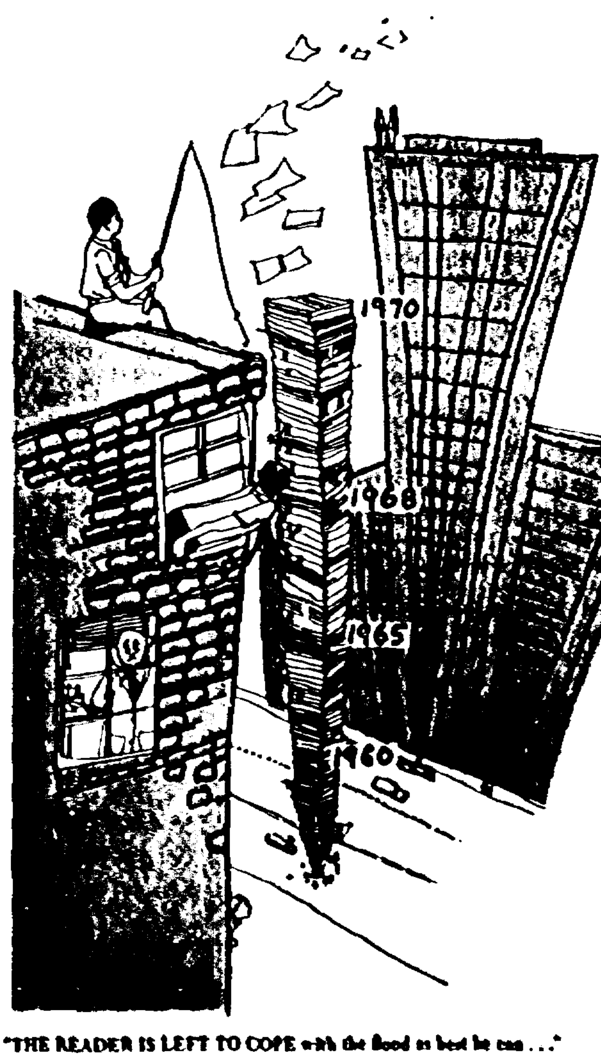
Every month, for AIP-published journals during the prepublication cycle and for other journals as they are received, the information-division staff will prepare, for each new paper, a "record" that contains basic information about the paper: author, journal, title, abstract, citations (that is, references to other literature), a list of "key words" and a special "AIP classification number."

These records are then transcribed onto magnetic tapes. Thus the cumulative file of all such records constitutes, in effect, a machine-searchable "physics almanac" that can be queried for a multitude of purposes and produces a variety of services. Specially formatted printed versions of all or part of the file can be widely distributed for ready reference. The file itself could answer specific questions both at AIP and at suitably equipped subscribing institutions.

Each item in the article record represents a "handle" that can retrieve the complete record. Thus one can ask for all articles published in a given journal or year or by a particular author or institution; papers that contain certain specific words in their title or abstract or that cite another paper or have a number of citations in common with a given paper; and finally, papers about a particular kind of physics.

Classification scheme

The classification procedure that the system as a whole will use was developed by AIP in coöperation with outside physicists specializing in various branches. It is a procedure for writing a formatted statement of what, objectively, a paper is about. The box on page 31 is an example and shows how the classification number is constructed using "verb expressions," represented as integers, followed by "nouns," represented as decimal num-



"THE READER IS LEFT TO COPE with the flood as best he can . . ."

bers. Typical verb expressions are: "the subject of primary interest is . . ." and "the method used is . . ." and "the host or environment is . . . ."

The more digits in a numerical noun, the more specific its meaning. For example ".2" means *particle*, ".28" means *hadron*, ".282" means *baryon*, ".2821" means *nucleon*, ".28211" means *proton*.

This particular way of "spelling" nouns has the advantage of exhibiting the word roots. Suppose, for instance, you want papers on hadrons. All nouns beginning with the digits ".28" belong to the class hadron. Knowing this, a request can easily be formulated. The same request, in clear language, would require a specification of all words included in the class hadron (meson, pion, kaon, baryon, hyperon and nucleon). In the example (right) the title is less explicit, from the viewpoint of information retrieval, than is the classification number

CLASSIFICATION EXAMPLE

For the paper "Evidence of Quarks in Air-Shower Cores"
0.1; 1.271; 2.9534; 4.24; 6.29

0	The document type is . . .
0.1	experimental;
1	The subject of primary interest is . . .
1.2	particle physics,
1.27	more precisely, a particle property,
1.271	specifically, its existence;
2	The method used is . . .
2.9	a technique,
2.95	more precisely, a particle technique,
2.953	still more precisely, a detection technique,
2.9534	specifically, track visualization;
4	The entity of primary interest is . . .
4.2	a particle,
4.24	more precisely, a hypothetical particle;
6	The host or environment is . . .
6.2	particles,
6.29	more precisely, cosmic rays.

What the file will do

One of the principles underlying the design of the system was that it should evolve gradually; not only will its coverage of physics literature be increased step by step, but also its services will become more sophisticated in stages. Thus we can improve the system as we go along.

The services that will be offered during 1970 and early 1971 are all straightforward products of the information file:

• *Current Physics Titles* (CPT), a current-awareness journal, initially in four sections that probably will be: particle, field and nuclear physics; atomic, molecular, chemical, plasma and fluid physics; solid-state physics; and optics, acoustics, astrophysics and geophysics. We expect the sections will be published every other week, with each section representing a printout of the accumulated records since the previous issue. The records will be arranged, under a new system of headings, in as many places as physicists would expect to find them. The journal will be produced through computer-controlled photocomposition, so that it will be of high typographic quality.

• A series of specialized bibliographies in several of the narrower fields of physics (updated periodically) as well as indexes for the primary journals published by AIP

• *Searchable Physics Information Notices* (SPIN), a magnetic-tape service that will allow organizations with adequate computer facilities to establish their own current file of physics information. The tapes will be issued monthly and will contain the records accumulated since the previous issue. The subscribing institution could use its own search programs or specifically designed AIP programs.

At a later date, the system will offer:

• File searches based on requests. This service would be of particular value to scientists writing reviews or data compilations. Considering the importance of this activity in evaluating and distilling the literature into a more meaningful and digestible form, additional means for encouraging the production of such articles are also being planned.

• Lists of articles tailored to the needs of groups of physicists working in specialities and who do not have local facilities for using the magnetic-tape (SPIN) service, as well as procedures for subdividing journals into packages that would better suit the needs of smaller interest groups.

• Microform copies of the primary articles, as a backup to CPT and SPIN.

Long-range prospects

We expect to improve the system on a continuing basis, rendering services as effectively, and as inexpensively, as possible. As a long-range prospect, we hope to offer a centralized service, with decentralized satellites, that would cope with all the information needs of the physics community as well as those of the broader national scientific and technical community for physics information.

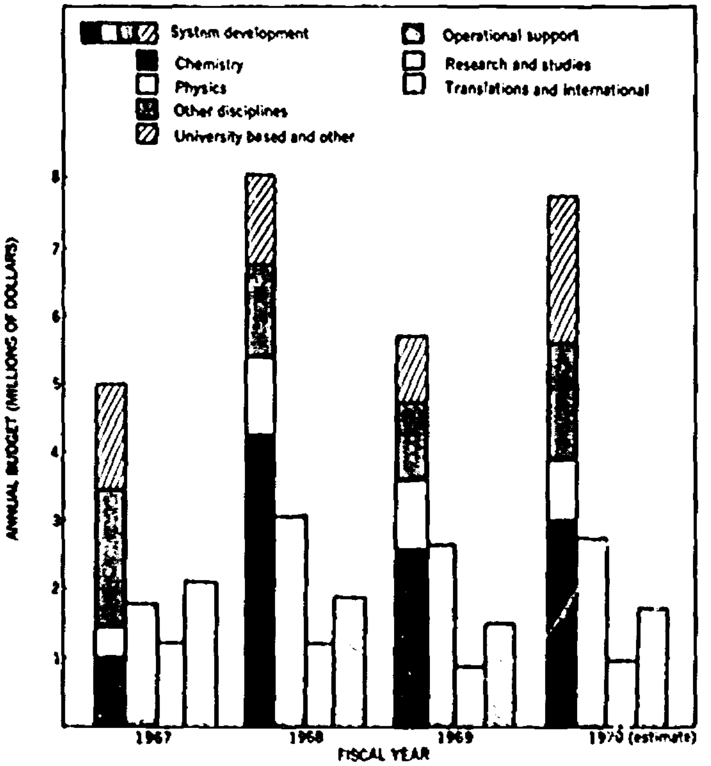
It would offer reference services and would obtain copies of hard-to-get material and refer questions it can not answer to sources that could. The system would also afford direct on-line access to the computer file from remote-access terminals in physics departments and other institutions. The centralized service also would have facilities for "scholars in residence," to supply clerical and reference aid for review writers.

Such a centralized facility would be linked into a network of "information centers" at various institutions and of similar facilities for other disciplines and for physics in other countries. This organization, with its information file and its broad spectrum of services, linked into a network of other information centers and services would constitute the "National Information System for Physics."

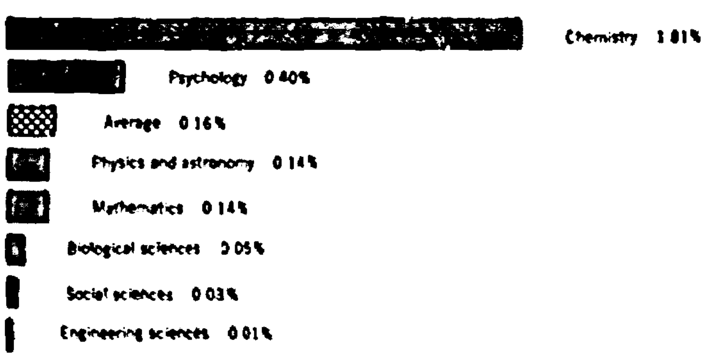
Value and cost

Each potential user must determine for himself what a service like this is worth. How many hours per week do you spend looking for information? How many hours would you save if you only had to look through one booklet, a short list or a response on your computer terminal? How much time would you save if the article was in one small collection or a numbered entry on a reel of microfilm that your librarian or secretary could copy?

These questions raise a number of imponderables. To put them into better perspective, consider that each published article represents about $60 000 worth of research investment; costs about $500 to be published and will cost about $15 to process and enter into the proposed physics-information file. The distribution price for listing that article in CPT will be only a fraction of a cent per copy. If, say, one out of a hundred articles is interesting to you and one out of a thousand important, would it be worthwhile for you to have it pinpointed? Similar considerations apply to the other services. Is a system that could accomplish these things worth the cost?



COST OF INFORMATION SYSTEMS. Chart compares NSF support of information programs by type and discipline during 1967-70. —FIG. 1



INFORMATION VS RESEARCH COST. Chart depicts ratio of NSF support for development of information systems to all federal support of research in the fiscal year 1968. The high percentage for chemistry is partly because most chemistry research is privately financed. —FIG. 2

The initial cost of development and pilot operation is being funded by the NSF, as part of its nationwide support of information services in scientific disciplines. Figures 1 and 2 show NSF expenditures for these purposes, both in absolute magnitude and in relation to total research support. Is it in the national interest to have the NSF support these programs? In 1968 NSF spent about $14 million for information activities in various scientific disciplines—less than 10% of the total was for physics. The improvement of efficiency in physics research and development activities is clearly in the national interest. The saving of a fraction of an hour per week by each of the 30 000 physicists in the National Register of Scientific and Technical Personnel, not to mention all the other users of physics information, would more than make up for all of the costs borne by the NSF.

In the future many of the operating services are expected to be self-supporting after the requested funding period ends in 1973. Some of the newer services would still need subsidies, and funds would still be required for further development of longer-ranged projects. The rate of NSF support, however, would probably decrease, and the additional cost would be offset by the greater values of the ultimate system.

A question

This program has been endorsed by the AIP governing board, which represents the member societies. But considering the magnitude of this undertaking, we would like the additional opinions of individual physicists: What do you think of our proposed system? Please write us and give us the benefits of your views and ideas on this matter.

References

1. V. Z. Williams, E. Hutchisson, H. C. Wolfe, PHYSICS TODAY, 19, no. 1, 45 (1966).
2. H. W. Koch, PHYSICS TODAY, 21, no. 4, 41 (1968).
3. *A Program for a National Information System for Physics*, American Institute of Physics, publication no. ID69R (August 1969). ☐

 Supported by NSF under GN-710

Appendix D

A TECHNICAL SOCIETY INFORMATION PROGRAM -- IEEE

Howard E. Tompkins
Director, Information Services
The Institute of Electrical and Electronics Engineers, Inc.

Prepared for the National Engineering Information Conference
June 25, 1969

Scientific and technological information dissemination has traditionally been a major objective of the Institute of Electrical and Electronics Engineers (IEEE) and its predecessor societies, The American Institute of Electrical Engineers (AIEE) and Institute of Radio Engineers (IRE), but that objective has largely been sought through activity in primary publication. First, let us get some background. Table 1 gives some relevant data on the IEEE:

Table 1. IEEE Numbers, February 1969

Members		163,000
(Student Grade 25,000)		
(All Other Grades 138,000)		
(Specialist) Groups		31
Local Units (Sections, Chapters)	>	1,000
Active Committees	>	3,000
(Involving over 25,000 Members)		
(Non-local) Conferences Sponsored per Year	>	100
Staff in New York	Approx.	300

It is the largest technical society in the world, provided you do not include the American Medical Association in that category. IEEE includes diverse members from research scientists to eminently practical technical field men, and many of its members now have very little contact with electricity or electronics as such.

A feature of considerable significance in shaping the actions of the Institute is the existence within its structure of 31 "Groups" having special technical interests such as Antennas and Propagation, Information Theory, Computers, Man-Machine Systems, Systems Science and Cybernetics, Magnetics, Electron Devices, Power, Instrumentation and Engineering Writing and Speech, etc. The IEEE Groups have a substantial degree of autonomy and independence, and are active centers of decision and action on matters of publication and information service.

The IEEE membership makes its needs and desires felt through the Groups, through local Sections and Group Chapters, and through Committees which are legion. Many of the Committees are involved in the management of conferences, and many of these are intersociety affairs which provide cross-disciplinary communication, for example, the Intersociety Energy Conversion Engineering Conference, cosponsored by about seven societies.

A staff of about 300 in New York "minds the store". Major segments of staff activity deal with the publications of the Institute, including their editorial preparation and their distribution. The Editorial activity is directed by Woody Gannett, who is here in the audience. The 38 primary periodicals of the IEEE have been--and still are--its major information dissemination activity. IEEE Spectrum reaches all the members (except those Student Members who exercise their preference to get the Student Journal), and contains articles of broad and general interest. A feature of Spectrum is an advance listing of every title that will appear in the other IEEE primary publications.

Proceedings of the IEEE is the major deeply technical journal of the Institute, and features special issues highlighting areas of significant and rapid technical advance in the field. Active efforts now under way are expected to increase significantly the number of review articles appearing in the Proceedings. Each of the 31 Groups has its Transactions, and some of them (such as IEEE Transactions on Computers, and IEEE Transactions on Aerospace and Electronic Systems) are major technical journals with circulation over 10,000. Others such as IEEE Transactions on Geoscience Electronics are small and highly specialized, and appear only once or twice per year.

Several Groups have worked together to sponsor the IEEE Journal of Solid-State Circuits and the IEEE Journal of Quantum Electronics. The latter journal, serving an active area of applied physics, gives further evidence of the difficulty of drawing a neat boundary between science and engineering.

Computer Group News covers matters of general interest in the computer World. Electrolatina (in Spanish and Portuguese) meets the special needs of IEE's Latin American membership.

I have listed and mentioned these primary journals to emphasize the important role they play in the dissemination of the new data of engineering and applied science. Of course, they do not re-present the old data in the manner of handbooks (at least not usually), but they do provide a major record of innovation, research, and development, and a medium for up-to-date review articles. We believe that primary journals will evolve, but not disappear. They will have new forms to supplement the old, as microfiche, and they will be produced using new techniques, but they will survive.

Other publishing activity of the IEEE is summarized in Table 2:

Table 2. Other IEEE Publications

7 Translated Journals
(Russian, Ukrainian, Japanese, Chinese)
4 Secondary Journals
c. 8 Standards Publications per Year
c. 60 Conference Publications per Year
Occasional Special Publications
n Newsletters, n Large

We shall return to the Secondary Journals in subsequent remarks. Of particular importance to our story is that a substantial body of conference publication exists, rivalling the periodicals in volume, and providing a needed parallel channel through which significant up-to-date technical data can be communicated to the specialized audience of greatest concern to it.

In Table 3 is shown a matrix presentation of the roles of selected parts of the technical information community, in an attempt to clarify their interrelationships. Individual members (MEM) of the technical community, alone or

Table 3. A Technical Information Matrix

	Data (Private Records)	Primary Public Records	Secondary Public Records
Creation	MEM	MEM PUB IE3	A&I SDI IE3
Storage	MEM	LIB SDI	LIB IE3
Distribution on Request	MEM	LIB SDI	LIB (IE3)
Distribution on Schedule	-	PUB IE3	A&I SDI IE3

MEM = Member of Community
PUB = Publisher
LIB = Library
SDI = Selective Dissemination of Information
A&I = Abstract & Index Service
IE3 = IEEE
( ) = 'Maybe'

in project groups, create new engineering data, store it in their notebooks, write informal memos about it, and tell their friends about it, with a degree of secrecy that is wildly dependent on individual circumstances. Eventually the individual community member gets together with a publisher (PUB) and creates a primary public record, which gets distributed according to some kind of schedule by the publisher, and stored in a library (LIB) to be distributed (or loaned) on request. To a sufficient degree, this model describes what happens to a technical report as well as to a technical paper.

If a "Selective Dissemination of Information" (SDI) service is operating in the subject area of the paper, that SDI service creates a secondary public record, a classified notice of the availability of the paper, and distributes it on schedule, according to profiles. (Criticality of relevance depends on how much information is given, and in how good a form, to permit the user to scan and evaluate the candidate item). Upon request, the SDI service dips into its storage (which may be the nearest friendly and capable library), and sends out a copy of the paper.

An "Abstracting and Indexing" (A&I) service typically operates in part like an SDI service, except that it is less speedy, and usually does not provide the library-like distribution of the primary public record (although some do).

Where now does the IEEE (IE3) fit into this pattern? Certainly as a publisher, creating and distributing the primary public records, more or less on schedule. Certainly also as a creator of some secondary public records, for its regular and specialized indexes. Beyond that, however, what should IEEE do? We disagree with those who say that the needed future systems are so incompatible with the present that we can't get there from here. As the SATCOM report has stated quite clearly, the technical societies have a definite responsibility for effective action and leadership in providing discipline-wide abstracting and indexing. To that end, IEEE established its Information Services Department in 1967. An early basic decision was made not to "go it alone" in the creation of secondary public records, but to use as applicable the long-established abstracting and indexing work of the Institution of Electrical Engineers (IEE), of London and Stevenage, England, which now goes under the name of INSPEC. To that end the IEEE has had for the past two years a joint publishing agreement with IEE covering the parts of Science Abstracts most relevant to the IEEE's scope. We also have made use of the services of EI on a contract basis, to avoid the buildup of a simply duplicative indexing staff. We are pleased to report that negotiations now almost complete are expected to extend IEEE-IEE cooperation into the indefinite future, and into the area of computer data bases.

A second basic decision by the IEEE is that it will maintain a well edited comprehensive data base, covering, first, its own published output, and second, the entire literature of electricity, electronics, computers, and control. Model 1 of that secondary data base now exists, and contains the 1968 IEEE published output. Model 2 will add the 1969 IEEE published output. Both of these will be converted to a permanent file format at the beginning of 1970, at which time we shall begin to accumulate data base records of the entire literature of the field, Electrical and Electronics Engineering, Computers, Control, and closely related areas of Applied Science and Mathematics.

IEEE will store that data base, and make it, or selected parts of it, available to subscribers on a regular schedule, for their direct use or, under suitable terms, for their inclusion in products they distribute. We are entering a thorny area here, full of pitfalls, but we expect to work out the problems one by one as they arise, for it is our deep-seated belief that the best interests of the community, the IEEE, and the IEE will be served by achieving a wide distribution of access to the IEEE data base through many media.

Finally, we come to an unsettled point. Shall we distribute information from the IEEE data base on request? In other words, shall we operate an on-line or quick turnaround information service? We fully expect to provide on-line direct inquiry facility for our staff, and our editors, and hopefully for our reviewers and our authors. Subsequently, when all the bugs have been found, and we understand what we have, I would like to extend this service to our members and to the community, but there are two questions, (1), the economic feasibility of such a service, and (2), whether a related organization such as the Engineering Societies Library should not provide such a service, in cooperation with us. Fortunately, we do not have to answer these questions all at once; events will help us decide.

Now, briefly, let me turn to the techniques we are developing: Author assisted indexing; selected citations integrated with subject and author index; computer assisted indexing; direct access for correction; computer controlled photocomposition; and inclusion of evaluative material. This program is supported one-half by the National Science Foundation, with Howard Falk as my principal colleague in the guidance of these efforts.

Author-assisted indexing is a natural procedure for IEEE to try, because our authors are indeed willing to cooperate, since it is a privilege to have one's paper accepted by one's peers and published in our journals. We ask our authors not only to assist by selecting the most appropriate classification number from the IEEE-IEE system for Science Abstracts, but also to supply index headings, and, perhaps most important of all, to identify those citations (references) they have made that are of particular value in retrieving their paper.

We plan to selectively include citations and other relational linkages among papers in our computer data base and in our indexes so that a user who knows of the existence of one article on a subject of interest to him can be led from it to later articles that extend, correct, or at least derive from it. The technique of using citations is not new--pure citation indexes may be purchased now. What we propose to do is provide good selectivity, to increase the signal-to-noise ratio and keep the size of the index within manageable bounds, and to provide the combination of subject, author, and citation indexing that proves by experiment and experience to be most effective. We also expect to put into the file such other relations as subject experts and information specialists are able to provide during evaluative use of the file.

In 1970, in the context of extending, deepening, and enriching our data base that has the IEE data base as one input, we plan to begin serious experiments in computer-assisted indexing. At least two major approaches will be considered, (1) Computer text processing of the title and abstract to yield a "pre-indexing" for examination and possible improvement by a staff editor-indexer, and (2) direct-access by the editor-indexer to the prior indexing of strongly cited papers to make use of that prior indexing to the fullest appropriate extent. It should be possible in many cases to index only the small difference or increment of knowledge appearing in the later paper, so we describe the concept as "incremental indexing."

Of course, direct access to file for indexing assistance will also be direct access for ease of making corrections, so we hope to have improved economical file maintenance as a byproduct of our new approaches. Clearly, we also have to consider the hazardous aspects of file security, or we will have too many people with direct access who can damage the file.

Scheduled output from the IEEE data base for a wide audience will be through the medium of computer controlled photocomposition. We have been using a Photon 901, but expect this year to switch over to an RCA Videocomp or to one of the other cathode-ray-tube composing machines. We will, of course, work through a service organization for the use of photocomposition equipment.

Some fine day in the not too distant future we expect to work seriously on the inclusion of evaluative material in the data base, to give the user a better idea of the audience for which the paper was intended, and an idea of the paper's significance. We do not know how to do this effectively, but should be in a position to try in two or three years. Perhaps a major part of the selectivity we provide will be in our choice of which data in a comprehensive data base we choose to enrich, and extend, and couple closely to the rest of the literature.

A paper in this conference would not be complete without specific reference to the problem of providing actual data to the user rather than just bibliographic information. As we tried to make clear in our presentation of a matrix of the Information Services World, we believe that effective data is always expressed in the form of a document of some kind. If the public has access to the data, a public primary data-bearing document exists. "Supplying the data" therefore becomes equivalent to "supplying the relevant primary document," or in some cases the appropriate sub-part of a primary document. We expect to index chapters and sections of lengthy documents where it appears bibliographically valuable to do so. We hope to continue to have effective liaison with on-demand document distribution organizations such as the Engineering Societies Library, and to work out a cooperative service with them.

Another approach may be provided by the rapid growth in importance of microfiche as a storage and retrieval medium. IEEE plans to use microfiche extensively beginning in 1970, or 1971 at the latest. IEEE microfiche, when available, will be numbered in a consistent manner, and fiche and frame number will be part of the IEEE data base, to permit automatic fiche and frame retrieval in one of the existing available computer-controlled fiche readers.

Meanwhile we shall be putting out the following products:

- Abstracts Journals or Journal Sections
- Titles Journals
- Year-end Indexes
- Selective Cumulative Indexes
- Bibliographies
- Data Base Segments at Retail or Wholesale, on Magnetic Tape
- Direct Access to Data Base

Many of these are on an experimental basis or addressed to a limited audience--and we shall be learning from their acceptance or rejection. As I indicated earlier, the last product mentioned, direct access to the data base, may be a product we encourage others to offer, using our data base as one of several they may need. We hope, in the process, to cooperate with other technical societies as is appropriate.

Having discussed our plans for the future, it is perhaps desirable to end on a note of realism--what we have accomplished to date. We have mechanized our index production of 1968 indexes. An IBM Magnetic Tape/Selectric Typewriter was used as input. The data tapes were then converted to computer tape form, and used to update our master file. When all material for a given index was correctly in the master file, it was exploded to index entry form, and formatted for proofreading, final editing, and photocomposition. The pages were then printed by conventional means. The process as we used it in 1968 was expensive, slow, and inflexible. We expect significant improvements in cost, speed, and flexibility for 1969, and should in 1970 have a satisfactory "permanent" system, competitive economically and operationally with manual techniques.

Finally, let me give a little more detail on our joint cooperation with the IEE of London. We jointly publish four secondary journals: EEA = Electrical & Electronics Abstracts; CCA = Computer & Control Abstracts; CPE = Current Papers in EE; and CPC = Current Papers on Computers & Control. They are photocomposed under computer control using a Lumitype 713 (Photon 713) in England. The two societies work together, using the advice and resources of their members, toward a truly useful classification scheme, and toward better methods of indexing and dissemination. Finally, arrangements are almost complete for us to use their data base as the basis for our enriched data base that will contain citations (and eventually evaluative material) as well as subject and author indexing. Both IEE and IEEE will market their computer data bases, beginning in 1970.

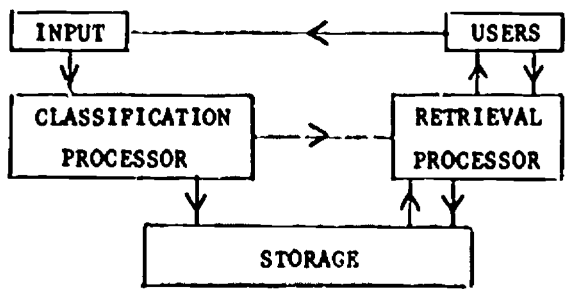
We hope that what we do shall prove valuable to the scientific and technical community that all of us serve. We also hope, and have tried to make it evident in what we have said, that we shall be but one of a significant number of information service organizations trying to meet the needs of the community--working together in a mixture of competition and appropriate cooperation with government and the profit sector that will keep us alive, well, and on our toes.

Appendix E

PRELIMINARY MODEL OF INFORMATION FLOW

R. Creighton Buck
University of Wisconsin

A highly simplified model of the flow of mathematical information may be diagrammed as follows.



In explanation, users are those mathematicians who produce or consume mathematics, as well as those non-mathematicians who consume it. Input consists of manuscripts, lectures (audio or video records) programs on tape or in punched card form, etc. The classification and retrieval processors are black boxes which identify the nature of a specific input, which address it, and which provide access to it. (Under the present system, portions of this are: editors, author abstracts, reviews, the journal _Mathematical Reviews_, the table of contents of a journal, the program of a meeting, etc.) Storage consists of preprints, published journals, microfilms, microfiche, magnetic tape, etc.

Through study of this idealized system, attempting to delineate and codify the nature of each of the boxes, and of each of the arrows, one may hope to arrive at both an improvement on the present system (realization of the diagram) and a process for continually improving on it, to take account of both human and technological changes in its components.

Every task in such a study can be related directly to one of the boxes, or to one of the arrows. In each case, the objective is to understand exactly what is being done now, explore other ways of doing it, and arrive at what seems the best combination in terms of cost and user needs and desires. The study might, for example, proceed as follows:

Task 1. Study the box USERS.

This would mean to identify the community of users and suppliers, both within the mathematical world and within the non-mathematical world, that uses or is interested in mathematics; the crucial thing is not to have a list of all the people who are in this general category, but to devise a means to identify this group, and to be able to communicate with them.

Task 2. Study the box INPUT, and the arrow joining it to USERS.

The former means to identify those things that shall be considered "mathematical" and worth putting into the system (or rather devising a means of making these decisions) and then to examine or create a collection of means by which people communicate their products to the system. Example: presently, anybody who wants to present a report to the AMS sends a very formalized abstract to the AMS office. Perhaps it should be recommended that there should be a single editor-in-chief for the AMS, to whom all manuscripts intended for "publication" in a generalized sense, be sent, and who in turn passes them on to one of a set of other editors, depending on the length, nature, etc., etc. (This overlaps slightly with the next task.)

Task 3. Study the arrow from INPUT to the box CLASSIFICATION PROCESSOR, and the nature of the latter.

This is a very big job. The combination of these comprises the whole process by which an individual document is examined, screened for suitability when this is relevant (i.e. refereed), classified for content on a multi-dimensional grid, and assigned for storage in a specified manner. Included in this is the production of various excerpts or projections of the document which will be used in various aspects of the retrieval system, including abstracts, reviews, and the like.

Task 4. Study the nature of the box STORAGE.

Among the various means of storage are included quick and slow publication, archival non-public tion (via micro print, film, fiche) with print-to-order facility, and even "return to author." The study should also include the examination of internal flow inside STORAGE, so that documents can move from one type of storage to another, as needed. There should also be a study of the storage aspects of the rest of the system itself, to allow for the possibility that some of the rest of the system can use the facilities involved in STORAGE.

Task 5. Study the box RETRIEVAL PROCESSOR, and the arrow joining it to USERS.

This is another big job. How does a user know he wants access to a certain document, and how does he get access to it? This includes the design of various things like Current Contents of Mathematical Journals, the Mathematical Offprint Service, Mathematical Reviews, etc., as well as means of pulling things out of storage and putting them in the hands of a user. The physical distribution of journals is only one way to do some of this. There may be many other ways of doing it better. There can be various depths of retrieval; first, one may have only author and title, then an abstract in very short form, or even a list of content labels, and then a more complete descriptive abstract or review, and then a complete document. The mechanism by which a user gains access to each of these might be different. (For example, there might be no publication in present form of any document, but with all documents available in microfiche form, so that what one buys is the coded location of the document in the file! The user then accesses the file at that location and extracts the hard copy print by facsimile.)

Task 6. Examine the total interaction of the system, and arrive at a management plan to allow for up-dating it.

PUBLICATIONS

of the

CONFERENCE BOARD OF THE MATHEMATICAL SCIENCES

Buildings and Facilities for the Mathematical Sciences,
by J. Sutherland Frame and John W. McLeod (1963).
ix + 170 pp., illustrated. $2.00.

Aspects of Undergraduate Training in the Mathematical Sciences,
by John Jewett and Clarence Lindquist; Volume I of the Report of the CBMS Survey Committee (1967).
xvi + 164 pp. $1.75.

Aspects of Graduate Training in the Mathematical Sciences,
by John Jewett, Lowell J. Paige, Henry O. Pollak and Gail S. Young; Volume II of the Report of the CBMS Survey Committee (1969). xxiv + 149 pp. $2.25.

Aspects of Professional Work in the Mathematical Sciences,
by Joseph P. LaSalle, C. Russell Phelps and Donald E. Richmond; Volume III of the Report of the CBMS Survey Committee (1970). viii + 144 pp. $2.50.

The above publications may be ordered directly from:

Conference Board of the Mathematical Sciences
834 Joseph Henry Building
2100 Pennsylvania Avenue, N.W.
Washington, D.C. 20037

The volume below, produced under CBMS auspices, is available from the publisher:

The Spirit and the Uses of the Mathematical Sciences,
edited by Thomas L. Saaty and F. Joachim Weyl for the Conference Board of the Mathematical Sciences. McGraw-Hill Book Company (1969). x + 301 pp. $8.95, hardback; $3.95, paperback.