Libraries and Information Technology: A National System Challenge (1972)

Pages
96

Size
5 x 9

ISBN
0309019389

Information Systems Panel; Computer Science and Engineering Board; National Academy of Sciences

LIBRARIES AND INFORMATION TECHNOLOGY

A National System Challenge

A Report to
The Council on Library Resources, Inc.

By
The Information Systems Panel
Computer Science and Engineering Board

NATIONAL ACADEMY OF SCIENCES
Washington, D.C. 1972

NOTICE: The study reported herein was undertaken under the aegis of the National Research Council with the express approval of the Governing Board of the NRC. Such approval indicated that the Board considered that the problem is of national significance; that elucidation of the problem required scientific or technical competence and that the resources of NRC were particularly suitable to the conduct of the project. The institutional responsibilities of the NRC were then discharged in the following manner:

The members of the study committee were selected for their individual scholarly competence and judgment with due consideration for the balance and breadth of disciplines. Responsibility for all aspects of this report rests with the study committee, to whom sincere appreciation is expressed.

Although the reports of study committees are not submitted for approval to the Academy membership or to the Council, each report is reviewed by a second group of scientists according to procedures established and monitored by the Academy's Report Review Committee. Such reviews are intended to determine, inter alia, whether the major questions and relevant points of view have been addressed and whether the reported findings, conclusions and recommendations arose from the available data and information. Distribution of the report is permitted only after satisfactory completion of this review process.

ISBN 0-309-01938-9

Available from

Printing and Publishing Office
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Library of Congress Catalog Card Number 71-183752

Printed in the United States of America

10 August 1971

Dr. Fred C. Cole
President
Council on Library Resources, Inc.
One Dupont Circle
Washington, D.C. 20036

Dear Dr. Cole:

The Computer Science and Engineering Board is pleased to submit the report *Libraries and Information Technology—A National System Challenge* by the Information Systems Panel.

The Panel examined the capacity of information technology to support information systems spread over public and private institutions. The interactions between technological and non-technological factors were also scrutinized.

The Board endorses the Panel's recommendations that experience be gained through pilot activities and that there be greater stress on scientific studies of networks which accommodate geographically dispersed information and users.

Beyond this, the Board recommends:

– that comprehensive and timely data on actual services and costs for the library system of the United States be obtained to guide future development and evaluation.
– that science policy join in welding the present fragmentary efforts into coherent national programs for improving the information handling capabilities of the United States.

In closing, we underscore Dr. Wigington's remark that the Panel's observations and recommendations "bear on activities that transcend the traditional library community, which cannot fruitfully be considered in isolation."

Sincerely yours,

Computer Science and Engineering Board
by Anthony G. Oettinger, *Chairman*

TO: The Computer Science and Engineering Board
FROM: Ronald L. Wigington, *Chairman*
Information Systems Panel
SUBJECT: Report Transmittal
DATE: 23 June 1971

The report *Libraries and Information Technology—A National System Challenge* is hereby transmitted to the Board. It is the product of our Study of the Applications of Computer to Libraries and Information Systems carried out with the support of the Council on Library Resources.

The report presents recommendations derived from the Panel's synthesis of facts, views, and opinions obtained from many sources, including visits to selected projects and installations, published and private information, and discussions with many individuals.

In addition to the Observations and Recommendations beginning on page 5 of the report, the Panel wishes to direct attention to the initial two Findings (page 10) which point out:

- The primary bar to development of national level computer-based library and information systems is no longer basically a technology feasibility problem. Rather it is the combination of complex institutional and organizational human-related problems and the inadequate economic/value system associated with these activities.
- The quantitative contribution of information to productivity or effectiveness of industry, government, and education is unknown; therefore, at the present state of knowledge, the construction of value/cost analyses is severely hampered.

The report is addressed to the Council on Library Resources for their use in encouraging the improvement of library facilities for the benefit of the people of the United States. However, the observations and recommendations also bear on activities that transcend the traditional library community, which cannot fruitfully be considered in isolation.

PREFACE

The need for a panel of the Computer Science and Engineering Board of the National Academy of Sciences to consider information systems was recognized in the spring of 1969. The Information Systems Panel was organized and started work at the beginning of 1970 with the sponsorship of the Council on Library Resources.

Many studies and committees had previously pointed to computer technology and the associated system science as the means by which the complex and growing national needs for information processing and information transfer of all types could be met. Among others, the Baker Committee Report in 1958 and the Weinberg Report in 1963 highlighted the national information needs in the areas of science and technology. More recently, in 1969, the SATCOM Report intensified the hopes for computer science and technology to play a primary role in making possible the timely and economic handling of scientific and technical information of importance to industrial success and national goals. Recent emphasis on social and environmental problems also calls for timely and accurate information from many sources in support of study, evaluation, and decision-making. In 1968, the report of the National Advisory Commission on Libraries broadened the statement of needs to the full range of information handled by libraries.

In these areas and others involving information handling, the science and engineering community faces a serious challenge to provide tools, techniques, and systems for the solution of some very difficult and important problems. While the potential of computers for handling information has long been recognized, the achievements, over the last several years, seem to be less than desired. Systems are slow in development, expensive in operation, less capable than desired, and generally insufficient to fully satisfy the national needs that have been identified.

For defining what is of national concern, the Panel used the following guidelines: (a) involvement of or the use of resources of national significance; or (b) significant impact on general United States society, government, or business; or (c) requirement for widespread coordinated action to achieve results, isolated local action not being sufficient.

In brief, the mission of the Information Systems Panel was designed to (a) assess the state and trends of computer and related technologies relative to the requirements of library and related information systems, (b) identify the roadblocks to more effective and rapid employment of these technologies for information handling, and (c) focus national-level attention on appropriate actions to correct deficiencies identified thereby.

The study was carried out through: on-site visits to selected libraries, research projects, and information system activities to discuss and analyze library and information systems research and to observe operational activity; discussions with individuals, not on the Panel, who are involved in the implementation of computer-based library and specialized information systems; and in-depth analysis by the Panel to develop the technical evaluation of the status and trends for computer-related information systems. The Panel is deeply indebted to the many individuals and institutions who cooperated in providing information, insight, and opinions to the work of the Panel. These contributions were made during site visits, individual discussions, and the process of review of the written report.

The report is organized in two sections. Following the Introduction, the main Observations and Conclusions of the study are presented and then the specific Findings that led to them. The remainder of the first section is a Discussion of Problems and Issues.

The second section of the report is a set of Annexes, which discuss three main topics on which the first section of the report is based. The dimensions of cost and usage of library and information systems are discussed, to the extent that information was available. The site visits to key activities and institutions are described and the key characteristics and development trends of information technology are presented.

In this way, the essential results of the study are placed early in the report for the convenience of the busy and hurried reader. Increasing detail is available in the later parts of the report.

INFORMATION SYSTEMS PANEL

F. T. BAKER, Manager, Applications Development, Federal Systems Division, IBM Corporation

JOSEPH J. EACHUS, Group Director, Advanced Systems and Technology Organization, Honeywell Information Systems, Inc.

DOUGLAS C. ENGELBART, Manager, Augmentation Research Center, Stanford Research Institute

GERALD SALTON, Chairman, Computer Science Department, Cornell University

JAMES E. SKIPPER, University Librarian, University of California at Berkeley

RONALD L. WIGINGTON, *Chairman*, Director of Research and Development, Chemical Abstracts Service

COMPUTER SCIENCE AND ENGINEERING BOARD
STAFF OFFICER

J. F. KETTLER, Assistant Executive Secretary, Computer Science and Engineering Board, National Academy of Sciences

ACRONYMS AND ABBREVIATIONS USED

AIM	Abridged Index Medicus
AFIPS	American Federation of Information Processing Societies
ANSI	American National Standards Institute
ARPA	Advanced Research Projects Agency
ASIS	American Society for Information Science
AUTODIN	A U.S. military communications network
BTL	Bell Telephone Laboratories
COBOL	Common Business Oriented Language
CPU	Central Processing Unit
DATRAN	Data Transmission Company
DDD	Direct Distance Dialing
DEC	Digital Equipment Corporation (now DIGITAL)
FJCC	Fall Joint Computer Conference
FORTRAN	Formula Translating Language for Scientific Computation
IBM	International Business Machines
INTREX	Information Transfer Experiments
MAC	Multiple Access Computer, or, among others, Machine Aided Cognition
MARC	Machine Readable Catalogue
MEDLARS	Medical Library and Retrieval System

MIT-TIP	Massachusetts Institute of Technology—Technical Information Project
OCR	Optical Character Recognition
SATCOM	*Scientific and Technical Communication—A Pressing National Problem and Recommendations for Its Solution*
SJCC	Spring Joint Computer Conference
SPIRES/ BALLOTS	Stanford Public (née Physics) Information Retrieval System/Bibliographic Automation of Large Library Operations Using Time Sharing
SRI	Stanford Research Institute
SUNY	State Universities of New York
TELEX	A name for a national teletype service
UCC	University Computing Company
WATS	Wide Area Telephone Service
WU	Western Union

CONTENTS

CHAPTER ONE

INTRODUCTION

The library system of the United States is a vital national resource for citizens of many endeavors, who seek the knowledge necessary to participate fully in our society. It has been recognized for over a quarter century, however, that the hoary ways of libraries are breaking down. The need for better ways, by now proclaimed by many, was expressed by Vannevar Bush:[1]

> Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose . . . .
>
> The difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present day interests, but rather that publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships.

Shortly after the Second World War, in which science and technology had played such a prominent role, Bush, with the above statement, urged scientists to turn to the task of making recorded knowledge more readily available for the benefit of society. And

[1] Vannevar Bush, "As We May Think," *Atlantic Monthly*, 176 (No. 1), 101–108, July 1945. Copyright ©1945, by the Atlantic Monthly Company, Boston, Mass. Reprinted with permission.

while, more than a quarter century later, national leadership to solve this problem has not emerged, many people still share Bush's vision and hope that science and technology will meet his challenge. New and powerful technology now exists to be used, but the basic mechanisms for exploiting recorded human knowledge remain mostly unchanged.

Many studies and reports have treated various aspects of information handling for national attention and action, usually with primary emphasis only on information related to science and technology. Chapter 8 of the SATCOM Report[2] summarizes the history of this national concern and cites many previous works on the subject. Information handling transcends national boundaries, since much of the information used in intellectual activity of all types has international origins and many of the results of that activity have international impacts. These aspects, as they relate to science and technology, have also been summarized in the SATCOM Report (Chapter 9). There is, therefore, no need to repeat this history here.

All the previous studies have concluded that: information resulting from scientific and technical work has high value in our technological society; means for handling it are inadequate; and modern technology, coupled with concerted national action, offers the opportunity for developing adequate means.

The new technology encompasses processes based on computers and compatible information-storage media, input and output devices, electrical communication systems, and various classes of image-reproduction techniques collectively termed "reprography." Computers are key elements of this technology because they provide the logical control mechanism to select alternative actions based on predefined selection conditions and on various characteristics of the information being handled.

The fact that the previous studies have concentrated on scientific information alone should not blind us to the broader opportunities presented by modern technology. Our entire society is increasingly dependent on comprehensive and readily available

[2]Committee on Scientific and Technical Communication, *Scientific and Technical Communication–A Pressing National Problem and Recommendations for Its Solution*, Washington, D.C.: National Academy of Sciences–National Academy of Engineering, 1969.

information. In this report the Panel adopts that broader view.

The traditional library system is the embodiment, in ancient technologies, of certain essential functions that any national information system, however distributed over public and private institutions, must perform. These functions include:

Selection: processes for choosing what should be obtained and stored.

Storage: processes for appropriately organizing the information to be stored, for storing the information, and for maintaining and managing the storehouses.

Access: processes for identifying and finding desired information in the storehouse.

Distribution: processes for transferring information from producer to storehouse, among storehouses, and from storehouses to users.

Control: processes for orchestrating the selection, storage, access, and distribution functions.

In traditional libraries, for example, the acquisition process includes aspects of control (the policy-setting and management processes governing acquisition), selection (executing policy on what to buy), and administration (ordering, monitoring orders, checking discounts, paying the bills, and so on). The library stack, now containing mostly printed bound books, embodies the storage function. Cataloging and indexing organize the information to be stored, and catalog or index look-up are means of finding pertinent information. Distribution is by physical transport of documents or, in some cases, photocopies of portions of documents.

New technology provides alternative embodiments of these processes. Programmed computer processes can support the selection and control functions. Evolving computer technology offers capability for the production and storage both of catalogs or indexes and of the primary records to aid the access function. Electrical telecommunications offer alternatives for providing machine support to all types of operations and to physical transport in the distribution functions.

This new technology can be brought to bear gradually as various aspects of each process come into the range of practicality;

there is no need to wait until an "ultimate solution" is available in all areas. Indeed, many localized and specialized applications of information technology are being demonstrated in the current information environment. However, to be adequate to the national information system challenge, the framework and planning for the evolution of an effective national system must be forward-looking in anticipating the trend of development of the pertinent technology.

Meanwhile, the costs of library systems and dissatisfaction with their performance continue to rise. The Panel therefore undertook to identify the deficiencies in the understanding and development of information handling technology that impede its effective use in meeting the totality of national information needs and to focus attention on the proper direction of resources to assure the development of the needed principles and capabilities. The Panel, therefore, has (a) examined the adequacy of present and developing information technology to support various functions in the categories listed above, taking into account the design and development problems likely to arise in the course of applying effective technology; (b) considered the interactions between technological and nontechnological factors, to determine what development strategies can lead to an information system capable of meeting the nation's needs; and (c) developed recommendations for actions needed to bridge the gap between the potential of new information technology and its effective use in the national information system.

The study included visits to selected projects and operations using computers in libraries and information processing, review of published information on relevant topics, and use of the experience of the participants in the study. Through these means, insights and experience from many sources were brought together. These sources include library automation projects, the computer and communication industries, academic information science research, science abstracting and indexing system development and operations, the commercial computer service industry, and the news processing and publishing industry.

CHAPTER TWO

OBSERVATIONS AND RECOMMENDATIONS

Two important observations emerged from the visits and discussions of the panel:

1. To enable information technology to be applied more effectively for improving national information handling capabilities, science policy must assume a much more effective role in stimulating public and private actions to weld the present localized and fragmentary efforts into nationally coherent programs.
2. Comprehensive, consistent, and timely data on actual services rendered and on costs for the library system of the United States must be obtained and presented to provide a firm basis for planning and decision-making in developing and evaluating coherent programs. All libraries—school, university, industrial, special, public, federal, national—must be included.

To surmount critical obstacles and to place new capabilities into a framework enabling them to be widely used, we make the following two recommendations:

1. *The present collection of localized and fragmented efforts must be guided toward harmonious integration through experience with a comprehensive pilot system.* A pilot system must be of a scale sufficiently large to model accurately those critical operational problems that typically fail to appear in small systems, yet whose absence produces misleading results. To be comprehensive, a pilot system must account for all the categories of functions in-

herent in a national information system. To be realistic, the pilot system must incorporate files of sufficient subject coverage and contain information published over a sufficient period of time, so that users will depend on it for their operational needs; it must cover sufficiently heterogeneous subject areas so that conflicts in special needs are encountered for resolution. To be of value in illuminating policy, economic analysis of the pilot operations must project them to a future full-scale context and not account solely for current operations in the pilot system. Finally, to be truly effective, the pilot activity must find an acceptable way to bring the many institutions and interests, public and private, into the productive partnership that will be necessary for eventual full-scale operations.

2. *To develop information systems consistent with geographic dispersion of information resources and information users, increased stress must be placed on scientific design and modeling studies of broadly based information networks.* These studies must consider explicitly the available alternative embodiments of all the necessary functions of information systems as well as the technical, economic, and policy parameters governing choices among these. At least the following technical aspects must be addressed for any type of network: information flow facilities and patterns, intellectual and physical access dynamics for individuals and for organizations, file maintenance patterns, file and process backup arrangements, and an appropriate balance between automatic and manual processes. Economic analysis must go well beyond the usual cost-oriented functional studies to encompass financial transfers within the network, charges for various classes of service, capital and development costs, and conditions sufficient for startup and viability of the network. Early results from such analysis are needed for guidance, to assure that the necessary aspects are included in the pilot system, and continuing analysis is needed to guide the eventual development of a full-scale system.

Because several projects are now in process, and good progress is being made in some of them, it is now possible and timely to respond to these two recommendations. Some of these existing projects are aimed at single libraries; some, cooperative efforts among the members of a group of libraries; some, aimed at single functions needed in an overall system; and others, related information-handling activities on a discipline-oriented basis or on a mission-oriented basis. Separately, they are not adequate to pilot all aspects of a national system and are hindered by the lack of the necessary national framework, but they form a base of much experience and provide candidate segments for the recommended national pilot system.

There are certain specific targets in the development and employment of computer science, engineering, and related technologies in information systems. The following recommendation is addressed to important technical areas needing emphasis identified by the Panel:

3. a. *Data Base Organization and Management* This is a critical area in the development of software for creating, managing, and manipulating the large files normally associated with information handling in an efficient and economical way; it is, however, still in an embryonic state. The computer field is beginning to pay more attemtion to this problem, and may soon give it as much emphasis as has been given in the past to programming languages and operating systems. *It is therefore timely to provide guidance to systems software projects so that satisfaction of the requirements of the library and information community will be viewed as a major, pace-setting goal of data management systems developers.*

b. *Large-Capacity Digital Storage Technology* Present digital bulk storage technology cannot be used to store all the contents of a large library economically, even if the information were now available in the necessary form. Much work is under way and excellent progress is being made. *Nevertheless, the requirements of the library community for large and low-cost storage facilities will probably exceed the needs of scientific and commercial applications. Great care must therefore be taken to ensure that development is not slowed down prematurely.*

c. *Distribution Technologies for Information Networks* Computer and communication technologies have great promise as media for distribution of processing services and information among generating, processing, storing, and using points. However, it is necessary to explore the specific conditions under which electrical transmission becomes economically competitive with local facilities and physical transport and/or enables sufficient savings to be realized in development and operations to cover its cost. *This is a specific factor to be examined for the full operational system and one to be tested in the pilot system.*

d. *Machine-Readable Catalogs* Catalogs and indexes are fundamental elements used in the intellectual and physical access to any information system. Although several projects have produced important results in this area, a coherent, total solution has not yet been constructed. *Steps must be taken to ensure that the cataloging and indexing of all useful information items will be brought to full operational status in computer form, available to and usable by all libraries.* Printed forms derived from the machine record will maintain compatibility with manual operations.

e. *Microform Technology* Microform technology will be important as a primary text storage medium for automated information systems until the costs of input and storage of digital forms come into an acceptable range. Machine-controlled microform collections must operate in conjunction with computer-based information access systems. *Hence specific attention must be given to ensure both the availability of information in microform and the necessary degree of compatibility between microform and computer technology.* At the same time, high-quality microform masters might serve as an interim storage medium for eventual conversion to digital form by means of advanced optical character reading methods. *The specific technical and cost requirements for this purpose must be determined; the adequacy of the relevant conversion technology must be assessed in detail.* The importance of an interim method for conversion of text information into machine readable form, to suffice until information can be routinely captured at the source in digital form, is emphasized by the cost of present input methods. The cost of input through keyboarding exceeds, by an order of magnitude, the cost of storing the information in today's large-capacity memory systems. Thus, even if large-scale memory costs were now low enough for storage of all text information, input costs alone would prevent its use except for text of high value and text already available in machine-readable form.

f. *Nonprinted Primary Media* A growing body of material is available that can best be used through computer processing. For example, Census Bureau information on magnetic tape has provided a new information source for scholarly activity. Since the libraries were unprepared to provide routine service for interested users, various *ad hoc* arrangements are being made in this specific case. *A plan should be developed for extending the general library functions to handle machine-readable sources of information, including semi-processed material (e.g., census tapes), highly refined and validated data (such as is becoming available from the National Standard Reference Data System), and published information as it becomes available in computer-readable form.* Various audiovisual media are yet another class of nonprinted information not adequately handled by conventional library systems.

g. *Education Key workers and managers involved in building and operating segments of the national information system must be provided supplementary knowledge and training in one or more areas outside their respective primary fields.* The following fields are included: library science, information science, computer science and engineering, economics, systems analysis, systems operation management and control, development management, and business administration. In addition, the characteristics of information used in the subject fields that are served by the information system must be understood so that the information-handling activities will be in harmony with the needs of each field, and, in many cases, subject-competent information specialists will be needed to aid users of such systems.

It would be inappropriate for this Panel to nominate a specific institution to lead the national effort or to select specific projects or institutions to furnish the segments of the recommended "pilot national systems." However, it is clear that these actions must be taken by an appropriate national body that has the stature and means to provide the necessary cohesiveness and continuity on a national scale. The criteria for guiding these selections and for choosing important areas of emphasis have been identified in the comments supporting the statement of recommendations.

If the necessary national impetus cannot be generated and joined by the relevant institutions as participants, to produce better operational results than are possible working separately, then no matter how many studies, or reports, or plans are made or good intentions expressed, the needed, expected, and possible improvements in information handling and supply through the use of computers and related technologies will remain a distant vision.

CHAPTER THREE

FINDINGS

In the course of its study and discussions, the Panel on Information Systems arrived at the following particular findings concerning the problem and some specific requirements for dealing with it.

1. The primary bar to development of national computer-based library and information systems is no longer basically a technology-feasibility problem. Rather, it is the combination of complex institutional and organizational human-related problems and the inadequate economic/value system associated with these activities. National leadership to solve these problems has not emerged.

2. The quantitative contribution of information to productivity or effectiveness of industry, government, and education is unknown; therefore, at the present state of knowledge, the construction of value/cost analyses is severely hampered.

3. Complete and reliable data are not available to give an accurate and up-to-date measure of the resources expended on library activities of all types and of the actual productive use of the existing library facilities. Attempts to obtain and provide this information have produced only incomplete results and are inadequate for national decision-making. However, the total annual expenditures for information-handling activity throughout the nation are estimated to be in the multibillion-dollar range.

4. The library and information community is large enough and

important enough to influence some products of the computer industry. The community could better create this influence if it were not so fragmented relative to technological objectives and requirements.

5. Difficult problems associated with property rights in access to and distribution of information must be dealt with or realization of the desired and technologically possible systems will be hindered and delayed.

6. Some useful automation has been brought to bear on information-collection management in some large libraries, by a few collective groups of libraries, and in certain industrial support libraries. For the most part, these are localized developments and not easily transferred to other locations.

7. As a system design problem, the long-term continuity of information systems is different from the requirements of systems designed for scientific calculations or for handling routine business records. This difference has considerable impact on the system principles to be considered. The information files—their representation and organization for effective access—are and should be the primary concerns at present.

8. In the long-term, digital computer-readable forms for information are probably the best forms. However, because of the presently incomplete penetration of this technology into publication activities and the economics of conversion into machine-readable forms, near-term and mid-range plans and developments must be based on image storage for most "full text" (primary) documents.

9. The strategy for development should accommodate multiple forms of information handling, coordinated so that the employment of machine-readable forms can gradually be increased as technological achievement and economic considerations allow. Neither an all-at-once nor a completely separate development strategy is appropriate.

10. Computer technology continues to improve on a performance/unit cost basis at a dramatic rate. A key requirement—information storage of sufficient capacity, at acceptable cost, and with practical access times—is being met for some purposes. The prognosis for favorable continuing development is excellent.

11. The technological feasibility of providing processing functions and of transporting large volumes of data by electrical means

is not in question. The presently unfavorable economics of use will improve as new plant is installed, new services thus becoming available, and as general usage rises.

12. The software for organizing and managing large computer systems files is still in an embryonic state. The computer field is beginning to pay considerable attention to this problem, as it did in the past with programming languages and operating systems, and should be guided by the needs of libraries and general information systems.

13. Interactive computer systems are being developed and will come into active use for supporting the intellectual activities of appropriately trained people. However, for some time yet to come, noninteractive batch processing will handle the bulk of the overall processing to be done in computer-based systems. These two processing methods apply to different portions of an overall information-processing system and for effectiveness must be compatibly interlinked with common files and control procedures.

14. Libraries generally are not set up to handle the traditional functions for digitally produced files (e.g., census tapes, numerical-data collections) as they have for printed publications, nor have they been financially able to equip themselves to meet evolving requirements. Similarly, various other nonprinted media, such as films and cassettes, are not part of the traditional library environment. Attention given to these new media has been minimal.

15. The design problems being faced in the automation of information activities are not the intellectual problems of single disciplines or fields; the solution to these problems requires merging the knowledge and experience of personnel among the fields involved.

16. It is technologically practical at this time to initiate large-scale development of systems to bring various libraries and information activities into a coherent national system, initially at a pilot operations stage but with intent and commitment to expand to a full-scale system.

17. A key economic problem relating to reduction of unit prices of equipment and information services to "affordable" levels is the achievement of sufficient market volume. Of course, lowering prices will aid in increasing market volume. Thus, the achievement of getting above the threshold of a self-sustaining market is critical.

CHAPTER FOUR

DISCUSSION OF PROBLEMS AND ISSUES

NATIONAL FRAMEWORK

To apply automation effectively to libraries and information systems nationwide, problems of many kinds must be solved in a comprehensive framework. Unless the appropriate environment and certain critical capabilities are created with a national system in mind, there is little hope that the magic of automation will soon be effective.[3] However, if a national pattern were established enabling both national and local developments to take place within it, extant and emerging information technology, which provides for improvement, could be applied on a broad scale. The multiplicity of efforts could then be mutually supporting, not fragmented or in conflict, as too many now are.

This visualization of needed and possible actions for improvement does not differ from the ambitions and intentions of past and current efforts. However, too often progress has been slower than expected or needed, particularly when viewed from a national perspective; the problems encountered have been more complex than originally understood; the costs for development and operations have appeared to be higher than were predicted; and, in some

[3] G. Salton, "On the Development of Libraries and Information Centers," *Library Journal*, October 15, 1970, 3433–3442.

instances, the presently available technology has seemed to be deficient. This present study explores the situation relating to these perceptions and identifies opportunities for national action.

The rapidly evolving and improving state of information technology leaves no doubt of the ultimate feasibility, as well as necessity, for taking the automated approach. Information technology, if properly applied, has characteristics that can replace labor-intensive repetitive operations, freeing personnel for more intellectually oriented functions. It provides the means for achieving "economies of scale," a principle that has been demonstrated in both the computer and communications industries as well as in large-scale manufacturing, which can combat faster-than-linear growth of costs with increasing volume of material to be handled. Further, the continuing, rapid technological improvement and cost reduction per unit of performance in all aspects of electronic technology make the use of this technology increasingly attractive for replacing older methods of information transfer and handling. Library institutions of all types must keep pace with these developments if they are to provide necessary new services and to continue to furnish existing services.

PANEL POINT OF VIEW

The Panel did not seek to find a way to buttress traditional libraries with new technology; rather it sought the best way to improve society's "recorded dialogue" system. In this system, present libraries of all types are the present embodiment of the memory that holds the recorded information resulting from the "dialogue" that has taken place among all parts of society. Certainly, if we are looking for the best way to improve society's recorded-information system, we must make the most of the resources of our current library systems and should plan for evolution from the current situation in the harmonious way that, among other things, does not needlessly abandon valuable assets.

The recorded-information system in a society is as critical a functional component of the society as is a "central nervous system" in an individual organism. It is absolutely essential to the proper evolution of society that the production, management, retrieval, and access of recorded information be made as effective as possible.

Viewed as but a part of society's recorded information system, today's libraries will become obsolete components if they cannot provide greater information accessibility to meet future requirements. The recorded information system of tomorrow may well distribute the various functions of a library—as we know it now—in such a way that ordering, cataloging, bibliographic searching, and physical accessing (and studying) are done in widely separated places and/or with distributed services.

During the period in which computer technology is being extensively integrated into the library system, there will be concurrent activity by many segments of our society, industriously integrating computer systems into their workaday world, including such "intellectual, knowledge-oriented" activities as their studying, planning, formulating, communicating, teaching, deliberating, negotiating, and managing.

On a large scale, at least for some periods of time and within some community of use, this will have an overwhelming effect on the quantity of significant, recorded communication that will need to be stored and made retrievable for subsequent access. In other words, the "library problem" can be made very much bigger and more complex by the very technology that provides hope for dealing with it.

Thus, the fact that technology can open the floodgates to the production of a vastly greater volume and diversity of recorded material must be recognized ahead of time and allowed for and guided, or—like freeways—the modernization facilities themselves will become congested. The assimilation process must be reoriented so that it is more timely, more complete, and less costly to satisfy the needs of every citizen. Also, only a small fraction of the content of publications has long-term archival value for immediate availability. Many other classes of materials have a more transient value or can be accessed more leisurely.

The exercise of judgments of value for selection and retention of materials on a priority basis may require wisdom exceeding Solomon's. Nevertheless, the compounding of "more's"—more books, journals, and reports, more places of availability, more needs for information from more fields, more ability to analyze old information and produce more by and for more people—will eventually saturate any system no matter how high its capacity or capability may be.

HETEROGENEITY OF "PROBLEM"

Although we use in this report, as others have, the singular generic term "problem" as a subject for attention, this is done with full realization that "solution" of the "library and information problem," as an automation-development activity, is not a single homogeneous undertaking. There is a wide range of functions, sizes, sources of financing, and operational environments for libraries and the information production and demand activities with which they interface. Acceptable costs in one situation are not tolerable in another. Nor does the appropriate operating mode for a small library provide the same unit cost and functional performance as for a huge library. Similarly, the method appropriate for serving a graduate student doing original research is not necessarily the same as quick-response support of a business executive. The Panel saw many of these variations during the visits to automation project sites.

THE EFFECTS OF INCREASING SIZE

Those entities serving a large group of people generally face difficulties stemming in part from size and increase of size. As the body of recorded knowledge expands—and this will be the case so long as civilization as we know it continues to exist—so does the burden of housing the record of that knowledge. As the community served increases in size, so does the diversity of interest from individual to individual. Thus, the fraction of the total of knowledge useful to some member of the community diminishes with the growth of the record. On one hand, an entity serving a community of static size has increasing requirements, and one serving a community increasing in size at a linear rate has requirements that increase at a much faster than linear rate. If the increased requirements are met by expanding the function in traditional library fashion, the cost of doing so increases not less than linearly with the requirements. On the other hand, those who defray the costs—whether they be a legislature, the managers of an endowment fund, or paying users (patrons) of library and information services—are unwilling and, beyond some point, unable, to allocate an ever-increasing fraction of the total available resources to some one function, regardless of its importance.

One concludes that the problems met in attempting to serve a large, growing, diverse community are not to be solved locally with more buildings and more employees. If a solution is to be achieved, it must be done either by *altering the problem* or *going beyond the local level*, or by both.

The establishment of the National Library of Medicine is an example of a serious attempt to do both in a specific area of interest. If successful in the long run—and there is no reason to believe it cannot be—the "national" center operating in conjunction with "regional" distribution points may well substitute for a hundred libraries or portions of libraries each a tenth the size of the national library and perform service in excess of the combined ability of the hundred.

The *alteration* of the problem here occurs in the change from "diverse" to "uniform" in the nature of the community being served. The *going beyond the local level* may be seen in the fact that the community served is geographically dispersed, not merely that the library is a national institution.

Evidence that the problem can be altered at the local level to the benefit of the local institution has been demonstrated by some universities, by some nonuniversity libraries, and by other information activities. In each case, alteration has been effected by departing from traditional methods to take advantage of electronic means for handling and conveying information.

Some success has been achieved in attempts to work at the global level without significantly altering the problem. There is vast duplication across the country in cataloging the same book or other piece of library material independently in many places even if it be done the same way. The Library of Congress' effort to remedy this, as authorized by Title II-C of the Higher Education Act of 1965, is commendable. It authorized the dissemination of bibliographical information in either conventional or nonconventional form. Of particular interest to computer-based systems is the portion of this program that led to the development of MARC II computer-readable bibliographical information.[4] However, its introduction has sometimes been frustrated by the incompatibility of the cataloging system used by a library that might otherwise benefit, the slow start-up time for extending a pilot operation to full scope, and by the threshold of usefulness (i.e., in scope of coverage), which must be exceeded if it is to be economically helpful to a recipient and thus be taken seriously.

The shared cataloging and bibliographic access systems being developed by the library cooperatives are other examples of steps being taken to improve the situation. Similarly, abstracting and indexing services will be a source of machine-readable information for access to stored information.

All these efforts, however, have suffered in the past from the lack of off-the-shelf input and output devices capable of handling an appropriate character set. They also have been hindered by the costs of storing or handling large files, injured by the growing pains of the adolescent computer industry, and lacking in comprehensive national cooperation toward a unified objective.

THE "MEMORY" ASPECTS OF THE LIBRARY FUNCTION

Considering the "national information problem" from the library point of view is useful because the library function contains the "memory" of national information systems. The current embodiment of this memory is in the wide range of the large national libraries, major academic and research libraries, federal libraries, industrial and commercial libraries, public libraries, school libraries, and so on, and various information collections used in support of industry, government, and education. The system has been built up over many years, based on the traditions of centuries, oriented mostly toward the acquisition and management of storehouses of knowledge printed on paper, and based on manual processes that require extensive human interpretation of situations and the exercise of judgment in handling them.

[4] H. B. Avram, J. F. Knapp, and L. J. Rather, *The MARC II Format: A Communications Format for Bibliographic Data*, Washington, D.C.: Information Systems Office, Library of Congress, 1968.

Thus, for the purpose of functional characterization in this report, present libraries of all types have been regarded as storehouses of recorded knowledge (or recorded dialogue) in graphic form to be used over a period of time (delayed information). However, the emphasis on "memory" must not imply only a "storage technology." Storage technology must be coupled with an "accessing technology" with both intellectual and physical components that increase the accessibility and usefulness of the information stored.

The examination of aspects of the "library function" as an analogy to the memory of a computer system provides insight into library problems and their solutions. With this point of view, the investigation is freed from considering only the mechanization of traditional methods. The suggestion arises that the library function could be extended to cover new forms of information and to use new methods for processing it. Also, the requirements for mechanized interfaces to the memory function are clarified. To eliminate the necessity for extensive, expensive, and time-consuming human interpretation through the application of automation, the entire system of procedures and mechanisms available for the supply of information to the memory function and the associated retrieval and distribution mechanisms need be addressed as an overall memory system design problem.

RESPONSIVENESS TO EVOLVING NEEDS

Since the memory institution has the responsibility to organize the existing knowledge in anticipation of unknown or unstated future requirements, it cannot exist solely on the basis of justification by current needs and organization to meet them alone. Thus, the motivation, design guidance, and support for building this resource for the future need specific attention, apart from the mechanisms and economics of using the resource once it exists for current use. Since the value measure for satisfying known needs does not exist now, it is even more futile at this time to derive a quantitative value for being ready for future requirements, whatever they may be.

An illustration of a shift of needs can be seen in the development of a specific application or mission-oriented interest, such as lunar and planetary exploration in the 1960's or "pollution" in the late 1960's and 1970's. In this kind of situation, information is needed from many fields—from the "hard" sciences and from economics, social sciences, politics and government, and other sources. Information that has been organized along traditional lines in the past must be reorganized and somehow made accessible via new intellectual routes. Information entities that were unrelated before must be brought together under new categorizations. This means the memory institution cannot use a rigid, static structure for organization of its information but must provide as much basis as is practical for automatic reorganization when new information needs are recognized.

SOME PROBLEM CHARACTERISTICS

In terms of an automation project, certain characteristics of the problem must be recognized. The economic and human factors have complex aspects that must be considered.

As a commodity to be managed, information is unusual in the following ways:

- Its quantity can be expressed in terms of characters or words or other "output" forms, and its recording medium can be measured in terms of pages or geometrical volume or weight, but its value varies with timeliness, with format to aid comprehension and further processing, with knowledge already possessed by the recipient, and with the relevance of what is delivered to the recipient to his concerns at the time of his receiving it.
- The extraction of information from mechanized storage forms by copying does not deplete the original resource, but the distribution of those copies may either increase or decrease the value of the original depending on whether the usefulness of the information is determined by universal, exclusive, or intermediate availability.
- The mechanical methods for handling information, whether computer forms, electronic communications, or microimage forms, are such that the information is not directly perceivable by human senses, and what is being handled and managed requires a high degree of indirect perception and a well-developed ability to deal with abstractions.

- Unlike food or other consumables, the apparent need for, and use of, information is regenerative, limited more by the means, cost, and time of obtaining and using it than by some absolute quantity that can saturate the need.

SOME SYSTEM CHARACTERISTICS

Some special system characteristics must be kept in mind during design and implementation of library and information systems:

- The overall system is complex, but it must serve, for the most part, persons without special training.
- The raw material input to the system is not now in the control of the system, i.e., the tradition is to accept everything that is published.
- The content of the system provides part of the source material for intellectual activity that provides the input to the system.
- The storehouse of knowledge that the traditional library function now represents is a cumulative, continually growing, non-self-purging file for which no widely acceptable purging criterion exists. This leads to a hierarchy of memory function, so that "the library problem" is really a multilevel problem.
- The problem of "getting started" is a serious one. To be useful, an information file must span both sufficient subject coverage and time so that the user can depend on it to contain the information necessary to useful conclusions and all significant existing information bearing upon those conclusions. Unless this requirement is satisfied in operational systems, users cannot, and will not, depend on the system. Unless a pilot system is chosen carefully so that this criterion can be satisfied on a relatively small scale, the pilot experience will not be realistic or small enough to be a reasonable pilot system.

TECHNOLOGY TRANSFER

Of particular concern has been the problem of "technology transfer," i.e., the problem of exporting the results of successful design and operation from experiments and pilot sites to other locations.

The problems of computer technology transfer were observed on two levels. As is true with much of the application software developments of the past and present, where the development has been carried on by an institution to handle its special requirements, the computer programs are simply not designed from the start to be transported, even between two installations having the same model of computer. They are incompletely specified and documented, sensitive to specific configurations of secondary storage, input/output equipment, and terminals, and dependent on specific versions of operating systems. Transfer between two types of computers is even more unlikely. With diligence and skill at the receiving institution, ideas can be transferred; portions of programs may be usable if common file forms are chosen; and one institution can learn from the struggles, achievements, and mistakes of another. The library-management and information-access systems that have been built so far have grown with the evolving understanding of the automation target. That understanding is yet incomplete and especially varies in extent and depth among the many workers in the field and other interested parties.

There are widespread misconceptions about transferability of computer processes, especially among those who have not tried it. The great flexibility, generality, and computational speed of computers have encouraged the development of a legend about the ability to take an isolated accomplishment and use it elsewhere with minimal, if any, work required to perform the transplant. If adequate prior planning has been done for the transfer, and with careful design control suitable for a specified operating system, the transfer can be made, but it will not occur automatically.

Thus, the aspect of technology transfer involving computer functions themselves can be improved by careful engineering for that purpose from the beginning once the system problems are understood and concepts have been proved out in prototype or pilot systems. However, much more difficult is system transfer involving the entire concept of doing business. A computer program system is only a vehicle for performing a set of functions, and, at the present state-of-the-art, such program systems mirror the operating environment, rules, and individuality of the specific institutions supported. Until a much greater compatibility of conventions and procedures is achieved in information processing, organization, storage, accessing, and managing, the complete transfer of systems will not be possible. There are alternatives to the actual transfer of systems such as are provided by networks of services to be distributed by means of electrical communications. While the economics of such operations are still questionable at this time, such an approach is an important future possibility.

THE LONG-TERM CASE FOR DIGITAL MEMORY

To achieve the full potential of computer-based information storage and access systems, it is necessary to get into permanent machine-interpretable form the total content to be manipulated. There is little disagreement among computer people that some type of digital representation describing the content of the information (as contrasted to, for example, an "image" of a page or a digitized scan of an image) is essential. However, disagreement arises among computer people and among the members of wider communities when the scope of the content to be put into that form is not defined, when the time and conditions for obtaining it in digital form are not specified, or when the handling of the information in one part of the overall national system is considered without the impacts on other relevant functions. Also, not everything is appropriate for conversion to digital form, such as rare books, works of art, high-resolution photographs (at least at present), and anything else concerning which the study of the actual object is part of the target of scholarly attention.

From observation of the present state and trends of memory technology, it is clear that high-value and limited-scope content can be handled in digital form today, but that the expected improvements in the near future will greatly widen the scope of practical situations. However, it is also clear that extreme situations, e.g., storing the entire holdings of the Library of Congress in digital form—i.e., 10^{14} to 10^{15} bits, to pick the single most extreme example imaginable—will not be practical in the immediate future, although development trends indicate that even this extreme is not out of reach in the very long term. It is neither desirable nor necessary to let the extreme situations prevent doing what is practical as the technological capability expands at practical levels. Thus, it is not necessary now to decide conclusively on either the desirability of handling the extreme cases or the ultimate technological possibility of doing so.

The reasons for getting the content into computer-interpretable form and distributing it in that form include:

1. Avoidance of unnecessary human participation in transferring the information from one form to another in the overall information-transfer system.
2. Permanence and regenerability of recording.
3. Transmission by electrical means with reduction in bandwidth requirements as compared to the wider bandwidth requirements for conventionally scanned images.
4. Suitability for manipulation by automatic processes for analysis and organization of content.
5. Flexibility and capability for the final users of information to explore, manipulate, and augment it in machine-aided systems.

The last two reasons apply to less conventional and longer-term desires and are dependent upon a general introduction of computing facilities into all aspects of society. However, to deny these possibilities and to fail to anticipate them would be shortsighted.

To speculate on the ability to store and use large computer-readable information files without being able to ensure that they can be prepared is of little or no value. Whether the full content of original documents is involved, or only bibliographic information, indexes, and document surrogates, the labor for original input into machine-handleable form is a serious functional and economic problem. In principle, when the information can be captured at the original source, in a form suitable for later machine-manipulatable re-use, this problem will be alleviated. However, most of today's publishing environment does not supply information in digital form.

To consider the total digital environment for all new material in libraries and information systems, one must assume that computers and computer-readable media have displaced the printing press and paper that have dominated formalized information transfer since Gutenberg. This would be a far-reaching assumption and will not be satisfied except in special environments for quite a long time, especially if an international view is taken that recognizes that the United States now has a significant lead in production and use of computer technology.[5] Because of this and because of the prevailing form of the extant library materials, a mixed digital/nondigital system for handling document content must be expected.

[5] Nicolas Jequier, "Computer Industry Gaps," *Science and Technology*, No. 93, 30–39, September 1969.

COMPUTER NETWORKING AS A POTENTIAL SYSTEM AND SERVICE MARKETPLACE

In order to achieve the full potential of automation in information handling, "computer networking" is needed, with its wider-market potential for sharing of hardware, software, and data-base resources by means of wideband, quick-response, digital communication networks of national extent. This would provide for the use of economies of scale in appropriate situations and for the effective delivery of local services. In such an environment, libraries and information activities will be able to share "utilities." Libraries need not be in the computer technology management and operating business, or in the business of doing original scholarship research and authorship to produce the books, reports, and journals they handle, or in generating their own electricity.

A large, flexible, and effective marketplace is needed for making available equipment and services for libraries and information-handling activities. Managers should be able to select from, and change among, processing capabilities and services delivered through electrical communication systems. Such competition and selection may make the difference between impractical possibility and practical feasibility in contemplating a significant degree of automation for libraries, particularly for the small or medium-size library. If such a marketplace were effective now, there would be much more opportunity for development of innovative methods and competition to solve existing problems of operating capability and cost without the necessity to undertake expensive system development by the libraries themselves.

THE ROLE OF INTERACTIVE PROCESSES

Interactive computer service will produce extremely high payoff in supporting the minute-by-minute intellectual endeavors of appropriately trained people, somewhat analogous to the ways privately owned automobiles support the minute-by-minute transportation activities of (trained) people. But, one does not use his private automobile for all transportation services; nor will interactive computer services be used for all information-manipulation activities. Ships, trains, trucks, airplanes, buses, etc., are extremely important parts of the transportation industry, and remote-batch processes, running at different levels of priority according to sensible scheduling algorithms, will be absolutely essential in the computer system.

The digital communications network and the systems connected for supporting computer-based activities and services must accommodate both interactive and batch processes operating in harmony on compatible information files.

WEAKNESSES IN THE COMPUTER SYSTEMS APPLICATION INDUSTRY

In current and past applications of computer-based technology there have been weaknesses in the supplying industry that have hindered the effective use of the existing level of technology. Looking ahead, the computer-systems industry must also mature significantly to provide the conditions for achieving the desired degree of improvement in supplying information to meet nationwide requirements.

Massive storage devices, digital communication systems, microform technology, for example, will be improved tremendously during the next decade, and for the long-term solution these improvements can (and must) be anticipated. The prognosis for these improvements is very good; we need not wait for hoped-for breakthroughs to move ahead.

More important than the expected technological improvements mentioned above is the evolution of a more mature "computer-systems industry."

Consider the building industry: Skilled, experienced firms of architects work with clients to develop overall plans to suit their needs. Special attention is given to the "user features," i.e., to the features that the buildings' users perceive. Many competitive firms of building contractors and subcontractors build what the architect planned. There are many conventions for carrying out monitorship responsibility, inspection functions, and other functions that provide organization and control for the overall process. In addition, there are established procedures for the processes of bidding, negotiating, and accepting.

In light of the improvements considered possible but not yet made, the number and complexity of knowledge-support systems to be developed, their complexity and rate of evolution and other factors, the computer-systems industry is not very well prepared to provide the really good architects and builders one would desire, however impressive its accomplishments over a relatively short period have been.

Most architects (of computer-based systems) want the building job or are even the sellers of the concrete and steel as well. Within the computer-systems industry, there is no discipline that compares with user-oriented building architecture. At present, computer-systems "architects" are more like the structural engineers, who know how to make the structure strong or reliable, but are not thoroughly trained for shaping the system elements so as to respond sensitively to the subjective and objective functional needs (much less to aesthetics), or even to accommodate the cost/payoff, need/value framework of the user.

Also, the "builders" in the computer-systems industry have been notoriously bad at meeting schedules and budgets, and their products are not easily maintained, modified, or transferred onto other sites.

These systems are often the most complex that man has designed, but if these complex systems are to serve in support of complex human operations, there is an absolute need for the quality of both the production process and of the product to be improved significantly. For instance, there must be cleaner conceptualization for system functions and system levels, and problem-descriptive means must improve. Also, the processes of design, and especially of management, must be improved significantly.

ECONOMICS OF INFORMATION: AN UNDEVELOPED FIELD OF KNOWLEDGE

The unsatisfactory state of knowledge of the economics of information is reflected in an analysis of the literature done recently by H. A. Olsen.[6] He remarked that the literature potentially relevant to this topic is extremely diffuse. Apparently there is little inter-

[6] Harold Anker Olsen, *The Economics of Information: Bibliography and Commentary on the Literature*, Washington, D.C.: ERIC Clearinghouse on Library and Information Sciences, ASIS, 1971.

action between the economics and the information communities, as shown by virtually no cross-referencing between them in the literature. The work done in this field is mostly on cost structure, with little on value. Too often, moreover, the basis for comparison between two situations or environments is not valid. Clearly there is an education and communication gap that leaves no basis, at this time, on which to make decisions based on complete value/cost analysis.

UNCERTAINTY OF ECONOMIC DIMENSIONS

Data on actual services rendered and on total costs for the library system of the United States are incomplete and inadequate for well-informed decision-making. Annex A discusses some of the data that are available and concludes that certainly we are considering a multibillion-dollar national expenditure.

Attempts to derive either a meaningful lower bound or a useful upper bound for these expenditures are not possible without extensive data gathering and analysis well beyond the means available for this study.

No one has been able to determine the value of information in terms that allow the normal management and entrepreneurial mechanisms to shape the developments to improve the timeliness, accuracy, and selectivity of information systems. While it is desirable to strive for handling information systems in ways guided by conventional economic value analyses, planning and decision-making cannot depend on such factors now.

ORGANIZATIONAL WEAKNESSES OF COOPERATIVES

A nontechnical factor to be considered relates to organizational structure and effectiveness in connection with library cooperatives. This factor is of particular importance because the benefits of the application of information technology depend on effective joint action and shared work.

A broadly based study by E. E. Olson,[7] using responses from 89 cooperatives, reveals a number of organizational, sociological,

[7] Edwin E. Olson, "Interlibrary Cooperation," Final Report Project No. 07-1084, Contract OEC-1-7-071084-5017, USOE (HEW), Sept. 1970.

and personnel factors relating to most library cooperatives. In general, Olson found that the cohesiveness of such collective activities is weak because of organizational and financial structure. The primary attention is on improving the *means* of cooperation (e.g., collective automated services); but, too often, the *ends* of cooperation seem to be to assist the member libraries in accomplishing their own goals rather than to enable the cooperative to respond as a whole unit to the changes in the total environment. Further, along with the attention to development of physical *means*, there does not seem to be a sufficient development of people needed to carry out the necessary development and operations.

Another source of instability noted in the Olson report is in the economic bases of the cooperatives, which in most instances are not under their control, making them dependent on the continuance of outside support.

PROBLEMS OF PROPERTY RIGHTS

It is a legal requirement that, in providing access to information, property rights of individuals and organizations be respected. Study of means of ensuring that these rights will be preserved has not been undertaken as part of the investigations reported here. However, it must be realized that failure to deal with these problems properly may seriously hinder and delay bringing the ultimately desired and technologically possible computer-based information systems into existence.

INNOVATION IN EXPENSE/REVENUE MECHANISMS

In addition to technological and managerial innovations in libraries and information activities, new approaches may be necessary in administering expenses and acquiring revenue. Consider the following speculations as illustrations of the degree of innovation that may be necessary.

Is it at all possible that the "information industry" could operate with full cost-accounting in a business-like way such that use and revenue are more directly related? Consideration of this concept was not in evidence during the site visits.

The term "information-utility industry" is brought forth regularly for analogy in discussions of this sort. To risk "old-hatting" a concept in exploring this line of thinking here, we observe that an electric-power utility, for its larger customers, must impose a monthly "demand-charge," which represents the costs of guaranteeing that the agreed-upon quantity of power will be available upon demand. The utility company must establish an expensive, general system of energy converters, generators, transmission lines, and distribution facilities, and for each big customer plant the utility company must install a specific system of lines, controls, transformers, and other equipment of this nature that is adequate to meet the customer's demands. The peak demand that a customer may make (i.e., the maximum number of megawatts he can draw during a given time of the week/day) is very carefully negotiated and monitored because of its economic significance in the business of supplying power.

Similarly, an information utility would have to charge big customers for "negotiated demand availability," as well as for specific services. For one example, an academic department would receive a hefty charge for maintaining a large, seldom-used collection in a condition of ready demand for current use.

Such a charge may exceed the current budget of a specific department, and it might be harder for budget allocators to give the money to that department than to the library. A danger is that the departmental decisions on how and where to spend the money might be at variance with some broader institutional need. However, this would be one way to put the decisions about financing library costs associated with current use where they are more relevantly related to cost/value tradeoffs, even if they are only qualitative, and should have the appropriate effect of distributing the budget-defense burden to the proper place.

Under such a system for allocating costs, the current services librarian could contract for support from the appropriate segments of the industry for collection-management mechanisms, cataloging and indexing services, business support systems, and for other systems componentry.

The current services portion of the libraries may collectively subscribe to and support first-level "digestion" agencies that handle the portion of the problem dealing with the initial organization of the material for storage and use, the preparation for adaptation to unknown future requirements, and the maintaining of basic collections in readiness for supply to service libraries. The revenue to support the "digestion agencies" would then be derived from service revenues, contributions from responsible industries with an interest in the maintenance of a complete collection maintained in a particular subject area, and governmental appropriation. There are existing activities that relate to local collection support in the backup/lending libraries, such as the Center for Research Libraries in Chicago, the cataloging activities of the Library of Congress, and the products of the abstracting and indexing services. Insight into the present operation of interlibrary loan activities is given by the recent study, *A Review of the Availability of Primary Scientific and Technical Documents within the United States*, supported by the Office of Education (HEW).[8]

INTERPLAY OF NONTECHNOLOGICAL FACTORS

There is a very complex interlocking among the economic, political, traditional, attitudinal, and other significant factors of the problems to be solved. Any solution of merit must cause or evoke large and pervasive changes, which may be expected to cause many fairly traumatic adjustments by individuals and among institutions. To provide for reasonable degrees of acceptance, the solution must be characterized by a framework that is simple and direct in its essential details. The complex changes and adjustments should follow from the many sectors and groups adapting relatively autonomously to local situations with unambiguous, simple, straightforward guidance derived from the overall framework—as contrasted with a plan that involvės a high degree of centralized (or interfaction) planning, coordinating, monitoring, and enforcing. The environment must be one that makes it seem inevitable that changes will occur and that all parties concerned must do their part.

NECESSARY NATIONAL PROGRAM CHARACTERISTICS

A national program must have both long-term and immediate objectives. It should be bold and massive and leave no doubt that the

[8] James L. Wood, *A Review of the Availability of Primary Scientific and Technical Documents within the United States*, Final Report Project No. 7-0930, Contract No. OEC-1-7-070930-5145, USOE (HEW). ED Numbers 046437, 046438, 046439 (3 volumes). Washington, D.C.: ERIC Clearinghouse, ASIS. Oct. 1969.

nation is committed to providing the necessary information services for all U.S. citizens in the future.

Current means for effecting collaboration on common problems are inefficient, and hence ineffective, even when the intent and desire to cooperate exist. A comprehensive and effective national information policy does not now exist despite valiant efforts to establish one. The leadership needed to unify all segments of the national information community has not yet emerged.

In working to achieve a national system one must keep in mind a general principle that problems of a global nature cannot be solved locally, and local problems are only clumsily handled globally. In this statement, the terms "global" and "local" have a meaning analogous to the concepts that apply to computer programs in describing the scope of definition of names for entities, forms of those entities, range of control of program execution, etc. One indication of effective computer program design is the degree to which the definitions and functions have been "sorted out" ("partitioned") among various chosen interfaces. These interfaces not only require appropriate arrangements for exchange of data and control, but they must be chosen at points of relative stability within the system and with the minimum possible interaction with interfaces at other levels of the system. This design attitude is important in achieving a localization of the impact of local change within the global framework (usually referred to by programmers as "transparency"). This is the essence of truly effective modularity.

These lessons have been learned in the field of computer programming through intense work and sometimes bitter experience. They have been learned in the environment of the most automatic of mechanisms, computer programs. Designers of computer-based library and information systems that are intended to have a high degree of automation must understand the generalization of these concepts if they are to be successful in building a national system that will have a long-term existence and that can be brought into being only in gradual stages.

PLANNING AND MANAGEMENT OF INNOVATION

In the planning and managing of national efforts, a misconception frequently shows up in activities that introduce new technology.

This misconception is not unique to the library and information systems field but it certainly applies to it; it could be perceived in the observations by the Panel, and it was known in the experiences of its members.

The "systems approach" is a popular concept widely quoted and used, and often misused, in the planning, development, and operations of entities represented by a multiplicity of interrelated functions. The technique grew out of the "system engineering" methods that originated in the late 1930's for the development of the nationwide telephone system, has been applied to both military and civilian situations of national extent, and has been enriched and extended by extensive intellectual work in mathematics, engineering, operations research, and management science. It is a powerful and essential tool for development and operations of nationwide scope and impact. But, like all tools, it can be improperly applied.

Among many characteristics of the systems approach is the strategy of having a conceptual model for the totality of the operations of interest, which identifies subproblems and their interrelationships. Such analysis of the "system" may be carried to several levels. Subproblems are then assigned to or assumed by separate groups for solution. The results are intended to be incorporated into the overall system. According to this strategy, specialized knowledge and skills can be brought to bear on specialized problems and undesirable duplication of effort can be avoided.

The erroneous notion is that the results will follow automatically if the formal steps of the theory are carried out. Major factors that can lead to failure follow:

1. The statement of subproblems to be solved is only the first step in design. Rarely is true separability of problem segments achieved.

2. Exploratory projects and developments undertaken in the absence of critical understanding or knowledge cannot be reliably counted upon to produce operationally satisfactory results and must be backed up by alternative methods until they do. Sometimes apparent duplication is the only way to avoid overconstraining a problem and thus preventing its solution, or for preventing premature commitment to an alternative yet unproved.

3. A conceptual model alone is not adequate for design control for assured subsystem compatibility. Very detailed and careful coordination is necessary to ensure that interpretations of the conceptual model are correct and remain consistent.

4. Even with careful planning and development management, various developments proceed at different paces, and provisions must be included to permit innovations to be introduced when they can be brought to fruition. In the real world, plans cannot be based on idealized conditions, and they must be capable of tolerating change with evolving experience.

5. Institutional inertia and competition are not usually taken into account in system design and development, but they are very real, sometimes overriding factors in the design and transitional problems to be solved.

Probably not many readers will contest the acceptability of the generalities stated above. They are almost truisms! Most involved personnel, it is believed, will expect that they are being observed. Nevertheless, advocating these methods on a national level, yet acting consistently with them in specific local planning and local decision-making, may be quite something else.

In the sampling of library and information-system activities involving computers, the recognition of many of these factors and principles was evident. However, it remains to be demonstrated that all parties are in fact acting consistently with them on a national basis. More critically, the perceptions of a national information system model differ from one environment to another. Thus, it should be no surprise that institutions may fail to cooperate, or, worse, may take conflicting actions based on these perceptions.

THE ROLE OF PILOT SYSTEMS IN DEVELOPMENT STRATEGY

In developing and installing new complex systems, it is frequently necessary to gain realistic experience to determine the complete specifications, the operating economics, and the operational experience to proceed to the full system needed. Thus the various stages of theoretical research, experimental research, pilot development, and finally operational development come into play.

The pilot development stage is very important for gaining realistic experience on a scale short of the full operational commitment. In some industries this is a well-established development strategy, and it is appropriate for the problem at hand—the application of information technology to the library and information system challenge. Such a pilot national system must treat both the technological and nontechnological aspects we have discussed.

ANNEX A

DIMENSIONS OF COST AND USAGE

INADEQUACY OF AVAILABLE DATA

The Panel found that data on actual services rendered and on total costs for the library system of the United States are incomplete and inadequate to characterize the total national system. Available figures are not on a consistent basis for all classes of libraries.

Some gross cost estimates are cited below to suggest the order of magnitude of the resources employed in the national library system. However, available quantitative descriptions of usage are superficial. The populations of areas served by public libraries,[9] numbers of registered borrowers (i.e., numbers of library cards issued), the enrollments of universities and the library seating capacities,[10] and other similar figures are available in the sea of numbers recorded about libraries of some types. However, none of them is very helpful in measuring the actual amount and type of usage of the information in libraries as a whole. A brief snapshot survey of one university library indicated that a very large fraction of the students who go to its library do so for reasons unrelated to the collection stored there. They do homework based

[9] *Bowker Annual of Library and Book Trade Information*, New York: R. R. Bowker Co., 1970.
[10] Robert B. Downs, *University Library Statistics*, Chicago: American Library Association, 1970.

on their own texts and notes, or use the facilities as a general social meeting place.[11]

One serious attempt to analyze various parameters of library activity, using time-series analysis, was undermined by the unreliability of available data.[12] The author of the report quite clearly and responsibly pointed out the serious data-validity problems he discovered just prior to completing his report. His warning gives a valuable insight into the hazards of measuring and evaluating library performance. In many cases, data that nonrespondents should have supplied is simply entered as zero in cumulative statistics.

A current study[13] is expected to provide fresh insight into the number and types of federal libraries and their activities. No such figures are available for industry, and studies undertaken to collect them have been abandoned before completion.[14]

SELECTED DATA FOR INSIGHT INTO DIMENSIONS

One lower bound on the size of the library system is given by the American Library Directory, 1970–71 edition, which records 27,180 libraries in the United States categorized as public, college, special, armed forces, law, medical, or religious.

An illustration of the order of magnitude of various library and information activities is given in Table 1.

Subject to the uncertainties inherent in the heterogeneous and fragmentary data given above, the Panel assumes that a lower bound for the amount the nation is now spending on formal library and information services is at least $3 billion dollars per year, an expenditure that deserves national attention. Others think that, if all pertinent expenditures were accounted for, the true figure would be much higher; one individual estimated the order of $10 billion dollars per year. In any case, the categories for which numbers are listed in the table represent, for the most

[11] Comment during discussion, W. Locke, Session on Computer in Service to Libraries of the Future, Spring Joint Computer Conference 1969, Boston.
[12] *On the Economics of Library Operations*, Princeton, N.J.: Mathematica, 30 June 1967.[13]
[13] Task Force on Automation, Federal Library Committee, private communication.
[14] Special Libraries Association, private communication.

TABLE 1 Gross Budgets of Some Library and Information Activities in the United States

Type	Number	Volumes (and other items)	Population Served (nearest 100,000)	Total Annual Budget ($)	Date of Data
All college and university libraries[a]	2,370	305,000,000	7,000,000	509,800,000	1967–68
Major academic libraries[b]	76	137,693,729	–	238,385,663	1969–70
Library of Congress[c]	1	59,890,000	–	43,856,300	1969–70
National Agricultural Library[d]	1	–	–	3,974,000	1969–70
National Library of Medicine[e]	1	1,312,956	–	21,500,000	1969–70
Public libraries (serving localities of population exceeding 25,000)[f]	1,135[g]	188,000,000	125,000,000	421,000,000	1968
Elementary and secondary school libraries[h]	49,158	171,585,746	37,300,000	–	1962–63
Special libraries (industry)[i]			unknown		
Government STINFO[j]	–	–	–	534,630,000	1968 (FY)
Dep. State, foreign dissemination[j]	–	–	–	295,000,000	1968 (FY)

[a] Bronson Price, *Library Statistics of Colleges and Universities–Analytic Report, Fall 1968*, Washington, D.C.: Gov't. Printing Office, 1970.
[b] Association of Research Libraries, Academic Library Statistics 1969/70.
[c] Annual report of the Library of Congress. In FY 71, including the Legislative Reference Service, the budget was $53,359,209; in FY 72, the expected budget is approximately $68,000,000. Private communication.
[d] Office of the Budget, National Agricultural Library.
[e] Office of the Budget, National Library of Medicine.
[f] Ruth L. Boaz, *Statistics of Public Libraries Serving Areas with at Least 25,000 Inhabitants (1968)*, Washington, D.C.: Gov't. Printing Office, 1970.
[g] Of the 1,135 public libraries of this class, the data listed are for the 1,057 libraries that responded to the questionnaire circulated by the U.S. Office of Education (HEW).
[h] Richard L. Darling, *Public School Library Statistics, 1962-63*, OE-15020-63, USOE (HEW), Sept. 1964.
[i] Special Libraries Association, private communication.
[j] Private communication.

part, activities supported by local or national taxes and are hence of public concern.

It would be desirable to be able to estimate an upper bound as well as a lower bound to the expenditures for information memory and access mechanisms, thereby bracketing the actual amount. The Panel was unable to find anything adequate for this purpose. The only existing attempt that might be considered relevant is an attempt by Machlup,[15] published in 1962, to derive an economic dimension for the "knowledge industry." The knowledge industry, the collection of activities that produce information rather than hard commodities, which includes the "knowledge" aspects of all private and public activities, would seem to be the portion of national effort that should be enhanced by an effective national library function and related information activities. Based on a very broad definition of "knowledge production," Machlup commented that "total knowledge production in 1958 was almost 29 percent of adjusted GNP—provided all our estimates are accepted, our conclusions granted, our omissions disregarded."

In his analysis, Machlup concentrated on the output of knowledge activities, as was appropriate for his study, and the role of the formalized storage of and access to information was not fully explicit, making it impossible to derive a meaningful estimate for the storage and access activity alone.

The Panel therefore concludes that the true bounds of the national information storage and access function are clearly not adequately defined and that the economic dimensions are poorly identified. However, it is clear that the magnitude is large enough to deserve national attention.

COST TRENDS

Data on university and college libraries are more available and complete than for other categories. A hint of trends is given in the data for 3 years as summarized in Table 2.

Using these data and computing the ratios of dollars per student and dollars per volume, we find a striking escalation of unit costs

[15]Fritz Machlup, *The Production and Distribution of Knowledge in the United States*, Princeton, N.J.: Princeton University Press, 1962, p. 362.

TABLE 2 University and College Library Growth

Year	Number of Libraries	Number of Volumes (nearest million)	Students Enrolled (nearest 100,000)	Budget (nearest million)
1962[a]	1,985	201,000,000	3,900,000	184,000,000
1966[a]	2,207	265,000,000	5,900,000	320,000,000
1968[b]	2,370	305,000,000	7,000,000	510,000,000

[a] See reference 16.
[b] See reference 17.

(Table 3), significantly exceeding increases in the Consumer Price Index.

The sparse numbers listed above must be used with great caution, but they suggest that costs are going out of control at an increasing rate. The budget crises of many large libraries tend to support this implication.[19] Compounding the difficulties of escalating costs, the amount of money available to pay those costs is leveling off and in some cases actually decreasing.

The dynamics of library growth in the academic sector are illustrated by statistical compilations issued by the Purdue University Library.[20] This study plots library growth factors from 1951 through 1968–69 and projects the growth curves to 1980.

TABLE 3 Unit Cost Escalation

Year	$/Student	$/Volume	Consumer Price Index[a]
1962	47	.92	105.4
1966	54	1.21	113.1
1968	73	1.67	121.2

[a] See reference 18.

[16] Douglas M. Knight and E. Shipley Nourse, ed., *Libraries at Large*, New York and London: R. R. Bowker Company, 1969.
[17] Bronson Price, *Library Statistics of Colleges and Universities–Analytic Report, Fall 1968*, Washington, D.C.: Gov't. Printing Office, 1970.
[18] Bureau of Labor Statistics, Bulletin 1660-27, October 1969 (Base period: 100% for 1957–59).
[19] The Association of Research Libraries, Minutes of the Seventy-Seventh Meeting, January 17, 1971, Los Angeles, California, p. 29–42.
[20] Oliver Dunn, *The Past and Likely Future of 58 Research Libraries 1951-1980: A Statistical Study of Growth and Change*, 6th Issue, Lafayette, Indiana: Purdue University Library, 1970.

The average composite of the 58 university libraries included in the Purdue study indicates that the number of volumes held doubles every 17 years; the number of volumes added to the collections each year has doubled every 9 to 12 years; and library operating expenses double every 7 years. The report concludes: "In short, the records of growth since 1951, including the most recent years, and the unfaltering growth of even the largest libraries, indicate that this growth may not soon decelerate."

TARGET FOR COST REDUCTION

Library staff salaries range approximately from 50 percent to 75 percent of university library operating costs.[21] A clue bearing on the nature of library operations and the nature of the labor involved is afforded by William Locke's analysis of a given university's library budget of approximately $2 million. It is a demonstrable fact that a considerable part of the workload in any library consists of repetitive tasks such as cataloging, searching files, shelving, preparing lists, checking on circulation, and other routine repetitive labor-intensive functions.[22] These are the functions in which early application of new technology is most promising for decreasing cost of operations.

INFLUENCING THE INFORMATION-TECHNOLOGY INDUSTRY

It has often been stated that the library and information-processing community is too small to have a significant impact on determining the characteristics of equipment and software to be marketed by the information-technology industry through normal market mechanisms. Thus the approach has been to "make do" by adapting equipment and processes produced for other purposes. In view of the total costs at stake, the Panel believes that a coherent library and information community would have the market potential to influence the nature of products and services available from the information technology industry.

[21] Robert B. Downs, *University Library Statistics*, Chicago: American Library Association, 1970.
[22] W. N. Locke, "Computer Costs for Large Libraries," *DATAMATION*, 16(No. 2): 69–74. February 1970.

LEVERAGE OF INFORMATION ON NATIONAL ECONOMY

Scrutiny of information activities in themselves yields a very incomplete picture of the economic dimensions and the importance of the information problem. These activities doubtless have massive leverage on the total conduct of industry, education, and government. Many studies and reports have recognized this importance unequivocally in the case of science information,[23] and similar impact would be expected for all types of recorded and organized information.

The continuing inadequacy of information services has required organizations to do much for themselves in inefficient and duplicative ways, expanding resources to collect information on their own because useful information that has already been collected cannot be found or delivered for use. Often, even if the required information is available and at hand, it is not in a form that is easily assimilated and organized or combined with information generated within the organization.

If the bulk of formally recorded information were readily available through responsive library and information services, the internal resources of organizations could better be utilized in the generation and organization of information from within, merging it with the information from outside, and making it available for effective use.

Just as transportation systems are one key to the development and conduct of commerce, information systems are increasingly essential to modern society. The long-term memory function in the information system, the "library function," can be very important if its current embodiments evolve to be appropriate to the modern environment of large volumes of information and need for quick response.

[23] W. O. Baker, C. Benjamin, C. P. Haskins, E. Hutchisson, W. C. Johnson, D. K. Price, H. Scoville, and A. Waterman, "Improving the Availability of Scientific and Technical Information in the United States," The White House, 1958; J. H. Crawford, Jr., G. Abdian, W. Fazar, S. Passman, R. B. Stegmaier, Jr., and J. Stern, "Scientific and Technological Communication in the Government: Task Force Report to Special Committee for Science and Technology," Washington, D.C.: President's Advisory Committee, 1962 (available from National Technical Information Service, Springfield, Va., AD 299 545); A. M. Weinberg, W. O. Baker, K. Cohen, J. H. Crawford, Jr., L. P. Hammett, A. Kalitinsky, G. W. King, W. T. Knox, J. Lederberg, M. O. Lee, J. W. Tukey, E. P. Wigner, and J. H. Kelley, "Science, Government, and Information," A Report of the President's Science Advisory Committee, The White House, 10 January 1963.

ANNEX B

SITE VISITS TO KEY ACTIVITIES AND INSTITUTIONS

EXPLANATION TO SITE HOSTS

By way of introduction to this report of site visits, the following outline of Panel interests, which was sent to sites prior to visits or discussed with our hosts upon arrival, is reproduced:

In discussing with you the planning for a visit of our Information Systems Panel of the Computer Science and Engineering Board, I promised to provide you with background on the Panel, its mission, and its approach. Enclosed is an excerpt* from a recent report to an open meeting of the CSEB which should outline these things, and help establish the environment for our visit.

I assume you are aware of the Computer Science and Engineering Board which is an operation of the National Academy of Sciences to study various aspects of the use of computers in our society and their impact on it. As a result, the CSEB is expected to provide advice to national decision-makers as one of the traditional roles of the National Academy of Sciences. The Information Systems Panel reports to the CSEB and has been organized to consider computer-related topics relating to information systems of all types.

*The excerpt discussed the background and mission of the Panel as outlined in the Preface to this report.

For case studies in our overall effort, we are interested in:

1. Case studies in operational information support activities where there has been vigorous use of computer aids.
2. Discussion of what has been learned during this experience—both good and bad.
3. Discussion of scaling factors that might make some of the techniques and concepts of importance on a wider scale.
4. Discussion of opinions on the state of computer technology as it applies to applications and as necessary for broader scale applications of analogous methods—both sufficiencies and deficiencies.
5. Opinions on what is important to consider in our task as I described.

For information transfer research and experimentation, we are interested in:

1. Understanding the objectives of a project and where it fits into the universal information transfer picture (model). Assuming success, what will be the impact?
2. Separating the conceptual from the visible means for investigating concepts (or inventing).
3. Understanding the "design of experiment" (methodology, measure, hypotheses).
4. Discussing the adequacy of the technology (hardware, software) for the research and experimentation [and] for extension to broad scale use.
5. Discussing the associated changes in environment which must occur if the techniques prove successful and should be replicated or propagated.
6. Discussing the means by which the results can be exported and applied (administrative, funding, commercial, model plans and directions).

I hope that this explains our mission and interests and contributes to an effective discussion. We would appreciate your assistance in ensuring all points of view from your organization are presented in our discussions: research, development, operations, and management.

ORGANIZATIONS VISITED

To provide an empirical base for its work, beyond published information and the first-hand experience of its members, the Panel made visits to several sites. The sites were chosen to sample a broad range of activities—not to attempt a comprehensive survey or evaluation of all present library automation or computer-based information projects. The organizations and projects visited are described here in the chronological order in which they were visited.

Bell Telephone Laboratories, Murray Hill, New Jersey, Main Technical Library

The main library, two other major collections, and a total of 20 information centers throughout the various locations of the company are managed with a set of computer-based files and programs. The system accepts on-line input at the time transactions take place and produces a variety of reports of activity and status. On-line query of information relating to the total collection is also used. Decisions about acquisitions, location of subcollections, and removal of inactive material are made weekly, using system-produced reports. Querying of requests for items and other direct support of circulation are provided. The system has become an integral part of library management and has provided much greater responsiveness to user needs and better collection control. There is very high "cost-consciousness" in the operation and it is considered to be cost-effective by management. The library serves a community that is very active in the use of information in support of research and development. Through a separate system, experimentation has been conducted on selective dissemination of internal reports with both originator and recipient-directed distribution.

Harvard University Library

Three units of this largest of U.S. academic libraries (more than 8,000,000 volumes) are using computer technology. Widener, the central library, is producing a book-form version of its unique shelf list. It has a partially computerized circulation system, and also uses computer programs to support ordering and book-fund accounting. With the acquisition of 60,000 new volumes per year, the cataloging and check-in load is very heavy, creating considerable backlog and delay. Because the MARC tapes would supply such a small percentage of cataloging input, they had not been usefully employed at the time of the Panel visit, although they were being acquired. In such a large environment, change occurs very slowly. Nearly all experimentation and development in Widener has been financed out of the regular library budget without foundation or federal support and therefore has necessarily been less extensive than in some other institutions. The Baker Library at the Harvard Business School has a computerized serial receipt system; computer-printed lists of current periodicals have replaced the corresponding entries in the card catalog. The Francis A. Countway Library of Medicine, site of a MEDLARS Search Center, has one of the AIM terminals provided by the Lister Hill National Center for Biomedical Communication, and is a part of the SUNY Biomedical Communication Network; the latter terminal is also used in a shared cataloging project with the National Library of Medicine and the Upstate Medical Center.

Project INTREX,[24] *Massachusetts Institute of Technology*

In distinct contrast to a traditional library environment, INTREX is an attempt to demonstrate highly interactive facilities for information-collection use. It can be viewed as a laboratory that includes a test collection of about 10,000 documents and it uses time-sharing to support access to that collection. Because equipment was unavailable from commercial sources at the inception of the project, much time and effort was spent in the early years of the project to design and construct the laboratory.

Even though the project assumes a future orientation in carrying out interactive retrieval experiments, it still must be based on an "image" technology for full text of documents. For wide-scale use of techniques modeled in this laboratory, reductions in the prices of mass memories and of wideband communications are needed.

[24] J. F. Reintjes, R. S. Marcus, P. Kugel, R. L. Kusik, D. R. Teicher, D. R. Haring, and J. K. Roberge, "A Progress Report on Project INTREX," AFIPS Conference Proceedings Vol. 34 (SJCC69) pp. 457–490, AFIPS Press, Montvale, N.J. 1969.

Project TIP, *Massachusetts Institute of Technology*

The TIP project is an operational pilot system for information access via various bibliographic methods. It has been developed on the MIT time-sharing system and is in its second version (operational) with the third version in the design stage at the time of the Panel visit. Various subject fields have been handled, including physics, MAC project reports, literature of interest to the Sloan School, and eighteenth century Italian opera.

One very noticeable desire of the system users was for the ability to manipulate the products of searches by forming private subfiles, annotating entries with personal additions, and reorganizing and reformulating retrieved material for report purposes. Since the MIT time-sharing system has very general text-editing and manipulating capabilities, the TIP system has exploited those capabilities.

Because the system builds so heavily on the unique MIT time-sharing computer system, it is not easily exportable to other locations.

National Library of Medicine

The entire institution includes activities in three areas of interest to this study: a large national library devoted to a single specialization; document analysis and indexing services; and experimentation with remotely accessed retrieval systems, carried out by the co-located Lister Hill National Center for Biomedical Communication. Computer systems have been deeply involved in these activities. MEDLARS II was intended to integrate the computer system support of many activities, including library acquisitions, document analysis and indexing, production of secondary publications, and provision of search services. The production of *Index Medicus* and retrieval searches by MEDLARS I provided a base of experience for the development of the new system.

The development work has been long and arduous, which is typical of many computer-based systems of such complexity. However, the basic requirements for support of the operations of the National Library of Medicine remain. As has also been illustrated in many other efforts contemporary with MEDLARS II, the designers, if given the opportunity to replay history, would increase the time and effort devoted to the early systems-analysis phase of the project. To handle the necessary requirements, the degree to which they are integrated has been relaxed and developments redirected accordingly.

The experimentation with remote on-line access involves terminals of several types and experimental data bases at several locations throughout the United States. Also, there is a project specifically concerned with various audiovisual media.

Library of Congress

The size and environment of the Library of Congress magnify the crises and problems faced by all information-storage institutions into epic proportions. Dating back as far as 1950, various studies and efforts have been commissioned to examine methods for coping with this growing heterogeneous collection. A notable report issued in 1963[25] included a comprehensive analysis of automation possibilities and proposed a massive automation effort. The effort was not undertaken, but projects performed by the Library of Congress staff and by contractors have continued to examine the problems and work on some of them.

Although it has not been given the formal status, the Library of Congress functions in many ways as the *de facto* national library. Since 1901, as a by-product of LC cataloging, bibliographic information has been made available to all libraries in card form, and thus it has become a focal point and standardizing force for such activity. In 1965, Title II-C of the Higher Education Act charged the Library of Congress with providing cataloging information in the national interest.

The best known result of computer applications at the Library of Congress is the MARC project. It has received worldwide attention and is producing bibliographic information in machine form for distribution to many libraries. At the time of the Panel visit, the input involved English-language monographs—about 50,000 items per year.* Extension to currently published books in German, French, Spanish, and Portugese is planned when resources become available.

[25]The Council on Library Resources, Inc., *Automation and the Library of Congress*, Washington, D.C.: Library of Congress, 1963. (Known as the "King Report")
*By mid-1971, the input rate had risen to 1,200 per week with an accumulated file exceeding 150,000 titles.

Institute of Library Research, University of California (Berkeley)

This long-term program of research and development in support of the library operations of the nine major University of California campuses originated in 1963. Not surprisingly, as the complexity of design of computer support for libraries has unfolded, the time needed for full operational penetration of computer technology into the area has lengthened, as in all projects of this type. By combining forces, the nine campuses have been able to mount a much more concentrated effort than any of them would have been capable of alone. However, financial problems caused by limitations on appropriations and by the impact of unbudgeted new academic programs have been disruptive.

The University of California at Berkeley community, in particular, ranges from undergraduates to its large group of Nobel laureates. Most of the advanced research in high-energy physics depends on the specialized information sources and facilities of the Radiation Laboratory. The main library is used in support of the other areas of graduate research and academic programs.

In the context of the nine campuses, early experience in transplanting systems from one location to another was found to be difficult, but present emphasis is on design for multiple use and concurrent automation of nine libraries to reduce total development costs. The first two modules for completion and use are acquisition and circulation control, which offer the greatest potential development savings when designed collectively for operation by each of the nine libraries.

SPIRES/BALLOTS *Project, Stanford University*

At this site, discussion centered on the design and implementation of an integrated data facility to support activities such as those envisioned in the SPIRES and BALLOTS information-handling projects. SPIRES stands for Stanford Public (née Physics) Information Retrieval System and is aimed at an augmented bibliographic retrieval capability. BALLOTS stands for Bibliographic Automation of Large Library Operations Using Time Sharing. The projects started separately in 1967 and were later combined. At the time of the Panel visit, the merged project was in its early stages and had established a vigorous and optimistic schedule. Extensive use of system-analysis tools was evident in planning and design. Both library automation and information-retrieval objectives are included in the total project.

Very specific views on technology needs (as compared to commercially available equipment in 1970) were expressed, including the need for mass digital storage with an order-of-magnitude improvement in capacity, speed (especially the avoidance of input/output wait due to disk arm movement), and unit cost of storage. For large-scale usage of terminals, costs must go down by a factor of 4 or 5. Similarly, present common-carrier rates for communication costs and poor long-distance line quality hinder large-scale networks extending beyond a few hundred miles. Data-base software, in general, is inadequate, especially in connection with recovery after system failure. A particularly crucial problem in design of data base management systems is the handling of multiple-indexed dynamic files.

Although provisions for handling source data in digital form are included in the concepts of this project, full-text storage on paper or other image media is anticipated to be necessary for a long time due to lack of necessary storage capacity at reasonable cost.

For exportability of results, networking by means of electrical communications appears very attractive for maximum sharing of resources with minimum need to install complete or partial systems in multiple locations.

Social and institutional changes for maximizing use of automation can be expected to be difficult in some instances. Immediate use of on-line interactive systems by all users is not anticipated; a trained intermediary (e.g., reference librarian or information specialist) may be necessary.

As with so many projects, current economic conditions are placing extreme burdens on the viability of the project.

IBM, *Los Gatos Laboratory*

An automated library-management system had been designed and installed for pilot use.[26] Its design was based on the level and type

[26]Council on Library Resources, Inc., *13th Annual Report* for the period ending June 30, 1969, Washington, D.C.

of activity of the library of a small college and was developed using an experimental time-sharing system. It is a comprehensive system for the operation and management of a library. The system is not available commercially, but IBM and others are continuing experimentation with it.

Augmented Human Intellect Research Center,
Stanford Research Institute

The activities at this center, which have built up over many years, are relevant to the substance of this study in three respects. Extensive capability for interaction with structured text has been the basis for development of processes for close support of human intellectual activity. The center is also preparing to be the information center for the ARPA experimental computer network. In addition, a feeling for the use of such a system was obtained in assembling some of the draft material for this report, illustrating the potential of a very powerful information-handling facility as well as the intellectual discipline necessary to obtain full benefit from automated support.

University of Chicago Regenstein Library

The University of Chicago Library system was chosen for a visit since it is an example of a large university operation that uses MARC tapes extensively. It has also been influential for a number of years in development of the system from the "user's point of view," and it is one of the major long-term library-automation projects in the United States. At the time of the visit, the MARC input represented over 20 percent of the cataloging input and was considered a definite economic advantage, even though the MARC information is modified for use in the Chicago system. The automated effort at this installation began in earnest in 1966. From the outset, the intent was to build a broad data base compatible with wide and diverse applications dependent on the constructed data base, counter to the view of some that accumulation of data should be limited to that necessary to support a limited number of library operations such as circulation and bookkeeping.

The library-automation programs use the computer installed in the university computing center. This arrangement is said to work quite satisfactorily since the personnel at the computing center are understanding, solicitous of the needs of the library personnel, and the library is on a pay-as-you-go basis with the center. Library-automated functions are processed on an IBM 360/65 computer. Previously the 360/30, 360/40, and 360/50 systems were used. Programming difficulties were experienced with each changeover.

It was commented that relatively large research libraries have on the order of one to one and a half million titles, with bibliographic entries estimated to require 600–800 characters each for machine storage. Storage is not a problem except for cost, and the cost picture will be improved by continuing technology development. However, the file organization and manipulation ability are recognized problems. Among other problems are adequate access ports to large-scale storage facilities and the lack of a character set sufficiently large to handle Russian and other foreign languages.

The Chicago library had approximately 80 catalogs in use at the time of the visit. One is on-line with 40,000 records. The large bibliographic file in the fall of 1970 was composed of 100,000 records and was being expanded at the rate of 50,000 entries per year.

As with many other projects, however, the job is more difficult, takes longer, and is more expensive than originally anticipated.

University Computing Company

The processing handled by this commercial activity includes both remote batch processing and on-line processing, with the bulk of the processing in batch mode. The University Computing Company (UCC) has realized savings by consolidating computers at one location, mainly through decreases in maintenance costs and effective management. In the view of company spokesmen, it is less expensive to provide data to a computer by telephone than it would be to install and use a computer close to the location of the user when all support costs of the "in-house" computer are taken into account. Because of this, they have concentrated their processing into a few large centers and have provided nationwide service by means of electrical communication.

The opinion was ventured by UCC personnel that the technology for the automation of many functions in libraries is here, but that the intellectual problem of organizing the large files that library content represents is considerable. Even if the hardware were equal to the task at this time, the problem of file organization and accession would still remain to be solved.

The on-line file-storage capacity now available at commercial service centers is well below the amount required for a large on-line catalog. However, batch processing with tape files is, of course, entirely practical for many purposes.

The UCC objectives for the DATRAN system, which the company seeks to establish as an alternative to existing common-carrier data transmission, were explained. DATRAN is to provide high-speed lines and leased-line facilities for dial-up entry. A new schedule of costs is to be devised based on the availability of 6-second increments instead of the 3-minute rates in effect in 1970. Very short connect times will be made available. Particular attention will be paid to achieving low transmission-error rates as required for effective data transfer.

The New York Times Information Bank

The New York Times was selected for attention as a large information system designed to serve in-house interests better and to be marketed as a profit-making enterprise.

At *The New York Times*, internal information needs are met through the *Times* morgue—the first stop in seeking historical or background material needed by the various research and staff writers. The *Times* is moving to automation in library services, having recognized that the present system is not staying abreast of needs and that its deficiencies are amenable to improvement through automation. This, in general, is the situation that faces most of the libraries of the nation and those other similar installations that have "memory" as a central function.

The major external hard-copy retrospective search instrument the *Times* provides is the *Index*. The heart of the automated system being installed is the abstract, the same informational form used in the *Index*. Access to the abstracts, as well as to full text, will be by subject, name, and geographic reference. Full text back-up will be available on demand from microfilm storage.

Indexing and abstracting of news stories will provide a digital data base for access to a microform collection of full text of the news stories. The capacity of the microform storage device is adequate to store many years of the *Times* and of other publications used for reference by the *Times* staff. The images will be distributed within the *Times* via high-resolution closed-circuit TV to terminals that can display either image information or alphabetical information distributed by computer. Forty terminals, sharing four simultaneous access ports to the image store, will be used in the *Times* for data input, editing, and querying.

Many queries are expected to be answered by the retrieval of index entries and abstracts without retrieving the full text (in image form). Access from outside the *Times* is only to the digital files, possibly backed up by local collections of microforms.

At the time of the Panel visit, the system was nearing full-scale system test.

IBM *Corporate Headquarters and Research Laboratories*

The main systems viewed were the Corporate Management Information System, the Advanced Administrative System, and R&D Networks of Computers. The first two systems are in use by IBM; the research and development effort connected with the third activity is pointed toward commercial application. In addition, and of particular value to the Information Systems Panel, is the Technical Materials Dissemination System used in support of the scientists and engineers in the corporate structure.

The Corporate Management Information System is a highly integrated network of 30 computers using 10 different languages and with terminals at several dispersed locations. This system, now coming into greater use, can supply the user who is cleared to query the stored material with responses on matters of concern to the top-level decision-making elements of the firm. Various means of display are available at the central headquarters at Armonk in a highly sophisticated array. While only about 80 files are physically located at the central headquarters, the system is capable of querying files located anywhere in the corporate-wide system. Access is by specific authorization and considerable emphasis is placed on file security and the enforcement of access procedures.

The Technical Materials Dissemination System supports approximately 4,000 IBM personnel. This is a self-supporting system to a large extent, charges being levied on the user personnel and divisions of the corporation. The computer-search system in use is called TEXT-PAC. Its operation is controlled by user profiles and incorporates 38 programs.

The Advanced Administrative System construction was begun in 1965 with specific requirements in mind. This is a real-time system organized around central files, with transactions conducted in a conversational mode throughout a 12-hour day to accommodate the time difference between west-coast and east-coast installations in the net. This appears to be a very large and expensive information system except when viewed in terms of the value of its use in managing a very large corporation. Traditional methods for this purpose had begun to be severely strained before the new system was installed.

Additional Discussions

In addition to the foregoing site visits, discussions were held with the Defense Documentation center and with participants in the Federal Library Automation Project sponsored by the U.S. Office of Education. Many more library and information-oriented automation projects were brought to the attention of the Panel by individual contact and through the many publications and reports of such projects.

DISCUSSION OF OBSERVATIONS

From these various projects, it is quite clear that useful automation has been brought to bear on the management of information collections and on aids to bibliographic access of industrial and other libraries specializing in current materials. This is particularly so where there is a well-developed computing base used for general computational and business support of the environment in which the library or other information activity fits, as exemplified by the BTL and IBM activities.

The progress of several groups has shown that these methods can be applied locally to large libraries and closely knit groups of libraries, provided that application is approached as a serious large-scale design and engineering project and that an appropriate multidisciplinary design team is employed.

There is no technological deficiency for automation of purchasing, inventory control, handling of bibliographic information, and circulation control. Economic payoff can come from saving labor normally used for repetitive tasks, from being able to manage a collection more efficiently, and from being able to use cataloging information supplied by others. There have been economic barriers in the financing of the high costs of local development of suitable software and there is little satisfactory evidence of multiple use of such systems to spread the cost. Furthermore, there has been considerable disruption of the continuity of library support systems by changes in hardware and operating systems in the computing centers on which the libraries depend.

Experience to date indicates that automation entails a series of technical, intellectual, and economic problems that must be solved by highly skilled people with combined competence in computer technology, bibliographical knowledge, and other fields relating to design and business operations. Expenses include resources for systems-analysis and design, hardware, file conversion, and the expense of operating the system. Because of economic factors it has not been practical to consider computer technology for large-scale text storage for information retrieval. However, libraries contain numerous files for the control of book ordering, serial receipts, binding, cataloging, and circulation, which are suitable for automation under appropriate circumstances. Computerization of these files can contribute to getting ready for longer-term endeavors.

An investment of $70,000 to $100,000 may be required to design the mechanization of a single local library operation, i.e., acquisitions or circulation, at each location where it is attempted. New money must be provided for these development costs, as they are in addition to the budgets required for operating regular library services. Moreover, funds must be found to pay for file conversion and to operate the system (hardware and staff costs). This suggests that only a few of the larger libraries can justify the expense of independent local automation. Alternative approaches are being developed by creating service agencies to provide automated programs for consortia, or regional groups of libraries such as those in Ohio, New England, and California.

One of the lessons from experience in automating information activities is that there should be no anticipation of an immediate cost saving from automation done only at the local level. However, the application of computer technology can produce more effective service for patrons and more efficient records for library management.

The MARC pilot bibliographic information service is "useful" to the extent that it supplies an appreciable percentage of bibliographic input of the receiving organization. In situations in which it does not, it can be regarded only as an experiment, the operational worth of which cannot really be evaluated. Both situations were observed in the Panel visits. Building the files, whether content files or control and access files, to a level of true usefulness requires a large investment without immediate payoff.

For the most part, the computer science and engineering industry knows how to provide for the individual mechanical functions that would serve the needs of library and information system users. However, more experience is needed to provide them cheaply, rapidly, and reliably in the environment of a large evolving system and to be able, in the first place, to design a large complex system in the existing multi-institutional environment.

MIT-TIP and the SRI system both show what can be done with extensive use of computer technology in information handling and what flexibility can be achieved in a research installation. These systems are "useful" as well as experimental, to the extent that the existing facilities will accommodate files of a size suitable for real applications. However, the economics of use outside the local specialized environments have not been established.

The INTREX Project at MIT demonstrates some of the expected hardware characteristics on a laboratory basis and provides a laboratory facility to test hypotheses about how persons may wish to interact with documents and other information, once the documents have been processed into a using facility.

In moving mechanized support into operational situations, the orientation and training of users must be carefully carried to the point at which they know exactly what the system expects of them, how it reacts when they do the unexpected and, finally, how to recover when the system leaves them in the lurch. The workers in the experimental and laboratory environments are "partisans" for the use of computers and communications and are very tolerant, even oblivious, of the unexpected and sometimes unreliable responses of machine systems. In the scientific or computer-oriented environments in which today's large-scale computer systems typically operate, the necessary training is both relatively easy to give and readily accepted. However, such a tolerant attitude cannot be expected of a casual library user and perhaps not even of most library operating personnel who may not be machine-oriented to begin with. The Panel observed these differences in the various environments visited.

The designers involved in most of the activities that the Panel visited cited an important problem, namely, the difficulty in organizing very large computerized files for rapid access and for maintenance, still retaining economy and reliability in operation and flexibility in adding new types of data as they become available. Typically, the large systems operating now are designed around particular file organizations, and any changes other than the most trivial are apt to mean expensive reprogramming. The overall problem of data base design and management is receiving increasing attention in computer-centered research and development, although general solutions are elusive.

A factor sometimes forgotten in the speculations on revolutionizing information-handling through automation is the massiveness of the general library problem. In such institutions as the Library of Congress and the Harvard Library, the physical size of the collections, the extent of the catalog card files, and the amount of activity around them, the amount of material awaiting processing, the glacial pace of manual methods, and the constant requirement for exercising human intellectual judgment in processing are all sobering to observe.

While these huge institutions are only one end of the library and information-system spectrum, they serve to emphasize the complexity and magnitude of the problem of managing and accessing the "national memory." These characteristics—size and complexity—are the very things that suggest the use of automation. However, usually given insufficient attention by computer-system designers is the "inertia" of such operations, the great care with which new methods must be introduced, the requirement for continued service during change, and the slow rate of evolution under such conditions.

Several large projects are applying computer technology to major research and academic libraries or regional cooperatives. The Panel visited only a few of them. In projects carried out at the University of Chicago, Institute of Library Research at the University of California (Berkeley), and the SPIRES/BALLOTS Project at Stanford University, the problems of large computer-system design for local functions are recognized although not yet completely solved.

While these examples represent a wide variation of environments, many common themes exist. The efforts are aimed at economically viable operational systems, not experiments. The computing power for the systems usually is obtained from general-purpose computing centers and the scheduling algorithms and charging algorithms often discriminate against the large-file, continuous-use processes needed in the information-processing systems. The development of multidisciplinary design staffs has been difficult. The system developments have turned out to be more difficult and have taken longer than originally expected. All these projects still require a significant amount of work to reach a fully operational status, and all are facing serious budget crises in routine operations and development activities.

As computing service activities develop into nationwide network operation, they provide a third alternative for libraries and information processors as compared to dedicated equipment or shared use with other parts of the host institution (e.g., the use of the university computing center by an academic library). For this reason, the Panel visited a company that is developing nationwide computing service on a commercial basis. Typical of several such endeavors, the equipment capability is adequate today, and soon will be expanded to provide some classes of data base service with remote access from all regions of the country. At this time, however, there seem to be few customers from the library and information community for data base service facilities from the computer utility industry. Various reasons were cited for this, such as high cost of preparing the data base, high cost of storing very large files on-line, variation in customer requirements, and especially problems of design in the basic functions of the potential customer well above the level of mechanization.

In contrast to other application areas, where computing and data-processing application programs can be developed by the computer utility and offered to many customers, only a very few library support processes are offered by software developers. While those that do exist may operate quite well, they have few active customers as compared to simpler applications such as scientific computation and administrative or business records.

Human knowledge has been recorded for centuries in the form of readable hard copy. Libraries have continually evolved around this technology. The use of computers in business, government, and educational institutions has presented the library function with a new challenge: information recorded in machine-readable form that is neither organized nor formatted as a publication would be.

A well-known example is the large set of machine-readable tape files from the 1970 U.S. census. A relatively small amount of data from this collection will be published in hard-copy form. Effective utilization of the bulk of the census information is difficult. First, purchasing the anticipated 2,000 reels of magnetic tape requires an investment of approximately $120,000. Second, the information is not arranged on the tapes for effective use. A considerable investment must be made to compact the data in a more efficient format. The third problem is the cost of computer time to run the tapes in response to inquiries. These issues present no basic technical difficulty; however, libraries generally are not organized or funded to effectively resolve this type of problem. If libraries do not respond to this general challenge, which the census information illustrates, other institutions will have to.

If libraries are to serve as part of the national information system, they must expand their interests and capabilities beyond those observed by the Panel to handle and to make available information from such digital media. Also, the handling of various audiovisual material by present methods can only be regarded as exploratory.

Although there are projects dealing with these problems, serious consideration of the nontraditional media was mentioned only occasionally in the Panel visits; most institutions have more problems involving the conventional medium than they can handle.

For information-access services, there has so far been no associated source of revenue, such as that provided by advertising in other information-transfer media, e.g., television entertainment, news, and primary serial publications. The tradition of public and institutional support of libraries has removed them from the mechanisms of a marketplace. Because, in addition to technological development, they must solve this problem of revenue acquisition in innovative ways, a double burden is imposed. So far, no significant progress has been made in innovation in revenue acquisition. Information centers offering computer-based search services are experimenting with charging algorithms, but this is a quite limited and specialized activity at present.

Matter to be represented in automated files includes bibliographic information, subject classification, and free text. They are complex and difficult to harmonize because of the large number of independent institutions involved. The organization and ordering rules necessary to sort the information for human lookup, so that related items are brought together, are much more complicated than simple alphabetic ordering.

There are enormous variances in the usage of any given items stored in a library or furnished by an information system. There is an age-old argument over the storage of little or never used material. Certainly anything done to segregate the active material from the inactive can help to lessen the amount of information that must be managed for maximum accessibility and responsiveness. Some major libraries now have off-site, high-density storage for this purpose, but it is never very popular with the library patrons to have part of a collection of interest to them removed from immediate availability. However, it is unlikely with any technology that the national memory will be a single-level storage.

Similarly, there are wide variations in the needs and demands of information seekers. Some persons never consult already stored and organized information. Others make almost constant use of it. If society is to be served, any solution to the library problem must take this into account and use methods of operation and financing that do not deter the frequent user and do entice the present non-user to learn and begin to get value from stored information.

The value to the nation and to society of providing access to information is not measurable in dollars at the present state of knowledge.[27] Consequently, it is not possible to include quantitative cost/benefit studies as part of this report, even though relative cost for equivalent service by different means is of high interest and importance, and although an affordable cost level is vital. In the absence of acceptable measures of value and until such measures are developed, national policy must continue to be made on the basis of qualitative judgment.

One of the main overall impressions gained by the Panel in the site visits and discussions was that "the national information problem" has not been perceived as a whole by most of those involved in trying to solve it. It is being worked on in a fragmented manner. To some, the main problem is access to documents; others emphasize the importance of fact or data access. Some will accept only real-time, on-line, interactive processes; others doubt that most people would work that way, even if they could afford it.

Some wish simply to obtain operating efficiency in the traditional framework; others press for totally new structuring of the problem. The extent to which automatic processes can penetrate into what is normally regarded as intellectual activity in information analysis is debated vigorously.

Current operating environments are used to judge research that should be viewed in a future context, and extrapolations from research environments are being made without full consideration of operational requirements and the transitional problem of getting from the present to the predicted possible future. "The problem" is variously defined as an intellectual crisis, an economic crisis, a technological deficiency, and an institutional-organizational failure.

Whether there is insufficient leadership, or too much divergence among leaders, the net effect is the same, leading to the fragmented appearance of activity as viewed at the national level.

[27] H. A. Olsen, *The Economics of Information: Bibliography and Commentary on the Literature*, Washington, D.C.: ERIC Clearinghouse on Library and Information Sciences, ASIS. January 1971.

Certainly the Panel does not claim the arrogance described so well by Churchman,[28] of being able to shed totally the distortion of view caused by the particular experiences of the individual Panel members. However, the various perceptions and the reports of activities relating to the efforts to solve the "national information problem," or some piece of it, were brought together into a mosaic. Insight was gained from the integration of these ideas and results of those activities to provide a basis for the findings, observations, and recommendations of this report.

[28] C. West Churchman, *The Systems Approach*, New York: Dell Publications, 1968, p. 28.

ANNEX C

INFORMATION TECHNOLOGY: KEY CHARACTERISTICS AND DEVELOPMENT TRENDS

FUNCTIONAL AREAS OF TECHNOLOGY APPLICATION

The specific system functions that technology must support are the following:

- Means for storage of information, appropriately organized.
- Mechanisms for input of the information for machine handling.
- Means for information transfer from producer to organizer, to and among storehouses and to the consumer; also related methods for maintenance and management of the storehouses.
- Means for intellectual and physical access.
- Mechanisms for output from machine form for human use.
- Mechanisms for manipulation of information and for control associated with all the above functions.

GENERAL TRENDS

The technology for use in all facets of information handling has been improving very rapidly over the past two decades, and that pace of improvement continues today. For the most part, this technology is based on electronics and photography and the many levels of equipment and system design necessary to employ them in useful systems. The trend of technological development shows every sign of being able to produce the necessary "nuts and bolts" from which to put together large-scale computer-based information systems of national extent.

Extensive coverage of the many facets of technology for information handling has been included in *National Document-Handling Systems for Science and Technology*[29] and in *Libraries at Large.*[30] In this report, emphasis is placed on important trends, key technological characteristics, and system-design principles. In all areas of this technology, the trend of costs per unit of capability has been a rapid decrease and this trend is continuing. For example, successive generations of computers have consistently been more "powerful" (storage capacity, speed, computational and logical capability) and less expensive in current dollars. For two decades, continuous improvement in the cost/performance ratio of electronic computational and data-processing systems has been sufficiently uniform that statistical extrapolation for 2, 3, and 4 years has yielded quite good forecasts. Two major influences have contributed to the improvement: competition among suppliers of such systems, and research and development programs assiduously pursued by government agencies and by private industry.

A note of caution about this successful and optimistic situation must be injected. During 1970, there was some curtailment of research and development programs both by the government and by private industry, particularly of longer-term programs whose benefits could be expected in the marketplace in 5 years. Unless there is sufficient resumption of research and development in 1971, some temporary effects of that curtailment will be difficult to avoid.

This note of caution must not discourage or prevent immediate operational system design and implementation of computer-based information-handling systems. More "nuts and bolts" than are used effectively are already on hand, and fundamental limits of the information technology are far beyond what is actually employed today in most information handling. Continuing vigorous development can ensure that the necessary capabilities will be achieved.

While the above comments have been made in the specific con-

[29] L. Carter, G. Cantley, J. T. Power, L. Schultz, R. H. Seiden, E. Wallace, R. Watson, and R. E. Wyllys. *National Document-Handling Systems for Science and Technology,* New York: John Wiley and Sons, Inc., 1967.

[30] Douglas M. Knight and E. Shipley Nourse, ed., *Libraries at Large,* New York and London: R. R. Bowker Company, 1969.

text of the computer industry, the same comments could be made with reference to other industries such as communications or reprography. The technological bases of all these industries are intertwined. All lead back to the basic sciences underlying electronics, photography, and associated mechanical devices, so that it is neither possible nor useful to try to divide the system elements into corresponding classes. The term "information technology" used in this report is intended to convey this broad characterization.

CONTINUITY CHARACTERISTICS AND PROBLEMS

In the past, computer-based information-handling or management systems have been designed on the basis of specific hardware, specific languages appropriate for that hardware, and specific operating systems. In many cases, particularly that of libraries, these "application systems" (as viewed by the computer field) have been operated on the central computing facilities of the institution of which the library is a part. As changes in hardware and operating systems have occurred rapidly and continuously, the foundations for the information activities have been unstable, causing continual redesign and operating problems. The basic reasons for this involve the continuity factors relating to the specific environment of library and information activities.

The continuity of scientific calculations rests in the algorithms in the programs to do those calculations. As programming languages have developed, especially FORTRAN, the exportability or movability of such algorithms has become a fairly routine process. Adaptations to changes in technology have been largely successful.

The continuity of data processing depends on both algorithms and the files they operate on. COBOL, for example, provides for the movability of business-calculation algorithms. The files involved, in the large majority of business situations, are relatively small (compared with published information files) and highly formatted and consistent in form (as contrasted with the variability of general information files). Both characteristics make the conversion problem from one system to another—when conversion is necessary—relatively easy. Of course, there are large corporate information systems with the size and complexity and continuity factors of general information systems. As the pervasiveness of the use of computer systems increases in business activities, more large-scale general information systems will develop.

The major continuity of library and information systems is in the files of the system and the portions of the software that interface with them for access to and management of the information, in contrast to the characteristics listed above for scientific and business calculations. The continued development of programming languages is aiding the design of "transferable processing algorithms," but the development of general data base organization and management software is still deficient relative to what can be purchased. The complex interaction between the computer-operating system and data base software compounds the problems. Thus, it is not enough to have ways to bridge between dissimilar hardware, but it is also necessary to be able to bridge between successive versions of the operating system for the same hardware.

The computer industry is far from having compatible operating systems (except where one manufacturer specifically emulates the software of another) and the data base management problem has hardly been touched for commercially available and supported systems. The software packages that are available as part of operating systems are "file" management systems, not "data base" management systems. They are a part of the corresponding operating system, and not yet sufficiently stable from one version of an operating system to another. The customized "data management" systems that have been available commercially up to now have been deficient for handling the very large files of many variable parameters that characterize information files.

FILE STORAGE

Whether the content is bibliographic information or actual information, the files involved are very large[31] as compared with today's

[31]For example, when the Ohio College Library Center expands to full operation in several years, it will have to provide access to six million unique items, located in multiple locations among 50–60 different libraries. At approximately 600 bytes of information on each item, a storage capacity of 3.6×10^9 bytes (2.9×10^9 bits) is needed for the catalog record alone.

As an extreme example, the information content of the Library of Congress is estimated to be in the range of 10^{14}–10^{15} bits, requiring 10^{11}–10^{12} bits for catalog files. (King report estimated 10^{14} and 10^{11}, respectively, in 1962.

"typical" file kept in a computer center. Very long computer runs are required if files are on magnetic tape, or, alternatively, an amount of direct-access storage is required that has been outside the bounds of economic feasibility. However, these technological barriers are falling.

Direct-access memory is the hardware component of utmost importance in an electronic information system. It now represents the major portion of the hardware cost for the central installation, and it's size determines how much information the system can make immediately available. It is necessary because access to information stored on less expensive magnetic tape is cripplingly slow.

In commercial data processing, there is a requirement for large volumes of machine-accessible storage, as presently typified by the rotating magnetic disk. A typical medium-sized data-processing system may have a total of 10^8 bits of direct-access storage, costing perhaps a tenth of a cent per bit. The very largest systems may have one hundred times as many bits (10^{10}), with the cost per bit down by a factor approaching 10. Based on these requirements, the competition among vendors, and known but unpublished research now in progress, a reduction of a factor of 2 or 3 in the cost to the user per bit of such storage can be expected within the next very few years. A forecast that the improvement factor will reach 10 within the decade is reasonable.

In addition to Panel views on this subject, an independent source of organized opinion is the Bernstein Report.[32] In a year-long effort, Bernstein and others of the Research and Development Division of the U.S. Naval Supply Command gathered and distilled expert opinion as to the future of information-processing technology. While the resulting forecasts are opinions and, as such, their accuracy is open to debate, the thoroughness with which the technological community was canvassed leaves no question that the opinions were the best available at the time. Unfortunately, these are the opinions held by experts in 1968 rather that those held today. There is no comparable record of opinion in the public domain that is more up-to-date.

On the topic of large-scale memory, the opinion of the tech-

[32]George B. Bernstein, *A Fifteen-Year Forecast of Information-Processing Technology,* Washington, D.C.: Naval Supply Systems Command, January 20, 1969.

nological community, as expressed by the Bernstein Report, is optimistic. Two examples from the report illustrate the expectations:

1. At a cost much less than 0.001 cent per bit, mass memory (direct-access storage) of 10^{11} to 10^{12} bits, average access time under 1 second.
   Earliest time: 1972
   Highest expectancy time: 1978
   Confidence level: 8, on a 1–9 scale
2. At a cost of less than 0.0001 cent per bit, mass memory of 10^{13} bits, access time under 1 second.
   Earliest time: 1974
   Highest expectancy time: 1977
   Latest expected time: 1980
   Confidence level: 5

During the April 1971 Intermag Conference in Denver,[33] some of the research and development expectations for future memories were discussed.

Examples of equipment that has been announced by the computer industry include:

1. Disk technology[34]
   Demountable storage units of 8×10^8 bits
   8-unit installation 64×10^8 bits
   Access time: 40 msec average (effective access time is reducible by request queuing in the controller)
   Cost (capital hardware only): less than 0.01 cent per bit
2. Large-scale storage[35]
   Demountable storage units of 5.0×10^{10}
   Large installation 2.8×10^{12} bits
   Access time: several seconds
   Cost (capital hardware only): approximately 0.0001 cent per bit

[33] IEEE, *IEEE Transactions on Magnetics,* New York, September 1971.
[34] IBM 3330 or equivalent.
[35] *Computerworld*, Vol. V, No. 16. 21 April 1971.

Storage of the order of magnitude of 10^{12} bits has existed for several years.[36]

Private reports indicate that new large-scale information-storage media are being widely worked on and will enter the systems marketplace in the next few years. Certainly the basic physical properties of materials provide for very much higher densities than now employed, whether magnetic phenomena, normal photographic processes, or laser-actuated storage mechanisms are considered. Unfortunately, publicly citable support for reliable prediction of such achievements is not yet available. Those working on these developments expect competition and are not providing full information for public consumption.

While firm commitments should not be made based only on rumors, it is clear that there is intense activity. Library and information-system designers should be alert to employ these developments when they are available. Within the next several years, it will be possible for memory systems of total capacity of 10^{13} bits to be assembled for use. Possibilities for development beyond that level of capacity exist.

INPUT TO MACHINE FORM

Even when the basic capability to store information at acceptable costs exists and suitable access mechanisms are available, the input technology must also be considered.

Until arrangements can be made for "source data capture," the material that enters digital form must be keyboarded specifically for computer-based information system use. The associated problems are well known and are being solved in the entry of information in the library community and in the abstracting and indexing services. A multiplicity of types of equipment for this purpose is becoming available, with the present state-of-the-art represented by keyboard to magnetic-tape recorders and clusters of keyboards supported by an on-line computer system. In some cases, typing documents with special fonts and converting to computer form

[36] Session on Bulk Memory Devices, AFIPS Conference Proceedings, Vol. 33, FJCC 68, pp. 1361–1397. Large-scale memories based on photographic, laser recording, and magnetic methods were discussed.

via optical character-reading equipment is used. In all cases, however, labor costs are high and error control is difficult.

There are two extremely important principles to be observed in the input operations: (a) The storage form of the information must not be tied to a specific keyboarding technology; rather, appropriate conventions must be used to allow available input mechanisms to tell the system what to store in the data base. (b) The input must not create files formatted and stylized for a single specific use, but should be oriented toward the maximum detail available so that all output forms of the information can be derived by mechanized processes.

These principles are gradually being learned as attention shifts from computerized publication systems to computer-based information systems.

Optical character recognition of general typography offers a possible alternative input mechanism from original documents although it is less attractive than generating machine-readable forms at the source. Present OCR devices have uncomfortably high mis-recognition rates when multiple-font recognition is required. While it is reasonable to predict improvement, it is not reasonable to predict perfection. Human involvement in page-by-page error-correcting will apparently remain necessary for a very long time. Additionally, there is the practical problem of presenting a page of a bound book to an optical scanner in such a way that the entire page is in sharp focus. However, that problem is the same for a camera as it is for a scanner, and we are led to the thought of the OCR device operating on a photographic image of the page rather than on the page itself. In particular, if the pages of a volume have been photographed for distribution or reproduction, some simple constraints would make the product serve as an "intermediate" storage medium and be suitable for optical character recognition. The constraints are that the photographic images are embedded in a framework that facilitates individual page images being located and moved by machine, and that the resolution is appreciably higher than the minimum requirement for human legibility.

The long-term hopes for true "source data capture" for general library and information systems rests on the evolution of the printing industry. The present dependence on paper as the primary medium of stored information is obviously a direct result of the form in which most information is now (and historically has been) produced. Conventional printing of all types and typewriter-originated documents depend on a paper technology. Until machine-readable information becomes routinely available as a part of the publishing process, whether as primary or by-product output, the preparation of information for use in computer-based systems will remain a serious economic problem.

Computer-controlled typesetting has been the scene of intense activity for many years. By now, there is no doubt that, except for high-quality photography, there is no aspect of publishing that cannot be computerized.[37] Even "picture processing" is the subject of intense research for possible future impact on information systems. The big question is not whether the technology is adequate functionally, but whether the process can be done economically. While the prevailing feeling of the publishers, as expressed at the ASIS Workshop, was one of caution and some frustration that the expectations had not yet really paid off in general, no one was ready to give up and all were convinced that those who learn to master the technology would profit from it. Thus, it is reasonable to expect that, as the publishers master the computerization of their environment, the machine-readable content of publications of all types will ultimately become available, solving a large part of the "input problem" as a by-product. The computer files available from the publishers will require further machine processing to prepare the content for use in computer-based information systems.

IMAGE TECHNOLOGY

Until full digital storage of documents is practical and input in that form is available from publishers, microforms will play an increasingly important role in storage and transfer of information. A particularly powerful combination at the present state of technology is full-text documents available on microform storage media, coupled with access routes through computed-based search sys-

[37]Robert M. Landau, ed., *Proceedings of the ASIS Workshop on Computer Composition,* Washington, D.C.: ASIS, 1971.

tems. The microforms may be the result of photographing hard copy or may be produced directly by computer-controlled microform recording equipment.

Additional purposes can be served by photographic images of pages of text. It may be desired to preserve the information content of a page beyond the lifetime of the paper on which the page was printed. For this purpose, the medium of which the image is made must be of very long-term stability. The serious problems facing libraries in long-term preservation of printed matter (on paper) were discussed by a committee of the Association of Research Libraries.[38]

At some time electrical transmission of a copy of a page to a remote point may be desired. Presenting the scanner with a photographic image is much more convenient than presenting the page itself. Since any transmission method degrades quality, the resolution of the image must be well above the threshold of human legibility if the quality of the received image is to remain above the threshold.

Finally, it may be desired to have additional copies of a page or pages at the site where the originals are, or at one or more other places. If there is a copying device that can both utilize and preserve in the copy the features needed by the other uses of the image, then generating useful copies from microform masters can be a very low-cost operation. For material of which multiple copies are desired, the cost of making the initial photographic image is then prorated among the copies.

The various required and desired properties of microform images and their embodiment include:

- The original image resolution must be high enough so that the degradation encountered in making a photographic copy of the original image, followed by using the copy as source for electrical transmission of the image, does not result in an illegible final image.

[38] Association of Research Libraries, "The Preservation of Deteriorating Books: An Examination of the Problems with Recommendations for a Solution," Minutes of the 65th Meeting, 24 Jan. 1965, p. 9–42.

- The medium that bears the image must be of very long life.
- The image must be as small as possible, consistent with the resolution requirements, the desired number of images per unit of medium, the mechanics of handling, and the optics of enlargement to viewable form.
- Images must be imbedded in a medium in such a way as to facilitate locating and moving them by machine.
- There must be a machine that can locate and move images.
- There must be a machine that will make copies retaining all the above properties except for some small degradation of image incurred by copying.

The key problem area in this segment of information technology remains the lack of standardization of image polarity, reduction rations, sizes of basic medium, and shape of image and ratios medium.[39] A particularly important capability is the computer-controllable addressing of individual images on the microform, now appearing in some new retrieval equipments. Mechanized handling of microform media enabling transfer of images from one package to another is an essential capability in the overall publishing–library–access and delivery system. For a specific application, a capability of this type is illustrated in the Composing Reduction Printer employed by the U.S. Patent Office to produce multiple-image aperture cards from 35-mm roll microfilm.[40]

Copies of microfiche can be produced for several cents per copy and of roll microfilm at a few dollars per 100-foot roll, if the installation operates at a volume high enough to spread the original cost of equipment over many copies. Clearly, microcopying is an inexpensive way to reproduce and deliver images of published information. Compensation of the originators of the publication being reproduced and distributed must be added to the physical-transfer costs. However, these fees would apply independent of the form of transfer and are affected by legal principles and contractual arrangements.

[39]Arthur Teplitz, "Microfilm and Reprography," Vol. 5, p. 87–112. *In* Carlos A. Cuadra, ed., Annual Review of Information Science and Technology, Chicago: Encyclopaedia Brittanica Inc., 1970.
[40]U.S. Patent Office, private communication.

An indication of the economics of microform distribution as compared to paper forms is in the price schedules from National Technical Information Service, Department of Commerce, which is charged with recovering reproduction and distribution costs. Typically a report that costs $3.00–$6.00 in paper copy can be obtained in microfiche form for 95 cents.

The delivery of images from a central store to a using site is also of concern. Two methods are available. One is a local collection viewable on a local microform reader. A great many such equipments are becoming available. Particularly noticeable at the AFIPS Conference (FJCC and SJCC) in 1969 and 1970 was the increasing display of microform retrieval and viewing equipment that rivaled the electronic computing, component, and instrumentation exhibits in number and variety. In the Spring of 1971, the comprehensive exhibiting of this equipment shifted to the meeting of the National Microfilm Association. For small collections the local microform collection is an eminently practical approach, whether the bibliographic and indexing access to the collection is handled via a local computer system or a remote one.

Several large-scale systems exist for storage and retrieval of the order of 10^7 pages of images on microforms. Magnetic-tape video recording of such images has also been used.[41]

The distribution of images in these systems is by means of high-resolution TV technology. Industrial television equipment with resolution of the order of 900–1,250 lines per frame is the routine state-of-the-art. However, it appears that, for the practical viewing of typical published full pages, a resolution of 2,000 lines per page is necessary, as is used in the INTREX laboratory. Through various means of "zooming" into partial-page viewing, the more readily available TV technology can be successfully employed.

The use of television for image distribution is practical for local networks served by a single storage facility, i.e., a single building or a college campus where a direct coaxial-cable distribution system is practical. Long-distance distribution, at present, runs into the very high costs faced by TV network transmissions. However, in the future, it seems reasonable for local communities to be

[41]Recent exhibits of AFIPS Conferences and the National Microfilm Association.

served through modern cable-TV networks, particularly if terminal-station image-buffering means are developed.

COMPUTERS

In the computer realm of information technology, at least passing notice should be given to the state of development and trends of the central mechanisms of computers themselves. Spurred on by the demands of scientific calculations and various military applications, basic central processor development has been pushed to a high level of achievement and far more computing power is available than is now necessary in ordinary information-handling systems. Of course, as machine processes are applied more and more to content analysis, automatic organization of information, and picture processing, cheap computational power becomes more important. However, the state of development of central processing units (CPU) is fully adequate for near- and mid-term purposes.

One aspect of CPU design needs some attention. This is the design of addressing mechanisms suitable for accessing very large hierarchical stores of information. The current emphasis on virtual-memory and cache-memory designs for computers actually implements only a "single level" storage hierarchy, even though multiple memory device types may be involved. As the data-management system architecture and software problems are solved, there will probably be requirements for improvements in basic information addressing capability.

One very important trend of development is the increasing power, decreasing cost, and diminishing size of minicomputers. They are very important for application to "intelligent terminals," for communication-network control and interfacing, and for low-cost small local handling of small files of machine-readable information.

The price-performance index for minicomputers has been improving 50 percent per year since they were introduced in 1963. New developments in semiconductor technology for both logic and memory are expected to continue this trend until 1975, reaching a cost of between $1,000 and $2,000 for a 4,000-word minicomputer of quite respectable performance. This would have considerable impact on "intelligent terminals," communication-network controllers, and for decentralizing many mechanized functions.[42,43]

At present, the large data-processing system outperforms the smaller one on a computation-per-dollar basis. It was at one time a rule of thumb that one could buy a factor of 10 in performance for a factor of 2 in cost. Another expression of cost improvement made possible by increasing the speed of operations is Grosch's law, which hypothesizes a square-root relationship. It was therefore advantageous to amalgamate many functions into a single centralized system, even though substantial communications costs may have been incurred by doing so. However, recently much more improvement in performance per dollar has been achieved in small computers than in large ones, and thus the relative advantage of the totally centralized computer system for constant communication costs has been diminished. It will be prudent in the future to give attention to this trend, rather than to assume without question that the totally centralized computer system is less costly than is the combination of a large centralized facility for common functions coupled to small computers for support of the local environment.

Because the cost optimization involves complex interrelationships among the costs of the central facility, communications, and terminal equipment, and the effective system architecture is limited by the software design and file organization (logical and physical), no single factor can be used to predict the cost impact of the improvement of any single piece.

The hardware advances in semiconductor memories will also provide advances in price performance of major processors (at least one is already announced commercially).[44,45] Also, solid-

[44] L. L. Vadasy, H. T. Chua, and A. S. Grove, "Semiconductor Random-access Memories," *IEEE Spectrum*, 8 (No. 5), 40–48, May 1971.
[42] David L. House and Russell A. Henzel, "Semiconductor Memories and Minicomputers," *COMPUTER*, 4 (No. 2), 23–29, March/April 1971.
[43] D. J. Theis and L. C. Hobbs, "The Minicomputers Revisited," *DATAMATION*, 17 (No. 10), 24–34, May 15, 1971.
[45] Joseph K. Wineke and Mitchell Spiegel, "Generation IV: The Shape of Systems to Come," *Computer Decisions*, 2 (No. 10), p. 18–23, Oct. 1970.

state devices can be expected ultimately to replace drums and disks in high-performance installations,[33] although stiff competition is expected from improvements in drum and disk technology.[45]

The general trend of costs of achieving individual electronic functions is sharply downward, but this does not always result in devices of lower cost. The individual-function cost reduction may be expected to continue, but probably with no sudden spectacular improvement of the magnitude represented by the change from discrete to integrated circuits, not yet fully exploited in presently available computers and terminal equipment.

The delegation of certain functions to local sites does not violate the principles of the "economy of scale" when these functions can be accomplished once for the benefit of many users and when the sharing of the more efficient large resource can be accomplished so that the overhead of serving each individual user does not dominate each interaction with the system. The partial decentralization of function does provide a solution for combating the "Parkinson's law of large systems" by avoiding the interruption of the central facility for performance of "trivial" tasks with high system overhead. As a design problem, however, the principles and detailed design guidelines for partitioning the functions between the central and local facility are largely undeveloped.

DATA COMMUNICATIONS

In looking to the future of computer-based library and information systems and network activity, data-communications technology plays a key role. Several classes of communication links are evident: between individual users and a system, from a bulk information-distribution point to a service distribution center, and among nodes in storage and retrieval networks. It is technologically feasible to provide any of these links through electrical communications networks. However, particularly as it affects today's conventional library environments, relatively little penetration of that technology has been achieved.

Mostly, this present lack of usage is due to the very high cost per unit of transmission capacity in today's data-communication environment.[46] Much more electronic capability has been demonstrated than is yet installed in the nationwide communication system. Improvements are being installed at a rapid rate. However, it is a massive undertaking, and time and much capital investment are necessary. Costs are also high due to the specific rate structures as they now exist. The adaptation of the nationwide communication network to data transmission from its originally engineered purpose of voice communication has been slow due both to the lag of installation of new capacity and the regulatory and business arrangements for financing and using it.

The primary requirement for achieving full use of capacity, at any level of technological development, to cut the unit cost of use as far as possible, is to get the facility fully loaded. Quick and easy means are necessary for switching from one customer to another. This is how the economics of scale have been achieved in the telephone system and in large multiprogrammed computer installations. The switching patterns and trunking patterns for data communications are different from those for voice communications, and costs will be reducible as the communication network can be adapted to these new patterns. To illustrate, connection set-up time for a conventional long-distance call may average 10–20 seconds, and a 1-minute segment of time is the shortest connection charging period that has been considered. One network being specifically engineered for data transmission expects a connection set-up time under 3 seconds and expects to charge on the basis of 6-second periods.[47]

The technological basis for extreme improvements in both performance and cost exists in new digital carrier systems (e.g., Bell System T1, T2, and successor developments),[48] circular wave guide for millimeter wave carrier systems,[49] and various broadcast microwave and earth-satellite relay systems for obtaining greatly increased transmission capacities. The associated improvements in

[46]For "voice grade" service, which is inadequate for bulk transmission but fully adequate for user terminal interaction in simple bibliographic access systems, a very complete estimating guide is given in Karl I. Nordling, "Analysis of Common Carrier Tariff Rates," *DATAMATION*, 17 (No. 9), 28–35, May 1, 1971.
[47]Paul Hersch, "Data Communication," *IEEE Spectrum*, 8 (No. 2), 47–60, February 1971.
[48]Hersch, *Loc cit.*
[49]*Time*, 12 April 1971, p. 58 (advertisement).

switching mechanisms, whether circuit-switching or message-switching, can be achieved by applying computer technology itself, as is being done in the electronic switching systems being installed.

The potential of cable television (CATV) for local distribution of information of all types is very high. In 1969, there were over 2,000 CATV stations serving approximately 4 million homes. Estimates of growth by 1980 for total number of connections range from 6 million to 17 million, depending on the legal and regulatory adjustments that are now the scene of vigorous struggles.[50] The technical capabilities of new CATV facilities provide two-way communication over 20 to 40 channels, which is an enormous amount of information capacity if effectively shared.[51]

Other recent references treating the subject of data communications (each containing references to much source material) are noted below.[52,53,54]

Until performance goes up and costs come down greatly, the bulk transfer of data among publishers, libraries, and other information activities will continue to be by shipping magnetic tape rather than by electrical communications. However, if that tape is viewed as a transmission medium rather than as a file medium internal to a computing system, and if system design is structured accordingly, then a switch can be made to electrical communications under such conditions and at such time that the costs and values balance. Therefore, network design for movement of data need not wait, and should not wait, until the necessary improvements in data transmission have been achieved. In fact, the existence of data to be transferred and the eventual existence of the facilities to transfer it are tightly interrelated in another example of the threshold effect of "getting started," i.e., the problems of getting a market to the level at which it can sustain itself.

[50]Archer S. Taylor, "The Future of Cable TV," *IEE Spectrum,* 6 (No. 11), 77–81, November 1969.
[51]Hersch, *Loc. cit.*
[52]Robert L. Simms, Jr., and Edward Fuchs, "Communications Technology," *in* Carlos A. Cuadra, ed., *Annual Review of Information Science and Technology,* Chicago: Encyclopaedia Britannica Inc., p. 113–140, 1970.
[53]Data Communications Survey–Our Readers Respond," *Computer Decisions,* February 1971.
[54]Data Communications: Part I–Floodtide at the FCC," *Management Science Publishing, Inc.* (A Subsidiary of *Automatic Data Processing*), Volume XIV, Number 24, November 23, 1970.

The kind of concurrent performance improvement and cost reduction that is possible is illustrated by some estimates made in connection with the ARPA experimental computer network.[55] Estimates were made on various means of transporting a million bits of information for a 1,400-mile distance without regard to delay time. Only communications costs are included and an effective 8 hours of use per 24-hour day is assumed in apportioning the relevant equipment costs. The comparison table (Table 4) is taken from the Spring Joint Computer Conference Proceedings.

In interpreting Table 4, several items should be noted. Economy in the use of telecommunications requires, and the 8-hour usage assumption implies, high utilization. At this utilization level, another factor of 7 in cost reduction (from the level attained with the 50 KB service) would break even with airmailing a full tape reel. Also, in some situations the value of the shorter delay of electrical communication, which allows performance gains or cost savings in other parts of the system, can justify a communication

TABLE 4 Cost per Megabit for Various Communication Media 1,400-Mile Distance

Media	Cost ($)	Conditions
Telegram	3,300.00	For 100 words at 30 bits/word, daytime
Night letter	565.00	For 100 words at 30 bits/word, overnight delivery
Computer console	374.00	18 baud average use, 300 baud DDD service line and data sets only
TELEX	204.00	50 baud teletype service
DDD (103A)	22.50	300 baud data sets, DDD daytime services
AUTODIN	8.20	2,400 baud message service, full use during working hours
DDD (202)	3.45	2,000 baud data sets
Letter	3.30	Airmail, 4 pages, 250 words/page, 30 bits/word
W.U. broadband	2.03	2,400 baud service, full duplex
WATS	1.54	2,000 baud, used 8 hr/working day
Leased line (201)	0.57	2,000 baud, commercial, full duplex
Data 50	0.47	50 KB dial service, utilized full duplex
Leased line (303)	0.23	50 KB, commercial, full duplex
Mail DEC tape	0.20	2.5 megabit tape, airmail
Mail IBM tape	0.034	100 megabit tape, airmail

[55] Lawrence G. Roberts and Barry D. Weisler, "Computer Network Development to Achieve Resource Sharing," *AFIPS Conference Proceedings,* 36 (SJCC 70), 543–549, Montvale, N.J., 1970.

cost higher than that encountered in the delayed-delivery method. Such a value is heavily system-dependent.

The 50 KB communication service used by ARPA is only an early form of high-data-rate communications facility, and the gains in performance/cost ratio by the newer facilities now being installed or planned (by both the Bell System and independents) will make drastic changes in the cost balance between electrical and nonelectrical communications.

OUTPUT FOR HUMAN USE

In the man/machine relationship, the specific capability of terminals to handle the full range of information required is of considerable importance. The computer and related technological requirements of library and information systems have exceeded those of scientific and business applications. The character set needed, for example, is many times larger than the 48-to-64-character sets that heretofore have been fully adequate for scientific calculations and business records.

Over many years, the printing industry has developed type fonts and page formats that are effective in presenting printed information in compact form for rapid and accurate human consumption. While there can be considerable argument over how much of the quality and stylistic variation is in fact necessary, there is no doubt that what is needed has far exceeded conventional computer output capabilities. Library and information systems must handle multiple alphabets of non-English language material and scientific notations. Even with elimination of stylistic variations of typography and tight restrictions on symbols from foreign alphabets, the minimum library character set for bibliographic information alone, as described by the Library of Congress, includes 176 different characters—significantly in excess of the 95 graphics (94 printable plus space) of the standard ANSI basic set.

While there have been many limitations in the past, it is now technically feasible to build equipments that can handle such a range of material. The combination of cathode-ray-tube technology, inexpensive "nonimpact" hard-copy devices, and the minicomputers of the near future provide a fully adequate technical basis for that purpose. If agreement on specific characteristics for user terminals could be reached among potential users, the market should be large enough to bring the cost down to an "affordable" level.

As noted earlier in connection with publication activities, computer-controlled photocomposition provides the technical capability to produce hard copy of full printing quality. Conventional hard copy, as an output, will be necessary for human use for a very long time, even if the full extent of forecasted computerization and use of electrical communications is achieved.

THE BASIC ECONOMIC PROBLEM OF "GETTING STARTED"

As we may observe in all facets of the technology, basic technical feasibility of equipment function is not the issue in most cases now, and continuing vigorous research and development in the technology can provide solutions to remaining problems. However, to establish economic feasibility in some areas, unit costs must be brought down. Market volume is the key to this. If the costs were down, there would be more usage and more demand thereby expanding the market. Clearly, the main problem in these areas is how to achieve a sufficient level of activity to create a self-sustaining market.