AFOSR 68-1403 Vol I

23

AD 670606

SC

JUN 27 1968

Science Communication, Inc.

1050 31st St., N.W. Washington, D. C. 20007

266

73

AD 670606

STUDY OF SCIENTIFIC AND TECHNICAL DATA ACTIVITIES IN THE UNITED STATES --

VOLUME I

PLAN FOR STUDY AND IMPLEMENTATION OF NATIONAL DATA SYSTEM CONCEPTS

Prepared for

Task Group on National Systems
Committee on Scientific and Technical Information
Federal Council for Science and Technology

DEC 1968

Final Report
Contract F44620-6?-C-0022
ARPA Order: 892 as Amended

SCIENCE COMMUNICATION, INC.
1050 31st Street, N.W.
Washington, D.C. 20007

Science Communication
Washington, D. C. 200 07
COSATI Data Systems Study
Final Report - F44620-67-C-0022

SPONSORSHIP ACKNOWLEDGEMENTS

The work reported herein was sponsored by the Task Group on National System(s), Committee on Scientific and Technical Information (COSATI) of the Federal Council for Science and Technology. Financial support was provided by the Advanced Research Projects Agency of the Department of Defense, with contract administration by the Air Force Office of Scientific Research. Work was initiated, under Air Force Contract F44620-67-C-0022, on 1 September 1966 and completed on 30 April 1968.

FEDERAL COUNCIL FOR SCIENCE AND TECHNOLOGY DEC 1968
COMMITTEE ON SCIENTIFIC AND TECHNICAL INFORMATION
EXECUTIVE OFFICE BUILDING
WASHINGTON, D.C. 20506

FOREWORD

The Task Group on National Systems for Scientific and Technical Information of the Committee on Scientific and Technical Information (COSATI) is sponsoring a series of studies on aspects of information systems and activities in the United States. This report by Science Communication, Inc., is the result of one such study.

COSATI feels that this report contains much valuable information and many thought-provoking recommendations. Both government and private communities should benefit by having the report widely distributed, and extensively reviewed and discussed. Hopefully professional societies, private groups and interested individuals will continue the analysis of scientific and technical data activities which has been well begun in this report.

Andrew A. Aines
Chairman

Science Communication
Washington, D. C. 200 07
COSATI Data Systems Study
Final Report - F44620-67-C-0022 30 April 1968

ABSTRACT

This volume presents a plan for study and implementation of national scientific and technical data system(s) concepts. The plan reported was developed as a part of a broader planning effort by the Task Group on National System(s) of the Committee on Scientific and Technical Information (COSATI). COSATI is a committee of the Federal Council for Science and Technology.

Major objectives of the plan are: (1) management of scientific and technical data resources in a manner optimal for maintenance of a strong science and technology, (2) improvement of existing data management programs and data handling services by better use of available technologies and methodologies, (3) development of the personnel, institutional, and methodological capabilities required to support future data-management and data-handling systems, and (4) identification of procedures and designation of responsibilities for actions to facilitate the development of new systems of data management and data handling.

The plan envisions the achievement of those objectives within a National Program for Scientific and Technical Data. Significant elements of the National Program include organization of a National Advisory Council for Scientific and Technical Data and establishment of two Program Offices - one for scientific data activities and one for technical data activities.

The plan presented in this volume is based in part on an extensive survey-study of data activities as currently conducted in government, industry, and the professions. The results of this background study are reported in Volume II of this report.

Science Communication
Washington, D. C. 200 07

COSATI Data Systems Study
Final Report - F44620-67-C-0022 30 April 1968

STUDY OF SCIENTIFIC AND TECHNICAL DATA ACTIVITIES IN THE UNITED STATES

TABLE OF CONTENTS

VOLUME I - PLAN FOR STUDY AND IMPLEMENTATION OF NATIONAL DATA SYSTEMS

Science Communication
Washington, D. C. 200 07
COSATI Data Systems Study
Final Report - F44620-67-C-0022 30 April 1968

VOLUME II - PRELIMINARY CENSUS OF SCIENTIFIC AND TECHNICAL DATA ACTIVITIES

(Bound Separately from Volume I)

PART C
(Bound Separately from Parts A & B)

PRELIMINARY CENSUS OF
SCIENTIFIC AND TECHNICAL DATA EFFORTS

Science Communication
Washington, D. C. 200 07

COSATI Data Systems Study
Final Report - F44620-67-C-0022 30 April 1968

ACCESSIBILITY OF DOCUMENTS CITED IN THIS REPORT

Many of the background documents for this study are reports of Government sponsored studies. Most of these documents are available from the Clearinghouse for Federal Scientific and Technical Information ("CFSTI"), Springfield, Va., 22151. In ordering Clearinghouse documents, use of the "PB" or "AD" numbers is suggested to expedite the processing. The other principal source of government-sponsored documents is the Superintendent of Documents, Government Printing Office ("GPO"), Washington, D.C. 20402.

It is the policy of the President's Science Advisory Committee and the Federal Council for Science and Technology, Committee on Scientific and Technical Information, to make their reports and reports sponsored by them readily available to the public. To assist the reader, therefore, the following information supplements the bibliographic references to such reports as they appear in this report:

1. Progress of the United States Government in Scientific and Technical Communications, Committee on Scientific and Technical Information of the Federal Council for Science and Technology, Executive Office of the President, 1965, PB 173 510. Available from CFSTI.

2. Recommendations for National Document Handling Systems in Science and Technology: Appendix A -- A Background Study -- Volumes I and II, System Development Corporation, Santa Monica, California, September 1965, AD 624 560, PB 168 267. Available from CFSTI.

3. A System Study of Abstracting and Indexing in the United States, System Development Corporation, Falls Church, Virginia, 16 December 1966, PB 174 249. Available from CFSTI.

4. Exploration of Oral/Informal Technical Communications Behavior, Semi-Annual Technical Report, American Institutes for Research, Silver Spring, Maryland, 15 March 1967, AD 650 219. Available from CFSTI.

5. Handling of Toxicological Information, A Report of the President's Science Advisory Committee, The White House, Washington, D.C., June 1966. Available from GPO.

6. Science, Government, and Information, A Report of the President's Science Advisory Committee, The White House, January 10, 1963, GPO (out of print).

7. Report of the Office of Science and Technology Ad Hoc Panel on Scientific and Technical Communications, J.C.R. Licklider, et al., 8 February 1965. Unpublished (out of print).

8. The Copyright Law as it Relates to National Information Systems and National Information Systems and National Programs -- a Study by the COSATI Ad Hoc Task Group on Legal Aspects Involved in National Information Systems, Washington, D.C., July 1967, PB 175 618. Available from CFSTI.

9. Progress of the United States Government in Scientific and Technical Information, Committee on Scientific and Technical Information (COSATI) of the Federal Council for Science and Technology, Washington, D.C., 1966, PB 176 535. Available from CFSTI.

10. Review of Proposal for a National Data Center, Statistical Evaluation Report No. 6, Edgar S. Dunn, Jr., Office of Statistical Standards, Bureau of the Budget, December 1965. Available from Bureau of the Budget, Executive Office Building, Washington, D.C. 20506.

11. President's Message on Communications Policy to the Congress of the United States. The White House, Washington, D.C., August 14, 1967. Available White House Press Office, Washington, D.C., 20506.

12. Scientific and Technological Communication in the Government, (The Crawford Report), Task Force Report to the President's Special Assistant for Science and Technology, Washington, D.C., April 1962, AD 299 545. Available from CFSTI.

13. Information Sciences Technology: First Report of Panel 2, Committee on Scientific and Technical Information of the Federal Council for Science and Technology, September 1965, PB 169 686. Available from CFSTI.

14. Presidential Message upon signing of the State Technical Services Act, P.L. 89-182, President Lyndon B. Johnson, September 14, 1965. Available from White House Press Office, Washington, D.C., 20506.

Science Communication
Washington, D. C. 200 07
COSATI Data Systems Study
Final Report - F44620-67-C-0022 30 April 1968

FOREWORD*

The electronic age is once more the age of the hunter, only now it is the hunt for information, for data.[1]

Machines can measure and produce more data and men can discover more variables and objects to measure than either man or machine can cope with.[2]

Of the three principal categories of information involved in scientific and technical information--data, procedures and methods, and conceptual framework, theories and ideas--only the first, data, are presently readily responsive to available machine storage and retrieval.[3]

The stand taken here is to suggest an alternative goal for information retrieval systems which deserves greater priority than the dispensing of information. This alternative is to assimilate and weld newly generated knowledge into a coherent overall image. ... Whereas the key word of most enterprises and projects in information retrieval is access, the keywords proposed here as an alternative are evaluation and synthesis.[4]

Since only a small fraction of the effort expended in collecting data is typically devoted to its analysis, a large amount of the informa' on it contains generally is undiscovered and unexploited.[5]

Unless destroyed or lost, data can be reused an infinite number of times. The flow of technical data is not only from time to time but from organization to organization.[6]

---

*A listing of sources for the quotations in this Foreward is at the end of this volume.

Information is an agency resource, a federal, national, and international resource. [7]

The resource can be maximized only if it can be made more than a jumble of fragments. Information must be assembled for the good of the Republic. [8]

Modern science and technology cost our society dearly, and our society is justified in demanding its money's worth. [9]

. . . we must manage our technical data resources with the same care that we manage our materiel, manpower, and financial resources. [10]

The U. S. Government is the free world's foremost recipient and generator of all types of information. Sound handling of information is the heart of fulfillment of virtually every major Federal responsibility. [11]

In a large sense, . . . industrial organizations exist as much for the purpose of processing information as for producing military hardware or consumer goods. [12]

The massive store of new scientific and technical knowledge. . . represents a national economic resource possibly equal in importance to our classical resources of land, minerals, manpower and capital. But to exploit this resource--this bank of intellectual capital--we must apply it to the needs of industry and society. [13]

The basic theme is that information is a resource to be managed. Its generation consumes time and money. Its use, proper use, conserves time and money. [14]

Think of the pile of paper represented by over 50 million engineering drawings both old and new, collected in DoD activities and repositories.[15]

But it is not these physical forms (documents) that truly describe the importance of this subject -- it is their knowledge content and the utility of this knowledge for decision-making, for operation and maintenance, and most important perhaps, for the generation of new knowledge.[16]

We cannot think of scientific and technical information as dead records to be bundled up; and stored away. It is perhaps a psychological error to speak in terms of "storage"; if it can be stored, it is dead and its storage is a waste of time and effort. What we are dealing with is live and growing; it must be added to, adjusted, and--above all--kept where it can be reached, examined, used.[17]

That scientific and technical information comprise the life blood of scientific research can scarcely be denied. And, just as the human body will die if the flood of blood closes so will the scientific body wither and die unless the knowledge generated by research flows freely within the research community.[18]

...the test of our generation will not be the accumulation of knowledge. In that we have already surpassed all the ages of mankind combined. Our test will be how well we apply that knowledge for the betterment of mankind...[19]

The fruits of science are now so abundant, and human problems so staggering and complex, that nations and the world can no longer afford the luxury of being casual about knowledge.[20]

We have made a conscious decision that the knowledge that we are generating with rather massive amounts of research and development money must be applied as rapidly as possible. We are no longer content, in other words, to tolerate the inefficiencies of the information-transfer system with which we lived most comfortably for many years... We want to make the system more efficient. This, I think, is all right, but let us bear in mind that it is something we have decided we want to do and not something that is intrinsic in the subject itself.[21]

I look with anticipation to the time when people selling information handling systems will drop the pretense that the requirement is obvious and go and try to determine whether this obvious requirement exists at all.[22]

I believe that the realization of national information systems is one of the most challenging areas of current activity, with great potential impact on not only our science and technology but also on the academic, commercial, and industrial sectors of our society.[23]

Savings almost beyond comprehension may become possible--savings in manpower, materiel and perhaps most important, in time. The savings will not be automatic; at times, they may even prove illusory because hidden costs develop. But over the long run, the savings will be real and substantial.[24]

The executive's present efforts toward finding an optimum system or set of systems to meet researchers' and research planners' needs for information from and about research and development demand stronger support to overcome the parochial interests of individual agencies. The latter tend to proliferate data collection classification, storage, and retrieval techniques.[25]

One cannot expect existing groups to develop willingness to cooperate in a scheme where the purpose is general, intangible and perhaps only of sentimental character. Our goals in setting up a national system must be expressed in terms of the roles and concrete outputs of all groups in that system.[26]

If man's collected knowledge is to become truly accessible, plans and programs must be made, priorities assigned and resources allocated. [27]

Our present (data) system has not emerged as a result of overall planning... [28]

Eventually, information system planners foresee the development of a network of specialized information centers with on-line, instantaneous retrieval capability utilizing remote consoles that perhaps will be placed on the desks of every project engineer. [29]

Why do we worry about the type of machine organization before we worry about the management of information per se? [30]

Because most of the schemes and devices for handling information are so new, their limitations are still not fully understood; in particular, it is not usually appreciated that the new systems generally retrieve documents rather than information. [31]

We need a way of switching information, not documents, to the user in as discriminating a manner as possible. The user should be informed, not overwhelmed. [32]

Back in the late 1930's, H. G. Wells predicted that some worldwide system would have to be adopted for the classification, storage, retrieval, and dissemination of scientific data. We don't have such a system yet; but if Wells had lived through our post-World War II scientific explosion, he would have been gratified that electronics and cybernetics are providing new means for coping with the continuous avalanche. [33]

... it requires enormous intellectual effort to devise a system for ordering such data. One has to know a lot of botany to be a Linneaeus; a successful Chemical Registry scheme is worth a Nobel Prize in chemistry. [34]

I. INTRODUCTION

A. Background

The information activities that support the nation's scientific and technological posture have been the subject of considerable discussion and study during the past decade. At the national level, the U. S. Senate Committee on Government Operations, the U. S. House of Representatives Committee on Science and Astronautics, and other committees concerned with science and technology have conducted numerous hearings and, in general, engendered within both the Federal Government and the broader scientific and technical community a more realistic appreciation of the attention warranted by scientific and technical information activities. Subsequently, efforts of the President's Science Advisory Committee and the Federal Council for Science and Technology have become a focal point that has helped to clarify responsibilities, problems and potential opportunities within a perspective which spans the professions, the Government, and the private sector. Between 1958 and 1965, several Panels and Ad Hoc Committees of these bodies examined various aspects of scientific and technical information activities and related problems.

In 1962, the Federal Council for Science and Technology established the Committee on Scientific Information under the auspices of the Office of Science and Technology, Executive Office of the President. In 1964, the name of the Committee was changed to Committee on Scientific and Technical Information (COSATI) to indicate that its scope of interest included technical, as well as scientific, information activities. The current organizational structure of COSATI, as shown in Exhibit I-1, indicates its extensive involvement in improving Federal scientific and technical information activities. In addition, COSATI, through its own leadership and initiative, is stimulating organizations in and out of the Government, in the United States and overseas, in science and technology and other fields, to seek new methods of improving their capability to communicate information efficiently and effectively both within and between communities. Perhaps the most ambitious of COSATI undertakings, to date, involves the work of a Task Group to study the complex area of planning for national information system(s). The Charter of this Task Group, drawn up in 1964 (see Exhibit I-2), identified two principal goals and objectives:

- Undertake those investigations needed to (a) inventory and evaluate the resources (people, libraries, and other services, equipment, materials and funds) currently being utilized in national and other domestic scientific and technical information activities; and (b) ascertain the information needs of users such as: scientists, engineers, managers, practitioners, and the technical public, as individuals and as groups, in and out of the government.
- Based upon these and other findings, prepare recommendations and plans for the development of national information system(s) to include actions for Government agencies, suggestions for actions by the private sector, and steps to move from current to advanced information systems.

The Task Group recognized very early the hurdles that lie in the path of implementing improved national systems. The COSATI Progress Report of 1965* noted that: (1) There should be no disruption of existing information channels; (2) Account must be taken of widely differing capabilities of existing systems and the realities of funding, long-established practices, rapid changes in information technology, and the differing needs of various segments of the user communities; and (3) The Government cannot direct the private activities that form a major element of the national information capability--that it can only encourage them to join forces in a national system. Recognition of these realities has defined the objectives of the Task Group on National Systems in all of its programs.

The Task Group has sponsored a set of complementary studies to accumulate background information and to assess its relevance to the requirements and feasibility factors relating to national scientific and technical information systems concepts. The first study examined the current status of document handling and made recommendations concerning a national document handling system.** A second study dealt in depth with

---

*<u>Progress of the United States Government in Scientific and Technical Communications</u>, Committee on Scientific and Technical Information of the Federal Council for Science and Technology, Executive Office of the President, 1965.

**<u>Recommendations for National Document Handling Systems in Science and Technology Appendix A--A Background Study--Volumes I and II</u>. System Development Corporation, Santa Monica, California, September 1965. Contract AF 19 (628) - 5166.

abstracting and indexing services in the United States.* Another study analyzed the structures and functions of informal information-communication systems.** Based on the results of these studies, other findings, and the experience of its members, the Task Group is formulating recommendations and plans for consideration and implementation of national information system concepts.

B. Objectives of the Study of Data Undertaken by the Task Group on National System(s)

Prior to 1966, relatively few comprehensive studies had been conducted of scientific and technical data systems. Previous studies and surveys had been restricted to the data activities of an agency, to a specific scientific or technical field, or to selected elements of a total data system -- e.g., data processing equipment. Consequently, the COSATI Task Group on National Systems has undertaken a broad-scope study that can be used to guide the formation of national policy with respect to systems for scientific and technical data collection, reduction, storage, retrieval, analysis, and dissemination. Specifically, the study by the Task Group is intended to:

- Assess the degree of attention that is being given to data on the national level:
- Clarify the role that scientific and technical data -- in various stages of refinement -- play in the technical decision process; and
- Formulate data system policies and/or actions that will benefit the interchange of technological know-how and the conduct of research.

For the purposes of the study, data are described as quantitative or qualitative representations of properties, characteristics, or attributes of objects, events, measurements, or observations. In common usage, data connote factual, as opposed to conceptual information; in addition, the term is

---

* A System Study of Abstracting and Indexing in the United States, System Development Corporation, Falls Church, Virginia, 16 December 1966, Contract NSF-C-464.

** Exploration of Oral/Informal Technical Communications Behavior. Semi-Annual Technical Report, American Institutes for Research, Silver Spring, Maryland, 15 March 1967, DAHC-04 67 C0004.

frequently used to denote the factual information content of a document, rather than the document or artifact itself. The latter distinction permits differentiation between this study and many previous studies which dealt with document handling systems or with abstracting, indexing, or other treatment of conceptual information content of documents. The scope of the study, roughly defined, includes activities involving the following types of scientific and technical data:

- Data acquired in the course of conducting experiments or examining natural phenomena, or in the course of performing tests according to prescribed procedures;
- Data which describe the characteristics or performance of a natural phenomenon, a material, a device, or a component; and
- Data which instruct, guide, or aid skilled or semi-skilled persons in the proper use, maintenance, or replacement of artifacts, or in techniques and procedures.

This study is intended to establish how the various types of scientific and technical data are acquired, stored, retrieved, packaged, and disseminated for specific types of users; why these packaging methods have been adopted; and what changes in methods are foreseen in the future. Special emphasis is placed on uses made of data by various functional groups (e.g., basic research, equipment and systems development, product application, etc.) and the degree of processing or refinement of data needed for such functional groups.

A further objective of the study, which is to be conducted in several phases, is to facilitate an open discourse and provide an opportunity for expression of the views and knowledge of the many individuals and organizations that will be key participants in the development, operation, and use of future data systems.

The scope of the Task Group study can be summarized as encompassing scientific and technical data activities which are potentially amenable to determination of requirements for national data systems or for other types of coordination which would improve our national scientific and technological posture.

C. Approach and Products of the Initial Phase of the Study

In September 1966, Science Communication, Inc. was awarded a contract to initiate this study. Specifically, Science Communication, Inc. undertook to conduct a preliminary survey of scientific and technical data activities, data-related problems, and data system needs within government, the professions, and industry. This initial phase of the COSATI Task Group Study was undertaken to produce a preliminary census of current data activities and a plan for further study and consideration of various concepts of national scientific and technical data systems.

Volume II of this Report summarizes the census-survey findings of this initial phase of the study. Volume I presents a recommended plan for further study and implementation of national data systems concepts, and a discussion of factors considered in the development of the plan.

The following Summary of Conclusions and Recommendations (Section II) highlights some of the more important conclusions reached in this initial phase of the study. It must be emphasized that these conclusions are the result of a very limited effort in terms of the ratio of the size of the study effort to the magnitude of the problem area examined. However, these conclusions, together with the other findings and recommendations in this Report, provide, for the first time, visibility and articulation of the data segment of scientific and technical information activity.

During the structuring and conduct of the survey activities, it was frequently necessary to make decisions as to which areas or modes of study would best meet the purposes of the study. These decisions, by necessity, involved a consideration of the underlying justifications for the study. A summary of some justifications and of current perspectives concerning the role of data is presented in Section III. Also, it was necessary to formulate an internally consistent set of concepts and terminology to guide and unify the survey efforts. A selected set of these concepts with their definitions is presented in Exhibit I-3. Additional discussion of the structuring concepts is contained in the Introduction to Volume II of this Report.

Volume II provides background information for consideration of the Plan presented in Volume I. Primary elements of this census include state-of-the-art descriptions of data activities in selected communities of science and technology, a preliminary census of formal data efforts in science and technology, and findings from selected survey-probes of data management and handling capabilities, activities, and problems in typical institutional or organizational settings. In total, findings from these complementary surveys provide a summary of the current state-of-the-art in scientific and technical data management and data handling efforts.

The four survey techniques used to develop the preliminary census of data activities were literature review, mail questionnaires, workshop discussions, and personal interviews with leading data specialists and managers. These same techniques were employed to derive the plan for continuation of the study. Concurrent conduct of these two aspects of the study permitted meaningful integration by encouraging the consideration of current facts and opinions in the evolution of recommendations for future data systems study and development.

This Report equips the COSATI Task Group on National Systems and other interested organizations with a preliminary definition of the challenge awaiting those who will assume the leadership in creating improved scientific and technical data systems In addition, it provides a focal point for reviews and discussions to further define required actions and enlist the participation of the many organizations and individuals who must contribute to the development of improved scientific and technical data management and handling systems. Suggestions for further background reading are enumerated in the Selected Reading List, Exhibit I-4.

II SUMMARY OF MAJOR CONCLUSIONS AND RECOMMENDATIONS

Man's physical awareness, and to a lesser extent, his total existence, is defined by his ability to sense and comprehend his environment and his relationship to it. Creation of symbolic representations of this environment and use of these representations in constructive communications is a vital function of modern societies. A significant change in the manner in which this function is performed in any significant segment of our society can alter the nature of human affairs. Science and technology are in the midst of such a change.

Our study has only begun to sketch this change, as evidenced in current scientific and technical data activities. However, continued change or transition in data management and data handling systems is inevitable and can be expected to continue to grow in significance. Consequently, steps should be taken to characterize the transition, provide visibility to its nature and importance, and to enlist the resources required to guide the change in the direction most beneficial to our national scientific and technological posture. If properly applied, means are currently available to improve current data management and data handling practices. In addition, it is feasible to initiate development of new data handling systems which promise quantum increases in the utility of our national scientific and technical data resource.

The conclusions and recommendation of this preliminary study are not highly prescriptive as to the configuration and functional structure of national data handling systems. Rather, primary emphasis is given to identification of actions which will evolve goals, competencies, and motivations which can be integrated into a comprehensive, yet decentralized program to achieve optimum utility from our national scientific and technical data resource. The recommended program should not, in fact cannot, be implemented on a crash basis; neither can its implementation be delayed if the U. S. intends to maintain its position of preeminence in science and technology. The major recommendations in this report are offered as a preliminary blueprint for establishment of a National Scientific and Technical Data Program. If the recommended program is initiated in FY 1969, national scientific and technical data systems could be a functional reality as early as FY 1975.

A. Scientific and Technical Data--Perspectives and Policy Implications

CONCLUSION: The utility of our national scientific and technical data resource can be substantially increased by improved management.

Although the Federal Government and non-government organizations in the U. S. jointly expend in excess of 20 billion dollars annually to support research and development efforts, few policies exist to guide the management of the resultant scientific and technical data. In fact, current policies and programs have resulted in an imbalance between the efforts expended to generate new data and the efforts expended to maintain, evaluate and make maximum use of available data.

RECOMMENDATION: The Executive Office of the President should issue a policy statement establishing the objectives of a national program to improve the management of scientific and technical data activities within government, the professions, and industry.

The statement should not only identify goals but designate responsibilities and identify means to achieve the program objectives. Specifically, the program should achieve redistribution of Federal expenditures so that an appropriate percentage of each agency's research and development budget is allocated to data activities. The program should involve upgrading of existing data systems and services, and development of the capabilities required for implementation of improved systems for future data management and handling. In addition, the program should provide funds to support exploratory study and planning of data management and data handling systems in those areas of science and technology which can most benefit from such study and planning. The total program should be structured and administered in a manner which will assure appropriate participation by all sectors of the scientific and technological community.

CONCLUSION: No effective means currently exists for coordinating and integrating the data management and data handling activities of the governmental, professional, and industrial sectors of science and technology.

Despite the extensive efforts of COSATI and the respective agencies of the Federal Government to coordinate intra-government scientific and technical information activities, no broad-gauge means has been established to coordinate these efforts with those of non-government organizations. The National Science Foundation is currently funding and working with professional societies to establish discipline-based information systems, and many other government agencies cooperate with a limited number of non-government organizations. For example, the Department of Defense co-sponsors meetings and working groups with data-oriented sub-groups of the American Ordnance Association and the National Security Industrial Association. However, existing arrangements are not adequate to support the implementation of the program required to achieve optimal use of the national scientific and technical data resource.

RECOMMENDATION: A National Advisory Council for Scientific and Technical Data should be established.

The Council membership should represent the various segments of scientific and technical data activity, both governmental and non-governmental. The Council should function principally as a review and consultative body and should be structured to permit the operation of Panels concerned with the following types of data activities: (1) Discipline-research (scientific) data activities; (2) Developmental-mission data activities; (3) Applications-product data activities; (4) General-purpose data activities; and (5) Data system technologies and development activities. Through operation of its

Panels, meetings such as a White House Conference and special studies, the Council would provide a forum dealing with data management and data handling system requirements. Based on inputs from this continuing forum and Council review of current data management practices and data system performance, the Council would evolve recommendations to guide the National Scientific and Technical Data Program. The Council should maintain a small, permanent staff which would function as the secretariat for the Advisory Panels and would monitor planning and other special studies initiated by the Council.

CONCLUSION: Scientific and technical data and data activities are exceedingly complex; national data programs and system development efforts must be capable of effectively recognizing and accommodating this complexity.

The extent of this complexity can be gained by considering the attributes of scientific data activities associated with discipline research as compared with those technical data activities associated with mission development or product application activities. Each of these types of data activities use data (even the same data) in different ways, prefer specific packaging methods, and are driven by different motivational factors in the generation, handling, and application of data.

RECOMMENDATION: National data programs and related policies must be implemented with due consideration of the diverse types of data activities which are conducted as an integral and vital part of science and technology.

Present knowledge indicates that the National Scientific and Technical Data Program should consist of at least two subprograms - one formatted to develop improved data management and handling systems for scientific data activities, and one formatted to the requirements of technical data activities. This would permit the total program the flexibility required for effective interaction between government, industry, and the professions. More specifically, it would permit the Federal Government to tailor its support of data system development efforts in accordance with the extent of the public interest.

In addition, the National Data Program must be structured to complement and build on existing data efforts, both governmental and non-governmental. Especially, it should provide a means for including a voice for the interests of the pre-existing data service programs in the mission-oriented Federal agencies and in commercial service firms such as publishers.

CONCLUSION: The full utility of scientific and technical data is not currently realized under existing data management and data handling policies.

Analogous to the preservation of academic freedom, the individual scientist has rightfully striven to preserve independence from external influences on the conduct of his scientific work. As a consequence, the scientific community has avoided centralized coordination of its activities, including conservation of the knowledge structures which constitute the bases of the various disciplines. It is assumed that the informal structures, such as the invisible college and the more formal mechanisms associated with publication of the scientific journal and monographs, provide adequate vehicles for communicating data and for maintenance of the unity which is so vital to a strong science. This assumption fails to give due weight to the changing character of the role which our society expects science to perform and the changes which result as scientists attempt to react to these new expectations. It also fails to note the inadequacies of the scientific paper, professional journal, and abstract publication as a set of tools for the communication and maintenance of the increasing pool of factual information or data being generated by scientists.

Although due to other causes, the data management and conservation practices currently followed in developmental or applications activities in science and technology are equally ineffective. For example, in spite of the large amounts expended to develop costly items of equipment for applications, such as defense and space exploration, data describing these equipments are not maintained in well-structured, full-indexed files readily accessible to other potential users.

RECOMMENDATION: Each scientific or technical community, including mission-oriented agencies, should reappraise its current procedures for managing and handling scientific and technical data, especially in regard to their adequacy for conservation of the data as a costly and potentially reusable resource.

Each community should establish a focal point and procedures to identify the characteristics of its data resources, to articulate data management objectives, and to formulate plans and programs to implement data handling systems development or other actions to achieve the goals of the community. In effect, each community should create a body (office, committees, etc.) to serve as a spokesman and coordinator for the cooperative data activities of the community. Such data activity coordinators might be housed in professional societies, trade associations, educational institutions, or government agencies; however, participation in the activities of the coordinating body should be open to all members of the community. In fact, the ability to assure a large involvement of the community in its activities should be a prerequisite for recognition of the coordinating body as the designated agent to participate in National Scientific and Technical Data Program development. In addition, such groups should explore the feasibility and potential utility of data indexes or inventories, data-source referral services, computer-composed handbooks, computer-maintained data banks, and other means of maintaining and increasing the utility of the data base serving a given community of users. The coordinating body should serve principally to act as an initiator for the development of such services; however, if other means are not feasible, the coordinating body could undertake provision of such services.

CONCLUSION: There is inadequate knowledge concerning the nature (quantity, quality, location, ownership, usefulness, etc.) of existing scientific and technical data to permit optimum design of national data management programs or data handling systems.

It is amazing that, despite the large expenditures for generation of data and level of sophistication generally acknowledged to exist in administration of scientific and technological activities, a comprehensive inventory of data does not currently exist for any of the major communities in science and technology. As a consequence, many important decisions, concerning either scientific and technological programs or data handling, are made with little detailed information concerning the available data resource. In many cases, Federal contracts or grants are awarded for generation of new data when the awarding agency is unable to supply the contractor or grantee with information concerning the nature, quality, availability, etc. of previous measurement results.

RECOMMENDATION: A National Index of Scientific and Technical Data should be developed. Such an index is essential if data management is to be planned on a systematic basis. Also, such an index would be immediately useful to scientists and technologists who currently expend as much as 30% of their working hours searching for data required to perform their job.

The index envisioned is not one single, comprehensive index; rather, it would consist of sub-indexes covering the data resources useful within individual communities. The initial index should cover existing scientific or technical data files, data reference tools such as handbooks, and computer processible data. Subsequently, it could be extended and refined to index data at the document level. The form of the index should be adapted to the requirements of the using communities. In many cases, the index will be most useful if compiled and maintained in a computer searchable form. The National Index could be pursued as a central element of a National Data Program with the index compilation activities in each scientific or technical community being coordinated by a selected organization such as a professional society or trade association. Commercial firms could disseminate the Index in either hardcopy or computer processible form.

CONCLUSION: Federal policy relative to scientific and technical data management must recognize and facilitate maximum use of the existing scientific and technical data resource.

Science has fostered wide dissemination and accessibility to scientific data with restrictions limited largely to those conventions required to acknowledge the generator. To a lesser extent, copyright has been employed to enable the publisher to maintain an economically viable channel for dissemination of data. In contrast, mission-oriented agencies of the Federal Government and commercial firms frequently restrict the distribution or accessibility of the data which they generate. In some instances, it is because the agency or firm does not wish to disclose its position; in other cases, the agency or firm has no desire to restrict the data, but has no incentive to expend the funds and/or effort required to make the data available to other users.

RECOMMENDATION: The Federal Government should establish a policy to encourage the accessibility of scientific and technical data to as many potential users as possible.

Such a policy would not conflict with full recognition of the property rights of individuals or organizations. Rather, it would be promulgated with a specific delineation of private data (data which an individual or organization does not desire to disclose or release), proprietary data (data which the owner or possessor will release under prescribed conditions such as payment of a fee), and public data (data for which ownership and possession is in the public domain). Government support should be given to efforts for removal of the barriers which result in data being restricted when, in fact, the owner or holder has no objections to use of the data by others.

Initially, the Federal Government should take actions to see that data generated under its sponsorship is managed so as to assure maximum use. For example, all Federal research and development programs should direct a minimum level of effort to this objective. This should be supplemented by a central clearinghouse which would support the special data husbandry operations required to move data from a restricted or limited use context (e.g., an agency project file) to one where the data has higher visibility and greater use potential (e.g., a government-issued index to data of potential interest to a specific user group, either government or non-government).

In addition, where commercially generated data have high utility for a large or significant segment of a scientific or technological community, the Federal Government should be prepared to underwrite the cost of organizing and disseminating the data. However, this should be undertaken only when costs or other factors preclude such actions by commercial service firms.

CONCLUSION: As data handling becomes increasingly automated, the need for standardization of data handling methods will become increasingly important to the National Scientific and Technical Data Program.

To a lesser extent, there will also be factors leading to increased need for some standardization of the form and quality of data. However, any steps towards standardization of data form and quality must be taken with caution and with a full appreciation of the implications for the conduct of scientific and technical work. As the major supporter of scientific and technical work, the Federal Government has a vital interest, as well as the means, to assure that standardization requirements are delineated and implemented.

RECOMMENDATION: The Federal Government should take action to assure development and application of standardized methods of handling basic scientific data, especially those automated methods broadly applicable to data systems in more than one field of research.

Scientists in specific areas of research must make the final determination of whether standardization of measurements and data is feasible or desirable. Whereas, Government-initiated standardization of data handling methods supporting research on a broad basis appears desirable, standardization of data handling methods supporting development or applications activities does not appear warranted except within specific Government development programs. Industry, through cooperative arrangements, should be encouraged to upgrade and standardize its developmental and applications data activities. In situations where it can be shown that standardization will contribute to a better integrated and stronger national scientific and technological competence, the Federal Government should, if required, subsidize standardization efforts. At a minimum, the Government should provide increased technical assistance and financial support to standardization activities.

CONCLUSION: The diverse connotations assigned by different communities, organizations, and individuals to scientific and technical data, data artifacts, and data management and handling efforts constitute severe barriers to systematic planning and evaluation.

For example, in engineering and other applicant-oriented activities, data is frequently used to connote all documentation required for accomplishment of the scientific and technical objectives of the project, program, or organization. Whereas, in research or discipline-oriented activities, data is used to connote factual information as contrasted to conceptual information. Further, a preliminary review of the currently accepted definitions in the various Government agencies are not consistent. One result of this non-standardization is inefficiency and increased costs incurred by contractors and other non-government organizations who deal with more than one government agency.

RECOMMENDATION: The Committee on Scientific and Technical Information (COSATI) should promulgate a set of definitions which delineate an internally consistent set of terms covering scientific and technical data activities.

These definitions, if carefully formulated, will provide a guide to encourage consistent usage of scientific or technical data terms throughout the Federal Government. The existence of an acceptable set of terms will facilitate the establishment of more precise and effective policies and procedures dealing with data activities. Specifically it will facilitate acquisition of data from non-government sources and will make it easier to communicate concerning data between government offices.

CONCLUSION: Just as science is international, scientific and technical data activity is often international in scope.

In many areas of science and technology, such as atmospheric science, it is very important to obtain data on a world-wide basis. Where these needs exist, many scientific and technological communities in the United States, through the International Unions and similar organizations, have become participants in international data activities. Currently, much international data activity involves multi-nation efforts to collect data on a world-wide basis. In many cases, these data will constitute a part of the data base which future national data systems must handle. Consequently, it is critical that U.S. participation in such activities be planned and conducted on the most informed basis possible. A current problem is that the attention which responsible offices have been able to give to this activity has not kept pace with the increasing volume and importance of international data activities.

Another aspect of this question is the immense size of the total data management effort; the U.S. cannot hope to independently perform this function for all areas of science and technology.

RECOMMENDATION: Offices in the Federal Government designated as responsible for representing U.S. interests in the area of international data activities should be strengthened not only to permit them to better represent U.S. interests, but also to enable them to establish better communications and working relationships with on-going activities in the U.S.

The requirement for effective coordination of U.S. involvement in international data activities and program development is expected to continue to increase as data management becomes more formalized and national data systems are developed. It will become increasingly important to guard against unilateral actions by individual organizations or communities.

Also, a means must be established to determine which areas of scientific and technical data management the U.S. will undertake jointly with other countries and which would best be pursued totally or largely by the U.S.

B. Current Issues and Problems-- Nature and Possible Resolutions

CONCLUSION: The inadequacy of classifical methods for structuring and communicating scientific and technical data in current working contexts has created unnecessary apprehension.

The more evident symptoms of this apprehension include fears, such as those voiced in the Weinberg Report, that science could lose its unity and effectiveness by fragmenting into a mass of repetitious findings, or worse, into conflicting specialties that are not recognized as mutually inconsistent. Subsequent study indicates that the existing apprehension represents the expected preamble to a significant change in data management and handling methods.

RECOMMENDATION: The currently evolving expressions of need for large-scale scientific and technical data handling systems should be viewed as a response to opportunity, not an act of desperation to avoid inundation by the flood of data.

Science and technology has not only generated massive quantities of data; it has also developed computers and other tools which offer unprecedented opportunities for improved management and handling of this data resource. In fact, these tools, if properly applied, offer the potential for quantum increases in the uses of data, thereby not only increasing the return to supporters of science and technology but also reducing the total cost involved. The nation should no longer tolerate only partial use of data; data is not consumed by use, but rather gain information value with reuse.

CONCLUSION: Current research and development administration, especially within Federally-sponsored programs, frequently gives preferential consideration to research or development to generate scientific and technical data over activities directed to assembly, evaluation, and application of existing data.

It has been stated on several occasions that the individual scientist is frequently encouraged by diverse factors, some sociological and some related to the current nature of research funding and administration, to undertake new measurements before fully digesting previous measurements. In many instances, individual scientists or technologists, as well as research and development projects, have found it easier to repeat measurements than to locate the results of previous measurements. Such regeneration of data can be very expensive, especially if it should require flight testing of a supersonic aircraft such as the RB-70 or launching of an instrumented satellite.

RECOMMENDATION: Each Federal research and development program should be required to allocate a minimum percentage of its budget to husbandry and conservation of the scientific or technical data generated by the program.

For example, basic research programs might allocate 10%, applied research programs 15%, and developmental programs 5%. These funds need not be identified as line items in the agencies' budgets, but an annual report of compliance should be made to a centralised review body such as the Bureau of the Budget or the Office of Science and Technology. The intent of this recommendation is to assure that a reasonable effort is expended to conserve data generated by Federal expenditures and to assure that it is readily accessible for application in either other Federal programs or in the other sectors of our society.

CONCLUSION: Although essentially the same problems are observed in data handling activities in the different fields of science and technology, no mechanism now exists for the coordination of efforts toward solution of these problems.

Perhaps, the reason for so little comparison of experiences is the belief widely held by the directors of data-document depositories, data-evaluation centers, etc. that the problems faced by each data handling effort are unique. Although this study has confirmed the uniqueness of some aspects of data handling efforts, it has also found that most efforts encounter similar difficulties in the areas of application of new technologies, financing of costly development efforts, and recruitment and retention of capable personnel. Also, present operation philosophies do not indicate an awareness of the potential interaction between data efforts serving a common community of science or technology. The current evaluation of the effectiveness of many of these efforts is that it is very low because they do not permit effective interaction between the data resource and the potential user.

RECOMMENDATION: The Federal Government should establish an information center to serve as a depository and dissemination agency for information dealing with design, development, operation and management of scientific and technical data systems.

The center should serve to support participants in the National Scientific and Technical Data Program, especially the National Advisory Council for Science and Technology. The services of this center should be extended to non-government as well as government offices. Such a center could be established by consolidating and augmenting some of the current information service activities of the NBS Research Information

Center and Advisory Service on Information Processing, the NSF Office of Science Information Service, and the Bureau of the Budget Management Study File.

In addition, professional societies, such as the American Society for Information Sciences, should establish panels or subgroups of data system professionals and should undertake development of publications and meetings to communicate developments concerning scientific and technical data management systems. In addition, the roles and functions of these existing efforts should be re-examined as part of the National Scientific and Technical Data System planning. Careful consideration should be given to ways in which the operations of the efforts serving a given community of users could be coordinated to contribute more to effective data management.

CONCLUSION: Current data service requirements are largely undefined.

Additionally, effective methods are not available for predicting future data requirements. This factor, as much as any other single factor, has restricted the development of large-scale data handling systems.

RECOMMENDATION: Existing data service centers such as the National Oceanographic Data Center and National Space Sciences Data Center, and new prototype data resource centers, should be used as test beds to study data service needs.

Factors to be examined should include the effect of usage levels and user satisfaction resulting from the availability of different configurations of data services. For example, remote console access to a centralized data bank should be compared with desk-top microfilm services. Other factors examined should include degree of data evaluation, format of data presentation, user charges, etc.

CONCLUSION: Current scientific and technical handling practices do not fully employ available technologies.

Despite extensive use of computers for performance of mathematical computations, science and technology have only recently begun to exploit the computer as a tool for structuring, storing, and maintaining large files or banks of scientific and technical data. A more mundane example of the lagging use of technology is evident in the current practices for composing and disseminating data documents or artifacts. For example, despite the technological capability to maintain handbooks essentially current, most handbooks are five years or more out of date. If significant advances are to be made in the application of new technologies, knowledge must be gained concerning the effectiveness of these new tools for performance of specific data management and handling functions within real-world work environments.

Application of currently available data handling technologies offers potential for substantial increases in the utility of the existing scientific and technical data resource. This increased utility can be achieved by two means: first, by performance of current activities in a more effective manner; and second, by using new technologies to conduct activities previously impossible. An example of the first means would include computer maintenance and searching of indexes to data. The second application, which offers the greater potential, is to use large automated data files to perform pattern recognition or other types of higher level analyses Another new use potentially exploitable is computer-aided design which brings the data and the computer into the daily work pattern of the scientist or technologist

RECOMMENDATION: The Federal Government, professional societies, trade associations, commercial publishers, and other collectors and disseminators of scientific or technical data should explore means of applying modern technology for more effective assembly and dissemination of scientific and technical data.

Areas to be examined should include use of computers to maintain the data base and compose handbooks. Also, microfilm and computer processable media should be more extensively used to disseminate data. In appropriate cases, data should be disseminated in more directly useful forms such as in combination with computer programs for design, diagnosis, or other applications of data.

The Federal Government should sponsor demonstration projects in which innovative data handling tools and media would be tested. Some of these demonstration projects should be in government programmatic contexts and some in non-government contexts. These demonstration projects should be conducted as controlled experiments with results carefully documented for educational and training purposes. Where possible, existing projects intimately associated with on-going research and development should be used as demonstration projects. A typical candidate project might be the National Institutes of Health Chemical/Biological Information Handling Program. This program was only recently initiated and is intended to develop an on-line data system to serve the researchers involved in the NIH Toxicology Research Program.

CONCLUSION: The lag time between data generation and dissemination using traditional publications is frequently from two to five years.

This lag time is caused by several forces and conditions which prevail primarily in the scientific community. First, the practicing scientific investigator is motivated to publish data only upon generation or verification of a significant theory or hypothesis; and data generated at the outset of a theoretical investigation may therefore not be published for several years. Secondly, the time lapse between preparation of a paper and actual publication may be as much as one year, because of the slow review process and the backlog of papers that exist in scientific fields. Thirdly, the additional effort required to publish data which do not relate directly to interim investigative conclusions, deters and sometimes eliminates its publication.

RECOMMENDATION: Programs should be developed to more directly couple experimentation, tests, etc. with data systems.

As on-line use of computers in scientific investigations becomes a widespread reality, automated data banks should be developed, particularly in the physical sciences, environmental, and geosciences. Pilot programs should be implemented to determine the feasibility of such data banks and to examine the associated problems, especially structuring and access aspects of such systems.

CONCLUSION: Although the total investment applied to generation of data concerning products and processes far exceeds that applied to generation of basic scientific data, inadequate effort is expended by the Federal Government to organize this data for secondary uses.

As an example, files and search procedures do not exist to permit a potential user to locate data describing previously developed equipment meeting a given set of performance characteristics. Such data normally cannot be located unless the searcher has informal knowledge concerning the probable location of the data.

RECOMMENDATION: Current efforts, such as the Department of Defense Engineering File, should be substantially accelerated, and other equipment development agencies without such systems should initiate study of their feasibility.

A logical start toward such systems would be an inventory operation to develop an index to the existing files. If this were done in a number of agencies, it would make a major contribution toward the National Index of Technical Data.

C. Systems Development -- Requirements and Implementation Concepts

CONCLUSION: It cannot be expected that existing groups will cooperate in efforts to develop national systems where the purpose is intangible or Federal domination might restrict the legitimate freedoms of scientific groups, commercial firms, etc.

Many scientists and technologists object, as a matter of principle, to the involvement of the Federal Government in planning or coordination of scientific and technical data management programs or data handling systems. It must also be noted, however, that an equal or greater number recognize that neither the individual scientist and technologist nor the professional organizations have the necessary resources or have exhibited a capability to assume responsibility for creating data management and data handling systems responsive to current needs.

RECOMMENDATION: A National Scientific and Technical Data Program must be planned and administered in a manner which accommodates the interests and capabilities of diverse groups and organizations.

The structures of data systems cannot be dictated by fiat from a top-level policy position. Rather, such structures must evolve from working-level responses to real needs. In fact, national systems are already developing in this fashion. The current need is for coordination and financial support of these evolving systems. Each scientific and technological community must be encouraged to contribute to formulation of goals and implementation plans for national systems. This can be facilitated by the establishment of an office or other type of organization to serve as a focal point for national level data system planning.

This organization should be located and staffed in a manner which permits participation not only from government but also professional societies, trade associations, industrial organizations, and educational institutions. Centralization of responsibility for national system development should be limited largely to programming and broad planning functions. Detailed planning implementation and operation should be on a decentralized basis.

CONCLUSION: The inability to define a single system structure responsive to all data management and data handling requirements does not constitute a valid justification for delaying consideration of new or improved data systems.

Data management systems employing modern technologies such as the computer are still only in the concept definition phase. Current knowledge as to how best to use these tools will not support a crash program to create a large-scale, highly automated and totally integrated system. First, more must be learned as to which functions are most important and how each function can best be performed.

RECOMMENDATION: The present should be recognized as a timely point for initiation of national systems planning and development efforts.

It is anticipated that at least six years will be required to develop national data systems serving specific communities in science and technology. This period can be profitably used to explore alterna system configurations, and relative effectiveness in serving specific data management requirements. This effort should produce a base of knowledge which would substantiate later decisions relative to the extent to which such specialized systems could be integrated into a more unified system.

CONCLUSION: Data management is in a state of transition which is being driven largely by the introduction of computer and other improved data handling methods.

For the foreseeable future, data management must continue to be a decentralized process directed by the scientists, technologists, and administrators responsible for specific scientific and technical endeavors. However, as data system management methods and systems are developed and implemented, a capability will be created for management of larger and more complex sets of data.

RECOMMENDATION: In the near future, efforts at the national level should be directed toward the development and test of systems or tools to facilitate better data management.

Initially such tools or systems should be designed to facilitate currently definable data management functions, such as identification of the location of relevant data. As soon as data management functions are defined, data management requirements should be analyzed and articulated for workers at all levels from the bench scientist to the administrator [illegible] ational scope scientific and technical efforts. This should be d[illegible] jointly by systems analysts and the workers involved in each lev[illegible] of activity.

CONCLUSION: The most valid requirements for development of national scale data handling systems exist for systems operating within scientific and technical communities rather than between communities.

This conclusion is derived from a consideration of the volume and frequency of intra-community communication of data versus inter-community communications. And to a greater extent, it derives from the feasibility of being able to effectively identify intra-community data management and handling requirements as compared to the feasibility of identifying such requirements on a multi-community scale.

RECOMMENDATION: National data system development efforts should be focused on individual communities.

These communities will probably be defined on several bases. In one case it might be on the basis of the common discipline; in another, it might be on the basis of a common mission objective; whereas a third might be based on an interest in a type of process such as metal fabrication.

CONCLUSION: Data handling systems are tools to facilitate data management.

Therefore, implementation of effective data handling systems is dependent upon a prior definition of data management objectives. Unfortunately, individuals currently attempting to develop data handling systems are frequently forced to proceed without adequate definitions of data management objectives. As a consequence, the data handling activity frequently fails to interact effectively with the scientific or technical program objectives. More specifically, many data evaluation and service centers do not engage the individual scientist or engineer within his normal daily work routine. In contrast, the computing center requires an explicit statement of the user's objectives before it can undertake to serve him. Consequently, the computing center has become a vital element in the normal work routine of the scientist and engineer.

A survey (Volume II of this report) of formal data efforts currently serving scientific and technological communities reveals that the data efforts serving any given community operate totally independently. In other words, they do not formally view themselves as facilitating accomplishment of a common set of data management objectives.

RECOMMENDATION: The total data management requirements of a community to be served should be examined prior to implementation of data handling systems as part of the National Scientific and Technical Data Program.

This examination should not only include the management requirements which generate a need for archival data handling, but also those requirements which generate needs for data transmission and data processing associated with generation or use of data. This approach will result in maximum service to the user because it not only will enable the user to have a voice in the system design, but will encourage the integrated application of different data handling efforts.

CONCLUSION: National scientific and technical data system development efforts must consider not only the scientific or technical field to be served by the system, but also the specific type or phase of activity to be served.

For example, the public interest (i.e., non-commercial interest) is high in discipline-research related data activities because such activities generate and maintain data widely useful in our current society. Whereas, non-public or commercial interest is greatest in product-applications data activities. Economic or profit-oriented incentives are easily discernible in the case of product-applications data activities; whereas, they are practically non-existent in discipline-research data activities. Another relevant factor is the stage of development of a data system. It is likely that many data systems could be self-supporting, once established. But the time and cost required to establish the system constitute the threshold barrier.

RECOMMENDATION: Federal support of national scientific and technical data programs and systems should be pro-rated according to the type of data activity served and the stage of the data program or system.

In effect, what is suggested is a cost-sharing plan whereby the Federal Government's share is high for systems or programs to serve discipline-research or non-commercial data activities and low for programs or systems to serve product-applications data activities. In either case, the share of cost borne by the Federal Government would decrease progressively as the data system advanced from program planning, to system development, and finally, to system operation. For example, the Federal Government might bear 100% of the cost of planning for a data system to serve discipline-research in chemistry, 75% of the development costs, and 25% or less of the operating cost. In contrast, the Federal Government might bear 90% of the cost of planning for a data system to serve the food processing industry, 50% of the development costs, and none of the operating costs.

CONCLUSION: To be effective, data service operations must be complementary to the normal work routines of the scientist or technologist.

Many of the currently operating data evaluation centers and other data service efforts are ineffectual because they are too far removed from the daily service needs of the worker. This occurs because the operations of these services do not begin until the generator of the data has recorded, analyzed, printed and disseminated the data. Often four or more years pass between the date when data are recorded and when a data evaluation center offers them to a secondary user.

RECOMMENDATION: A part of the National Scientific and Technical Data Program should be the development of integrated data resource and service centers.

Data resource centers which incorporate to one facility several of the data handling systems and services which the scientist or technologist now must use separately should be tested. The data resource center could provide the user ready access to data acquisition facilities, computing equipment, automated archives of relevant data, archives of computer routines, reactive display consoles, automatic report generators, and long-distance communication terminals. If established within a project or other context where workers were engaged in a joint effort, the center could test techniques for communication from worker to worker as well as from worker to a data resource. Such resource centers could also be used to test the feasibility of on-line data reduction during experiments or tests, and the concepts of working data files and archival data files, concurrently accessible to a worker at his individual console. Preliminary studies indicate that such centers could be developed relatively economically.

CONCLUSION: Data systems complement other information systems; however, it is short-sighted to view data systems as simple extensions of document handling systems.

This short-sighted view fails to give due consideration to the extensive interactions between the scientist or technologist and his data prior to publication. In recent years this interaction has increasingly involved use of the computer in analysis and evaluation of the data. This view also fails to give due weight to the large volumes of data which are exchanged through channels other than publication, e.g., the data (specifications, engineering drawings, test reports, etc.) flow which occurs within a program to design and develop a satellite launch vehicle.

During the past decade, an imbalance has developed between the emphasis which the Federal Government and other organizations have given to study and development of document handling systems as compared to the emphasis given to factual information or data handling systems. Practically all of the more than 20 plans for national scientific and technical information systems put forth during the past decade have dealt exclusively with the problem of handling documents. Few of these plans seriously considered the extent to which documents perform optimally as the vehicle for the two major functions of information systems--communicating and archiving knowledge.

RECOMMENDATION: Data management and handling systems in their ultimate form should be viewed as providing a capability for a totally new level of interaction between the scientist or technologist and the accumulated data resource.

This ultimate goal cannot be quickly realized, however much of the required effort is already being expended. What is involved is not a radical change in level of effort; rather, it is coordination and better direction of current efforts, supplemented on a selective basis. Existing data programs could be integrated to form the major volume of operations in a national data system. As an example of a simple initial step, document handling systems could initiate indexing of the data content of documents processed.

D. Systems Capabilities - Assessments and Remedial Actions

CONCLUSION: Current research and development directed specifically to study of critical factors important to develop of large-scale scientific and technical data handling systems is totally inadequate.

A wide disparity currently exists between the technical capabilities of data processing and transmission devices and knowledge as to how best to apply these capabilities in scientific and technical data handling systems. Projects MAC and INTREX at the Massachusetts Institute of Technology, the laboratory automation project at the California Institute of Technology, and the Information Resource Center element of ILLIAC-III at the University of Illinois are representative of the types of studies which are needed. These three projects illustrate the varying levels at which the national systems concept should be studied.

RECOMMENDATION: The Federal Government should budget at least one-tenth of one percent of its total annual expenditure on research and development for research on techniques and procedures for managing and handling scientific and technical data.

This research should provide general support to data management and data handling activities and should not be directed to development of methods or tools for specialized applications. In order to assure efficient use of these funds, the administration of this research program should be centered in one agency, such as the Institute for Computer Technology at the National Bureau of Standards.

CONCLUSION: Current personnel and institutional capabilities are not adequate to support a crash program to develop a national scope scientific and technical data handling system.

Survey of professional societies, trade associations, computer service centers, etc. and discussions with leading data specialists have revealed a low incidence of serious consideration of, or work toward, establishment of large-scale scientific and technical data handling systems. Exceptions to this general observation were noted only in operational- or mission-oriented areas of activity such as weather forecasting, air pollution control, etc.

Outside of limited programs in Government agencies such as the Department of Defense, formal educational and training programs for scientific and technical data specialists, scientists, and managers are, for all practical purposes, nonexistent. In addition, sociological factors and career management practices currently discourage the more capable scientists and engineers from engaging in scientific and technical data handling efforts

RECOMMENDATION: Information and data managers should be developed from two sources - one is the current population of working scientists, engineers, data processing specialists, etc.; the second is the current and future population of students in colleges and universities.

As an interim measure, the Federal Government and other employers of scientists and engineers should search for individuals interested in scientific and technical data systems and should support the special training required for those individuals to become proficient in analysis, design and operation of modern data systems. Adequate training programs are not currently available; and since the amount of training needed now and in the near future is substantial, employers should foster such training in any institutional setting where it can be conducted successfully. In addition, employers must provide incentive for their employees to undertake such training by establishing job positions and career development opportunities. In contrast to the interim solution, the long-term solution hinges on the colleges' and universities' developing the capability of introducing all students to modern data management systems, regardless of whether the student later becomes a data system specialist or a scientist or engineer who will be a user of such systems.

CONCLUSION: Although data switching networks and computers are frequently mentioned in juxtaposition to one another, automated data service networks, for all practical purposes, do not currently exist within science and technology.

Among the several reasons for this are: (1) An inability to define user needs which provide economically justifiable requirements for such data service networks; (2) The current high costs of data transmission and remote access terminals; and (3) The difficulty of structuring and maintaining centralized data banks of sufficient breadth to serve diverse user groups.

RECOMMENDATION: Appropriate organizations should test the effectiveness of centrally supported, decentralized data resource centers as an alternative systems concept to data switching networks.

For example, the National Institutes of Health might support and provide centralized data collection and selected programming services for a series of data resource centers located at leading medical research centers, or the Department of Transportation might support data resource centers at laboratories involved in highway safety research. Each local data resource center would be configured so as to use data files and manipulation programs furnished by the central service unit in the sponsoring agency. The users of the data resource center would augment the basic data file with locally generated data or data assembled because of high utility in local work. Periodically, the central service unit would obtain read-outs of locally generated data to ascertain if it should be packaged for distribution to other local resource centers. The initial tests of the local data resource center concept should be conducted as controlled experiments with cost and effectiveness parameters carefully documented and analyzed so as to provide guidance for planning of national data systems.

A major objective of these tests should be development of data file structuring and access methods which are considered the key barriers to large-scale data handling systems of the future.

CONCLUSION: There is an almost complete absence of criteria for the evaluation of the economic performance of current data handling efforts.

Most past efforts to apply cost-benefit criteria to measure the effectiveness of data efforts have been inconclusive, due to difficulties in quantifying the benefits from the operation of such efforts.

RECOMMENDATION: Cost-effectiveness should not be the principal criterion to determine whether or not efforts should be initiated to explore the feasibility of improved data handling systems.

Until effective methods of data management and handling are demonstrated, effectiveness, rather than low-cost, should be the major aim of development objectives.

III. SCIENTIFIC AND TECHNICAL DATA ACTIVITIES -- AN OVERVIEW OF PERSPECTIVES, OPPORTUNITIES, AND STUDY REQUIREMENTS

An attempt to observe, measure, comprehend, and ultimately control his environment constitutes a continuing mission of man. Man's ability to accumulate, communicate, and apply information paces pursuit of this mission. Modern society requires that each person make numerous decisions every day which are based, at least in part, on data which are the results of controlled tests or observations. For example, a picnic might be scheduled or cancelled on the basis of weather observations, or the insect repellent for use at the picnic site might be selected on the basis of toxicity measurements of the available repellents.

Certainly, the ability to accumulate, communicate, and apply information paces the progress of our technological society. The United States has been a leader, not only in generation of scientific and technical data, but also in effective application of this knowledge resource. In fact, a distinguishing national characteristic of the United States has been its capacity to organize for effective exploitation of scientific and technological knowledge. This national attribute has contributed to unprecedented gains in gross national product, creation of a powerful defense armada, eradication of many diseases, and has in many ways contributed to an improved quality of American life.

For the past decade, the world scientific and technical community has become increasingly aware of the possibility of substantially improving the systems used to generate, communicate, and apply scientific and technical information. A recurring problem faced by every individual or organization attempting to study these possible improvements is the sheer complexity and amorphous character of scientific and technical data activity. It is difficult to visualize the totality of the subject and to find effective means to define and classify it for discussion or study purposes. Certainly, individuals from different backgrounds or working from different motivations diverge significantly in their identification of the key elements, configurations, and issues concerning scientific and technical data activities.

It is axiomatic that the scientific and technical data systems of the future must be based on a clear understanding of current patterns and practices. However, the complexity of the present scientific and technical data systems is difficult to describe analytically. Projection of a more effective, future system from this ill-defined reference base is obviously even more difficult. The present systems have not emerged as a result of systematic planning. Rather, as data needs arose, individual groups and agencies established their own information and data systems which continue to operate more or less independently.

Recently, many individuals and organizations have concluded that uncoordinated development and operation of scientific and technical data systems should not continue. Such conclusions have been, to a large extent, based on informed judgment rather than findings from formal analyses. It is quite likely that these conclusions would be confirmed by analytical studies, such as cost-benefit analyses. However, the results of such formal analyses methods, with their highly stylized formats, are not necessarily the optimal means for describing problems and opportunities confronting those who desire to improve scientific and technical data management. Consequently, this entire study has not followed the formal conventions of systems analyses. Rather, it has adopted a more descriptive style. This approach is displayed in Part A of Volume II in the form of state-of-the-art write-ups or scenarios covering data activities in ten selected areas of science and technology. The following sub-sections also reflect this approach.

The following sub-sections deal with some of the underlying factors which lead to conduct of this study and formulation of the recommendations and implementation plan presented in subsequent sections of this volume. Hopefully, these brief discussions will enable the reader to identify the perspective which he currently holds, and simultaneously, permit him to orient his perspective relative to the perspectives of others who are concerned with scientific and technical data activities. The discussions in the following sub-sections are not, by any means, considered <u>precepts</u> for scientific and technical data system development efforts. Rather, they constitute <u>recepts</u> which, hopefully, will evoke discussion and responsive actions that will lead to establishment of validated <u>precepts</u> which will provide effective guidance for future scientific and technical data systems.

A. Data--A Vital Resource

The Institute for Basic Standards of the National Bureau of Standards reports that approximately 20 billion measurements are made each day in the United States. Over $25 billion has been invested in instruments to make these measurements, and roughly $10 billion is expended each year on personnel to operate these instruments and related systems.* In light of these figures, it can be said that measurement or data-generating activities constitute a sizable and important segment of our economy.

Table III-A-1, also reported by the National Bureau of Standards,* indicates the costs in money and time of these measurement activities with respect to the national economy:

TABLE III-A-1
RELATIONSHIP OF MEASUREMENT ACTIVITIES TO THE NATIONAL ECONOMY

Economic Sector	Final demand (GNP) ($ billions)	Cost of measurement ($ billions)	Man years spent on measurement (thousands)
Manufacturing	225	7.8	845
Construction, Mining and Farming	21	1.1	120
Transportation, Communication, and Utilities	39	0.9	98
Medical and Educational Services	28	1.4	103
Government and Other Services	83	2.7	139
TOTALS	396	13.9	1305

*Huntoon, R. D., The Measurement System of the United States, Institute for Basic Standards, National Bureau of Standards, Washington, D. C., 1966, NCSL 66, page 89.

These figures are based on 1963 census information; if it were available, current information would no doubt indicate a further growth in level of measurement activities, because our society is becoming more technologically oriented.

The generation, handling, and application of scientific and technical data constitute a principal element in all scientific and technological activity, and also a very expensive one. The total U. S. investment in research and development since World War II totals well over $100 billion. For Fiscal Year 1967, the National Science Foundation estimates that more than $20 billion was spent on research and development, with the Federal Government expenditure in this area totaling 16.5 billion, and the remainder expended in the private sector.* As yet, no reliable estimate has been made as to what percentage of this amount was spent on related data activities. It has been estimated, however, that scientists and engineers spend anywhere from 10% to 20% of their working time acquiring data. A conservative estimate, therefore, of the amount of Federal money being spent for just this one facet of the entire data handling process--that of data gathering--is approximately $3 billion annually.

To date, attempts to obtain precise totals for the costs of scientific and technical data activities have been unsatisfactory because of the inability to separate these costs from those expenditures for other functions involved in the performance of scientific and technical work. However, the National Science Foundation has made an estimate of the Federal expenditures for collection and handling of general-purpose scientific data for the Fiscal Years 1962 through 1968. As the following definition indicates, general-purpose scientific data represents only a fraction of all scientific and technical data; and correspondingly, represents only a fraction of the total Federal expenditure for collection and handling of scientific and technical data.

> "General-purpose scientific data are defined as newly gathered statistics, observations, readings, specimens, and other facts.... from such activities as surveys, field investigations, laboratory analyses, or compilations of operating records which can be applied to useful, general scientific purposes.... This definition thus excludes data gathered from the R & D process or

---

*Federal Funds for Research, Development, and Other Scientific Activities, Fiscal Years 1966, 1967, and 1968, Volume XVI: National Science Foundation, NSF 67-19.

from scientific and technical information activities. Also excluded are data used only for internal operating or administrative purposes."*

The estimated total cost for collection of general-purpose scientific data during Fiscal Year 1968 is $412 million. In the same report, the National Science Foundation reported that since 1962, these expenditures have increased at the same annual rate as total research and development expenditures, i. e., approximately 11%. Table III-A-2 shows these facts:

TABLE III-A-2
FEDERAL OBLIGATIONS FOR COLLECTION OF GENERAL-PURPOSE SCIENTIFIC DATA, FISCAL YEARS 1962-68

Year	Obligations (millions of dollars)	Annual percent change
1962 ------------------------	$220	--
1963 ------------------------	268	22
1964 ------------------------	309	15
1965 ------------------------	343	11
1966 ------------------------	325	-5
1967 ------------------------	369	13
1968 ------------------------	412	12

The above figures give some indication of the costs involved in the generation and handling of scientific and technical data. It should be noted, however, that such costs do not necessarily reflect the current value of the data. For example, the Department of Defense estimates that it annually expends approximately $2 billion to acquire items of scientific and technical data. However, the current value of its investment in 50 million engineering drawings, 225,000 technical manuals, and other items of data is difficult to express, for it represents the vital reservoir of engineering knowledge upon which the continued effectiveness of our defense system depends.

Similarly, it is impossible to assign dollar values to the data which were utilized to discover the key to genetic coding and biological

---

*Federal Funds for Research, Development, and Other Scientific Activities, Fiscal Years 1966, 1967, and 1968, Volume XVI: National Science Foundation, NSF 67-19.

cell reproduction. The value of a specific bit or set of data is not a constant quantity. Much data, unlike other resources, are not consumed by use but are reused many times with their value being augmented by use. Paradoxically, some data, such as weather observations, are a perishable commodity with a short useful life. Data critical to one application may have no application to the needs of another individual, discipline or mission. These attributes of data complicate any attempt to rationally appraise the value of our current scientific and technical data resource.

A given element or set of data often provides information of value only when combined or presented in the context of other data and information. The full utility of our data resource, therefore, can be realized only if it can first be organized and then applied toward some desirable objective. As President Johnson has stated, "the test of our generation will not be the accumulation of knowledge. In that, we have already surpassed all the ages of mankind combined. Our test will be how we apply that knowledge for the betterment of mankind."

Perhaps, scientific and technical data are amenable to appraisal and management as an economic resource. As a commodity, data are packaged and distributed in complex networks highly analogous to some of our more familiar commodity distribution methods. Like the development and use of a tool, the production of data consumes time and money; and its proper use conserves time and money. Data have many of the attributes of a commodity or resource, but frequently their true cost and value are not currently known to those responsible for their management. Only recently has attention been directed to the economic value added to data by organization and evaluation. In one illustrative case, Dr. Herbert Hollomon testified before a Congressional subcommittee that it was anticipated that each dollar spent on producing standard reference data would save the economy 20 to 200 dollars.*

The ever increasing production of scientific and technical data has been described as a healthy sign of progress, demonstrating at the same time an unprecedented wave of scientific, technological and economic achievement. In addition, the concept of scientific and technical data as a national resource has evolved as our generation has observed numerous new applications of scientif[illegible] d technical

---

*<u>A Bill to Provide a Standard Reference Data System. Hearings Before the Subcommittee on Science, Research, and Development</u>, Committee on Science and Astronautics, U. S. House of Representatives, 89th Congress, 2nd Session, June 28-30, 1966, p. 28.

knowledge. In the future, scientific and technical data as an economic resource may become equal in importance to the classic resources of land, materials, labor, and capital. If current assessments of the potential worth of our national resource of scientific and technical data are accurate, several questions must be asked concerning the current attention given to the management of this resource. What are the benefits we should be able to realize from our current and future investment in this resource? Are there effective means and procedures for channeling the use of this resource for public benefit? Does current management of this resource aid, in an optimal manner, the decision-making and communication processes involved in the conduct of research and development, and the technological application of knowledge gained from scientific and technical efforts?

It seems apparent that numerous significant economic and social benefits could be achieved if mechanisms could be developed to better channel scientific and technical data into applications which are of public benefit. It is possible that a more effective utilization of even a small fraction of the existing data resource could increase significantly our rate of economic growth, provide new employment opportunities, enhance the international competitive position of U. S. industry, and add to the quality of American life.

The Committee on Government Operations of the United States Senate has not only provided an accurate description of the current scientific and technical information resource, but has also provided guidance regarding appropriate actions to be taken and results to be anticipated:

> "Information is an agency resource, a Federal, national and international resource. . . .
>
> "Modern information technology has made it possible to place much of the accumulated knowledge of the human race within the reach of a man's fingertips, so to speak. The potentialities of this access to power are awesome, in terms of improving the well-being of our own and other people, as well as in terms of improved education for young and old alike.
>
> "If man's collected knowledge is to become truly accessible, plans and programs must be made, priorities assigned, and resources allocated.

"Savings almost beyond comprehension may be possible--savings in manpower, material, and perhaps most important, in time. The savings will not be automatic; at times, they may even prove illusory because hidden costs develop. But over the long run, the savings will be real and substantial."*

---

*Summary of Activities Toward Interagency Coordination. U. S. Senate Report 369. June 24, 1965.

B. Data Management -- An Essential Part of Science and Technology

That special activity which we call science began as a classification of facts. Scientists have a proclivity to classify facts into patterns, to associate facts with each other and thus understand the connections between them. This characteristic of science is well expressed by de Solla Price:

> "Science seems to have a very special structure all of its own and it is certainly this peculiar structure that has given it the strength, the rapid growth far transcending the rest of scholarship, and perhaps even the utility that makes modern science so valued by society that it holds the strings of economic and political power of nations."*

Scientific communities are beginning to recognize that their continued progress will be materially affected by the effectiveness of their management of factual information or data. They are realizing that progress in science depends in large measure on having efficient formal and informal means of communication for both the exchange of conceptual ideas and of factual information or data. They are also becoming increasingly aware that scientific progress is served by, and is dependent upon maintaining, expanding, and refining the structure and body of scientific data.

R. E. Gibson has put in concise words the vital relationship between the structure of knowledge and its value to the scientific community:

> "'Order,' said Alexander Pope, 'is Heaven's first law.' It is also the essence of modern scientific research and the comprehension of its results. Although the formulation of the grand theories of knowledge is reserved for the occasional man of genius, the way must be paved for him by the codifier and the teacher. I have suggested that the orderly growth of knowledge is threatened by an unbalance between our resources for establishing new facts.... and our resources for the codification and systematic exposition of these facts.... In the field

---

*de Solla Price, Derek J., "Communication in Science: The Ends--Philosophy and Forecast." Ciba Foundation Symposium on Communication in Science: Documentation and Automation, 1967. Anthony de Rueck, Julie Knight, eds., J. and A. Churchill Ltd. London, pp. 199-209.

of codification, indexing, and retrieval of knowledge, a systems engineering type of attack of sufficient magnitude gives promise of readjusting this balance. In the field of systematic exposition, perhaps most important is the key role played by the professor and teacher in the growth of knowledge. Moral and financial support to those whose talents excel in this area, who write books rather than papers will, like the 'seed sown on good ground', be returned, fifty, sixty, or even one hundred fold."*

Figure III-B-1 is structured to display typical stages in the refinement of scientific data. The pyramidal structure in Figure III-B-1 can also be viewed as illustrative of the compression or reduction in volume which accompanies data refinement. A given scientist may be involved in activities at any of the seven levels, but most scientists devote much of their time to activities on the first four tiers of Figure III-B-1. The end results of these efforts are made available principally through publication of the scientific paper.

Historically, the scientific paper has been the principal means of conveying conceptual and factual information, granting accreditation and acknowledgement to the author, and providing a permanent record of the reported findings. Through this vehicle, the scientific community slowly built a foundation of factual information that indicated the progress that had been achieved and identified the leading edge from which new research was initiated. The process of establishing and building this foundation required careful husbandry of the scientific evidence which, in due course, became a part of the structure of the foundation.

As the corpus of factual information in the sciences grew, handbooks and other reference tools began to appear, separating out from scientific treatises the growing volume of evaluated data. With these reference tools at their disposal, the scientists and other users gained access to scientific data that had been checked and rechecked for its validity by their peers. The process of putting together these references of scientific data played a major role in further developing and strengthening the structure of scientific knowledge.

---

*Gibson, R. E., "Impact of Government R & D Programs on the Growth of Knowledge." Presented at the 50th Anniversary Meeting of the Division of Industrial and Engineering Chemistry, American Chemical Society, September 9, 1958.

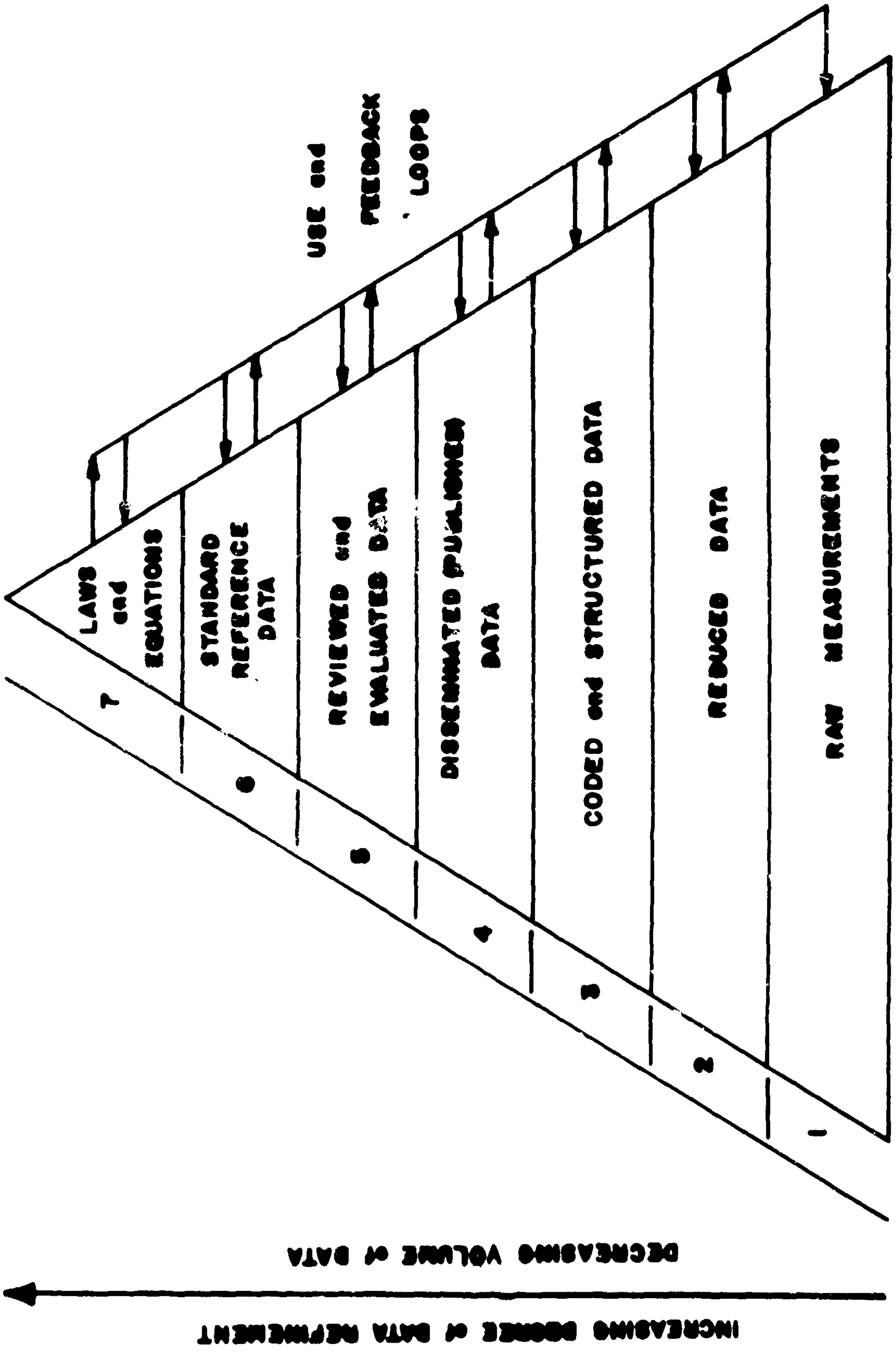


FIGURE III-B-1 STAGES OF REFINEMENT OF SCIENTIFIC DATA

As recently as a half century ago, it was possible to undertake a single comprehensive effort to extract and compile all or nearly all of the useful data stored in the scientific literature. In fact, the International Critical Tables of Numerical Data: Physics, Chemistry and Technology were assembled between 1919 and 1934. Approximately a decade ago, it became evident that somehow a mechanism should be devised for extracting, evaluating, and disseminating numerical property values on a continuing basis for all of the physical sciences; if not, it was feared that the data determined at great expense would be hopelessly lost in the morass of journal literature. Recognition of this situation lead to the establishment of the Office of Critical Tables within the National Academy of Sciences, the National Standard Reference Data Program, and indirectly, the Committee on Data for Science and Technology (CODATA) of the International Council of Scientific Unions. The current efforts of these and other groups, however, continue to be outpaced by the rapid generation and dissemination of scientific data. Perhaps, this occurrence verifies the contention of Gibson, quoted earlier, that the orderly growth of knowledge is threatened by an imbalance between our resources for establishing new facts and our resources for the codification and systematic exposition of these facts.

The systematic interpretation and evaluation of scientific data is a demanding task requiring the highest scientific competence. This is especially true when interpretations and evaluations are attempted on a broader and more involved basis. Figure III-B-2 illustrates several levels at which data can be analyzed. Most of the daily work routine of a typical scientist is devoted to analysis of data at levels 1 and 2; i.e., analyses of isolated bits of data resulting from research measurements or the correlation of such measurements against other similar or related measurements within his discipline. Multidimensional analyses of single-discipline data collections (level 3) are conducted in data evaluation centers such as the Thermophysical Properties Research Center, the NBS Crystal Data Center, and the Electronic Properties Information Center. In addition, the more sophisticated data evaluation centers in highly structured disciplines such as thermodynamics have developed a capability for theoretical projections or simulations of single-discipline data collections (level 4 analyses).

In recent years, it has become increasingly apparent that the performance of multidimensional analyses of data collections covering two or more disciplines (levels 5 and 6) might result in significant

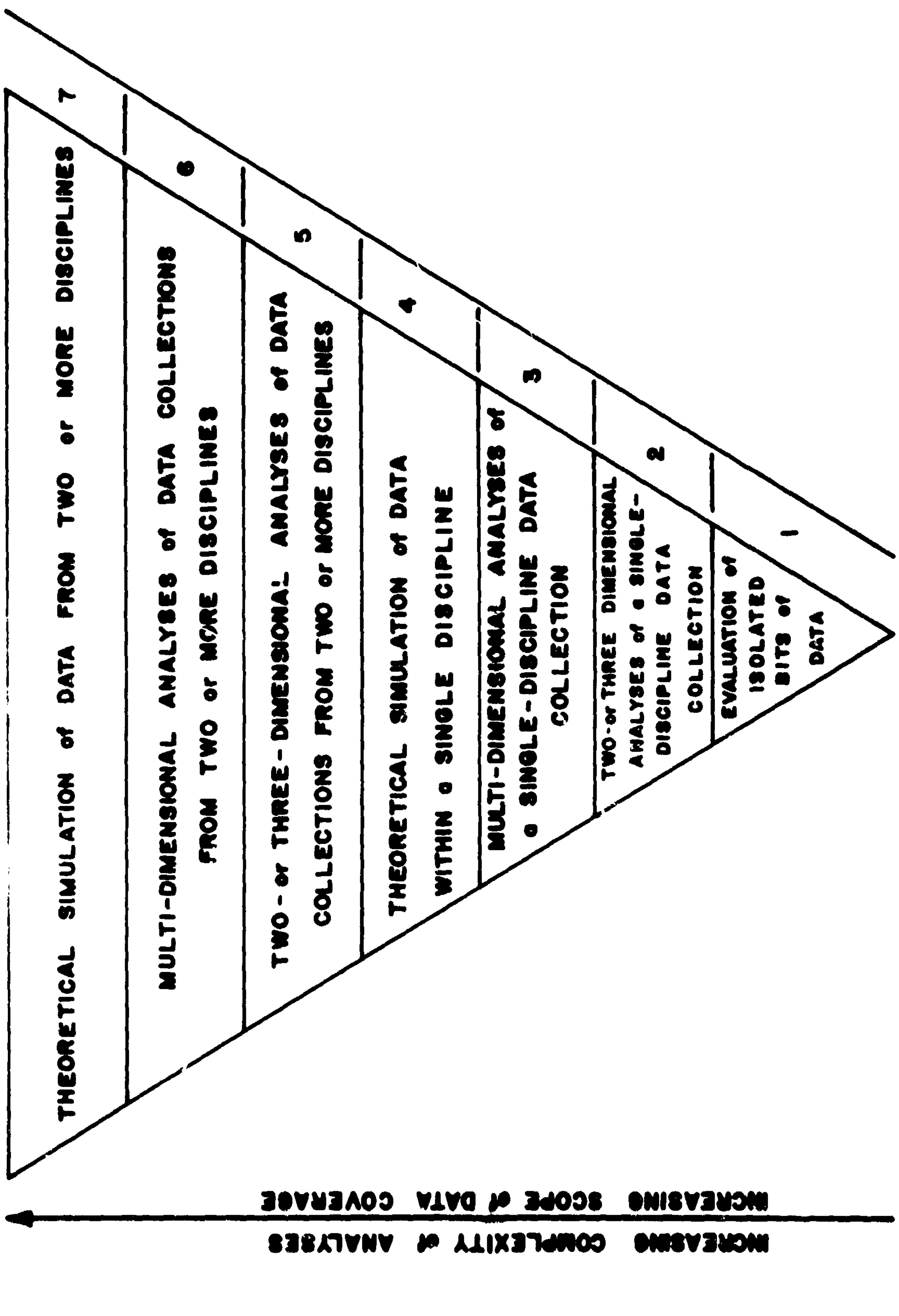


FIGURE III-B-2 LEVELS AT WHICH THE SCIENTIST INTERACTS WITH AND ANALYZES SCIENTIFIC DATA increases in useful scientific knowledge. For example, the Panel on Handling of Toxicological Information of the President's Science Advisory Committee recognized the need for:

> "A comprehensive and exhaustive system for storage and retrieval of valid information on the interaction between chemical substances and biological systems. The existence of this system would perhaps allow the creation of those new broad scientific conceptualizations which will speed the progress of toxicology and pharmacology by quantum jumps."*

The establishment of such a system obviously would be facilitated by the existence of structured and analyzed collections of data in both chemistry and the biomedical sciences. Preparatory steps of this kind would aid the implementation of the Panel's major recommendation that a computer-based system for handling toxicological information be established.

In some fields of health and the social sciences, collections of general-purpose scientific data are available which are multidiscipline in coverage. Such data collections have found use in epidemiological correlations and other higher-level analysis methods. For example, collections of both meteorological data and morbidity data are utilized, together with air pollutant emissions data, by the Public Health Service and other organizations to analyze correlations among those factors controlling the effect of air pollution on human populations.

Although definite benefits can be credited to systematic structuring and analysis of scientific data, it should be recognized that real-world factors do not always make it easy to manage or conduct such structuring efforts. For example, the United Kingdom, Office of Scientific and Technical Information, found it necessary to classify fields as to their amenability to systematic structuring of data activities. The classifications found useful for coordination and management of data programs were as follows:

> "(a) Fields where working scientists recognize the need for organized data and where data activities are well advanced (Crystallographic, nuclear, and thermodynamic data are prime examples).

---

*Handling of Toxicological Information, A Report of the President's Science Advisory Committee, The White House. Washington, D. C. June 1966.

(b) Fields where working scientists recognize the need for organized data but where individual workers are discouraged by the sheer volume of data or of specialized effort required. (Mass spectroscopy is a good example).
(c) Fields where data have been too fragmented or of insufficiently high quality for satisfactory data projects to be started, but where recent instrumental advances have suddenly changed the picture--e. g., cartographical data since the advent of computer-controlled map making machines.
(d) Fields where systematic data activities are still impossible, or where scientific workers are unaware of, or apathetic about, the value of organized data. *

Harvey Brooks says, in his paper which is included in a report by the National Academy of Sciences to the House Committee on Science and Astronautics:

> "Numerous observers have commented upon the differences between the communication systems within science and those within technology. Science has an elaborate system of public documentation with strong sanctions operating on the individual scientist to make full use of and give proper credit for previous work relevant to his own. . . . Within technology the communication pattern tends to be more localized and more confined to organizational channels. **

In a talk delivered to the Institute on the Transfer of Technical Information, Dr. B. W. Adkinson enumerates differences exhibited by foreign information services operated to serve principally either a scientist or a technologist user population. ***

---

* Journal of Chemical Documentation, Vol. 7, No. 1, February 1967, p. 20.

** Applied Science and Technological Progress, p. 38-39, U. S. Government Printing Office, 1967.

*** Adkinson, B. W. "Technical Information in Foreign Countries," Talk delivered to the Institute on the Transfer of Technical Information, Sponsored by the American University--Center for Technology and Administration, Washington, D. C., October 23-25, 1967.

The basic scientist works to compress and refine data; whereas the technologist seeks the most expedient path to usable data. Studies of the information needs of the technologist have revealed that he wants factual information or data which can be used to answer specific questions. In this regard, de Solla Price states:

> "The answers to such questions are to be found not in an archive of knitted papers but in a data bank. This is not the stuff that is packed down and made into textbooks and monographs; it is more like the tables of refractive indexes and specific gravities that scientists expect to find in their encyclopaedic handbooks. Unfortunately it requires enormous original intellectual effort to devise a system for ordering such data. One has to know a lot of botany to be a Linnaeus; a successful Chemical Registry scheme is worth a Nobel Prize in Chemistry "*

The technologist differs from the scientist in that his interactions with data are much less predictable. To the technologist, data are most useful when they are packaged in an instructional or other form tailored to his specific use. The technologist relates data to a problem rather than to a niche in a knowledge structure. Resulting from these differences, the volume of specialized artifacts generated to communicate data among technologists far exceeds the volume of journal articles, reference handbooks, and other artifacts used to maintain and communicate the store of scientific data.

Since the technologist maintains a high level of association between data and its application to the solution of specific problems, the data which he generates is packaged in forms (engineering drawings, test reports, standards, specifications, etc.) tailored for use in a specific situation. Consequently, it is frequently difficult to identify data relevant to the solution of a similar problem in another technological context. Also, when relevant data is identified, it frequently must be extracted and reformatted for its application to the second

---

*de Solla Price, Derek J. "Communication in Science: The Ends--Philosophy and Forecast." Reprinted from Ciba Foundation Symposium on Communication in Science: Documentation and Automation, 1967, p. 209. (Edited by Anthony de Reuck and Julie Knight. Published by J. &. A. Churchill Ltd., London.)

problem. As a consequence, the work patterns of the technologist, to date, have not created a well structured data resource. This has made data management activities in technology difficult and costly. The current high costs of data management in technology are tolerated because the decisions (e.g., selection of a design for an Apollo rocket motor) based on a technologist's accumulation and evaluation of diverse data are frequently the prime determinants of the success or failure of expensive projects.

As science and technology have interacted to strengthen and extend the capabilities of one another, their value to society has been increasingly recognized. The increasing breadth and complexity of missions assigned science and technology have required new organizational concepts involving large teams of specialists each contributing knowledge and perspectives of specific disciplines. As a consequence, significant changes occurred in the work environment and work functions of the individual scientist or technologist. For example, the process of research and development is now frequently conducted as a group endeavor, rather than an individual undertaking. The application of the scientist's classical information ordering and communicating procedures to this new environment does not constitute an optimal match.

The inadequacy of classical methods for structuring and communicating scientific and technical knowledge in current working contexts has created intellectual and managerial concern. The more evident symptoms of this concern include fears, such as those voiced by the Weinberg Report,* that science could lose its unity and effectiveness by fragmenting into a mass of repetitious findings, or worse, into conflicting specialties that are not recognized as being mutually inconsistent. Another concern voiced was that our society could fail to obtain a fair return on its massive investments in science. In regard to technology, Professor James Brian Quinn went directly to the root-cause of the concern when he stated:

> "Technology is knowledge... systematically applied to useful purposes... Thus when we talk in terms of policies

---

*Science, Government, and Information, A Report of the President's Science Advisory Committee, The White House, U. S. Government Printing Office, January 10, 1963.

> for technological development, we must think in terms of policies for an essentially intellectual process. This is something of a new concept for national policy makers and for all managers. The concept should include policies to stimulate both (1) the creation of knowledge for practical purposes, and (2) the use and transfer of knowledge for practical purposes. *

The challenging problem now facing both scientific and technological communities is how to maintain the ever increasing pool of scientific and technical data in a useful form. The handbooks and reference tools that once performed these functions in a satisfactory manner now have difficulty in accomplishing their task due to the increased scientific and technological activity that daily pours out new data. By the time these reference tools have gone through the rigorous editorial and evaluation processes required to assure accuracy and validity, they are, in part, made obsolete by new findings.

Computers have helped in the performance of evaluation and structuring processes by making computations that validated theories and gave credence to previously unevaluated data. The challenge, however, remains as to how to construct new and effective methods of collecting, structuring, storing, and retrieving scientific and technical data. Specifically, the challenge and the opportunity is that of actively investigating how new data handling systems can perform not only the function of storing and recording the corpus of scientific and technological data, but how they can materially assist in the management effort required to reinforce and extend the peculiar structure that constitutes the core of scientific and technical knowledge.

One of the chief aims of scientific data management should be to strengthen and accelerate the development of well structured data bases. This is as true for the softer sciences (e.g., social and behavioral sciences) which have less well defined structures, as it is for the more formally structured sciences, such as chemistry and physics. Automated data handling systems represent the potential and, in some cases, the actual means to maintain the increasing volume of data in science and technology in a viable form.

---

*Professor James Brian Quinn, "The Impact of the Policies of Government on the Creation and Use of Technology for Economic Growth," in Proceedi[illegible] Symposium: Technology and World Trade, U.S. Depar[illegible] Commerce, National Bureau of Standards, Novembe[illegible] 1966, p. 98.

In many ways, they offer a more effective medium by which to store and retrieve scientific data than do the journal article and related conventional means of data handling. Recognizing the potential of such tools to better perform these vital functions opens up the possibilities of data banks functioning as catalytic systems expediting and improving not only the organization and evaluation of data but also its communication among scientists and/or technologists.

C Data Handling Systems -- A Sub-Set of Information Handling Tools for Data Management Applications

1. Interrelations Between Document Handling Systems and Data Handling Systems

The changing concept of data management was discussed in the previous section. This section discusses the types of data handling systems which have emerged to date to facilitate data management. These systems, like data itself, have developed primarily as specialized elements within larger systems for the recording, archiving, evaluation, retrieval and dissemination of scientific or technological information. Therefore, current data handling systems are closely related institutionally, administratively, and conceptually to document and other information handling systems. They also have similar data acquisition approaches and sources. However, the management needs and automation amenabilities of data have caused data handling systems to assume distinct characteristics and potentials, demanding that they be viewed and managed as increasingly discrete sub-systems within the larger framework of scientific information management and handling.

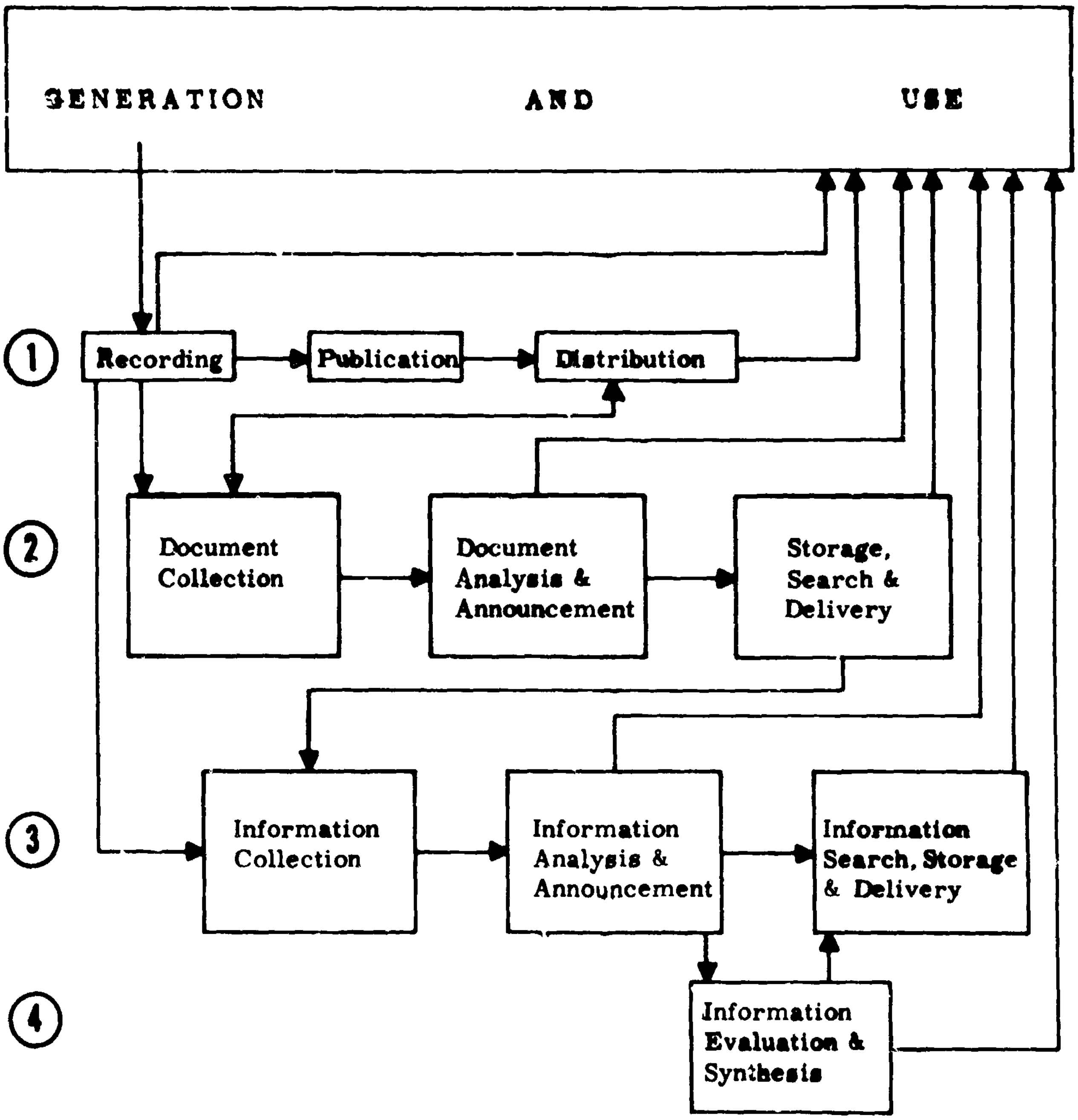
Numerous attempts have been made to depict the flow of scientific and technical information, and systems which facilitate the flow, within variously defined fields of scientific and technical activity. One such portrayal, that developed for the flow of biomedical information by R. H. Orr, et al.; * is a useful one for discussion of the place of data handling systems in the overall information handling and transfer scheme. The principal elements of Orr's model are reproduced in Figure III-C-1. Levels 1 and 2 portray information flow through the journal, book, and technical report literature. Levels 3 and 4 present the flow and analysis of information (data) itself.

Data handling is a sub-system pervasive throughout this entire information flow pattern. Formal data efforts identified in the census (Volume II, Part C of this study) fit quite comfortably into this scheme: Data Collection Networks and Data Publishing Programs are a sub-system of the first level; Data-Document Depositories constitute a sub-element of the second

* Orr, Richard H., et al., Communication Problems in Biomedical Research Report of a Study, reprinted from Federation Proceedings, Volume 23, September-October 1964, pp. 1146-1154.

FIGURE III-C-1

CLASSICAL FLOW OF INFORMATION*



*Adapted from "The Biomedical Information Complex Viewed as a System," Orr, Richard H., et al., in Communication Problems in Biomedical Research: Report of a Study, reprinted from Federation Proceedings, Volume 23, September-October 1964, pp. 1146-1154.

level, and Data Service Centers operate on levels 3 and 4. In fact, it is readily apparent that some of the classes of data efforts included in the census are specialized document processing operations, i. e., operations which process documents with a high content of data.

A basic premise of Orr's schematic depiction of information flow is that.

> "In general, information processing starts where document processing leaves off and depends upon prior accomplishment of the basic operations of document processing. (Documents must be collected, analyzed, sorted, and retrieved before the information they contain can be processed.)*

In other words, the collection, analysis, and distribution for use of information (data), per se, is viewed as a selective, refined analysis and redissemination of information already recorded in the journal, book or technical report literature. Although this concept of the relationship between document handling systems and information or data processing is widely held among documentalists, the effectiveness with which a document handling system feeds or interacts with an information or data processing operation is seldom considered either as a design criterion or as a measure of operational effectiveness. This situation is readily apparent when one considers the dearth of indexing of data content currently performed as a part of document handling system operations.

Formal data efforts, which organize and evaluate previously published data for use by others, already constitute an established sub-system of the larger scientific and technical information handling system. However, direct flow from recording to transfer and use, by-passing formal publication, is a feature of much data flow. In fact, it is striking to note the vast amount of data flow which is automatically excluded from the picture when a classical template of information flow, such as that in Figure III-C-1, is laid on top of real-world patterns of scientific and technological activity. This exclusion often occurs when plans are developed for implementing national scientific information

---

*Orr, R. H., et al., "The Biomedical Information Complex Viewed as a System," Communication Problems in Biomedical Research: Report of a Study, in Federation Proceedings, Volume 23, No. 5, September-October 1964, p. 1141.

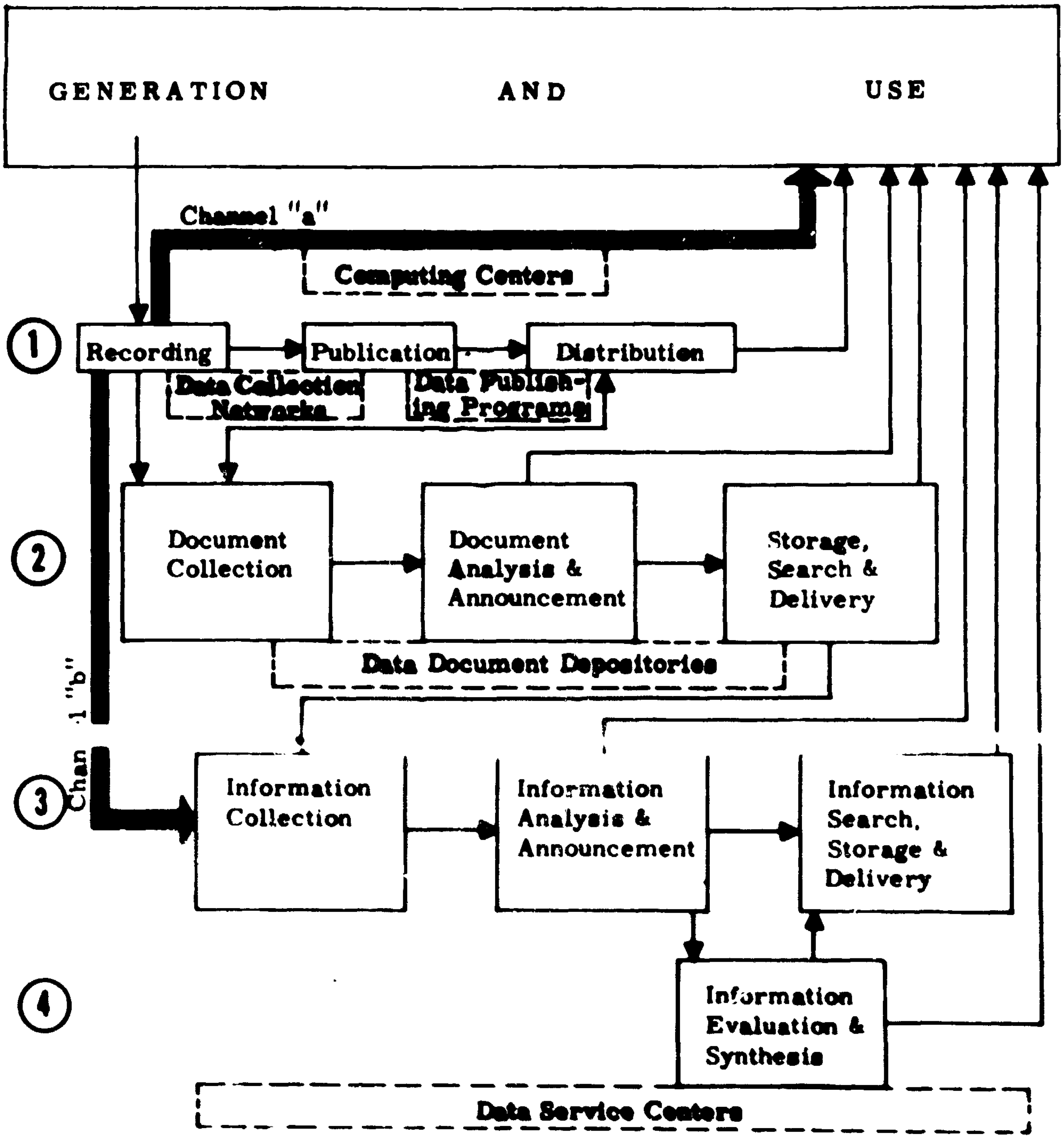
systems or when more limited means are considered for improvement of the use of scientific and technical information. Frequently, it is not realized that data management needs have spawned data handling operations and flow patterns which are significantly independent of the classic publication channels and associated information or document handling systems. Principal factors contributing to this trend are the large volume of data, its high amenability to computer processing, and the high incidence of situations where data is generated and applied within a limited sphere of scientific or technological activity.

The data flow depicted by channel "a" in Figure III-C-2 is a very significant one when the data generator and user are closely associated. In fact, it is becoming much more common in all contexts because current publication practices often preclude the publication of much data even when the generator desires to publish. Channel "b" depicts a data flow pattern which is of considerable significance in an area which has been designated by the National Science Foundation as general purpose scientific data. This data is not normally generated as part of a research or development project, rather it is collected as a support function for activities which may be either operations or research and development. Examples of this class of data are weather observations and demographic information.

Data volume and other factors have fostered the flow of data along channel "a" in Figure III-C-2 where the computer assumes prominence as an intermediary processor of data. A measure of the prominence of this intermediary is found in the broad statistic that there are 40,000 computer centers operating today, with many of the largest and fastest computers dedicated to scientific data processing. The volume and cost of data flowing into and out of these working data processing centers dwarfs that handled by formal data service centers oriented to the archiving and organization of data and information, as indicated by levels 2, 3 and 4 in the schematic of Figure III-C-2. This data flow sub-system, channel "a", is often submerged in the context of a given mission and is frequently excluded from examination of scientific and technical information flow, and yet, it is absolutely vital to efficient real world data management. The foregoing is not to imply that data communication from one worker to another via the classical publication channels is not important but to emphasize that the non-classical data flow sub-systems have become highly active, specialized and sophisticated.

FIGURE III-C-2

DATA FLOW AND PROCESSING SUPERIMPOSED ON CLASSICAL INFORMATION FLOW



Already the working difficulties experienced between the scientist or technologist and his supporting data processors have become significant. As the body of data used grows along with the sophistication of its manipulation and application, these difficulties will become increasingly acute Management of this problem requires the strengthening and acceleration of the development of well-structured data bases with sufficient technical validity to allow the technologist and his data processor to be able to draw upon well conceived data resources. Tremendous losses in time and money are now experienced due to the necessity for each mission group to develop independently its own data files or bank and its own processing, storage and retrieval approaches.

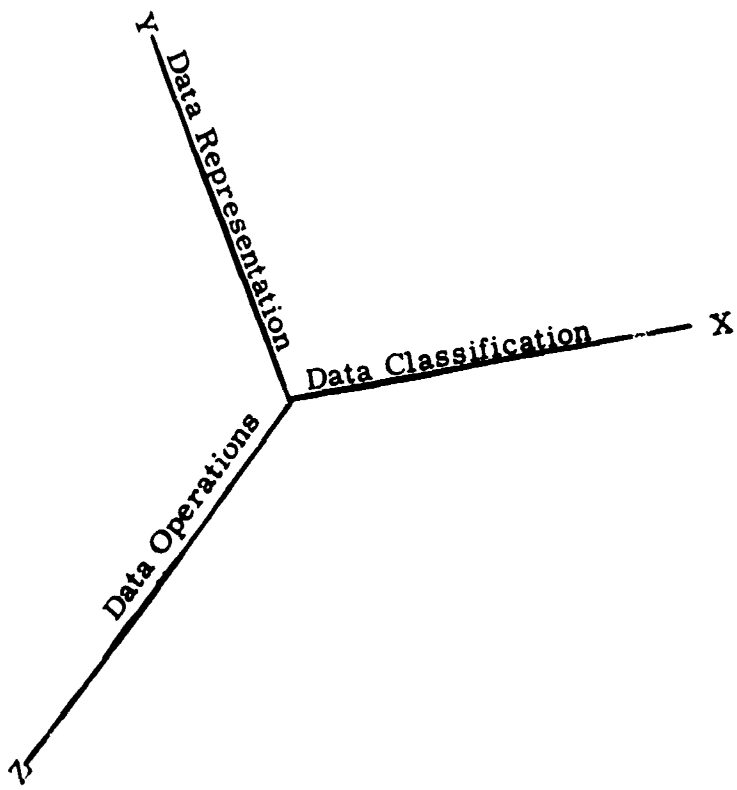
2. Elements of Data Handling Systems

Data handling systems are rightfully considered as tools to facilitate data management. However, one attempting to select, design or implement a data handling system to facilitate achievement of a specific data management objective or goal quickly realizes that it is not a simple, straight-forward process. Rigorous and systematic procedures for this selection process have not been demonstrated. However, in their presentation "Introduction, Definitions, and the Information Universe,"* Vincent and Weik put forth a useful conceptual scheme for delineating the constituent elements of information handling systems for descriptive or comparative purposes. One can use this scheme to illustrate the many elements which can be configured into data handling systems currently or are potentially applicable to future data management requirements. If one also considers additional factors not included in the Vincent and Weik model, such as performance criteria, the task of analyzing a data handling system becomes increasingly complex.

In the modified Vincent and Weik scheme, all data handling systems can be described through the specification of various elements arrayed

* Vincent, Col. Dale L., and Weik, Martin H., "Introduction, Definitions, and the Information Universe," paper presented at the Fundamentals of Information Retrieval Systems and Techniques Seminar, American Management Association, New York, New York, 15 January 1968.

upon three coordinate axes:

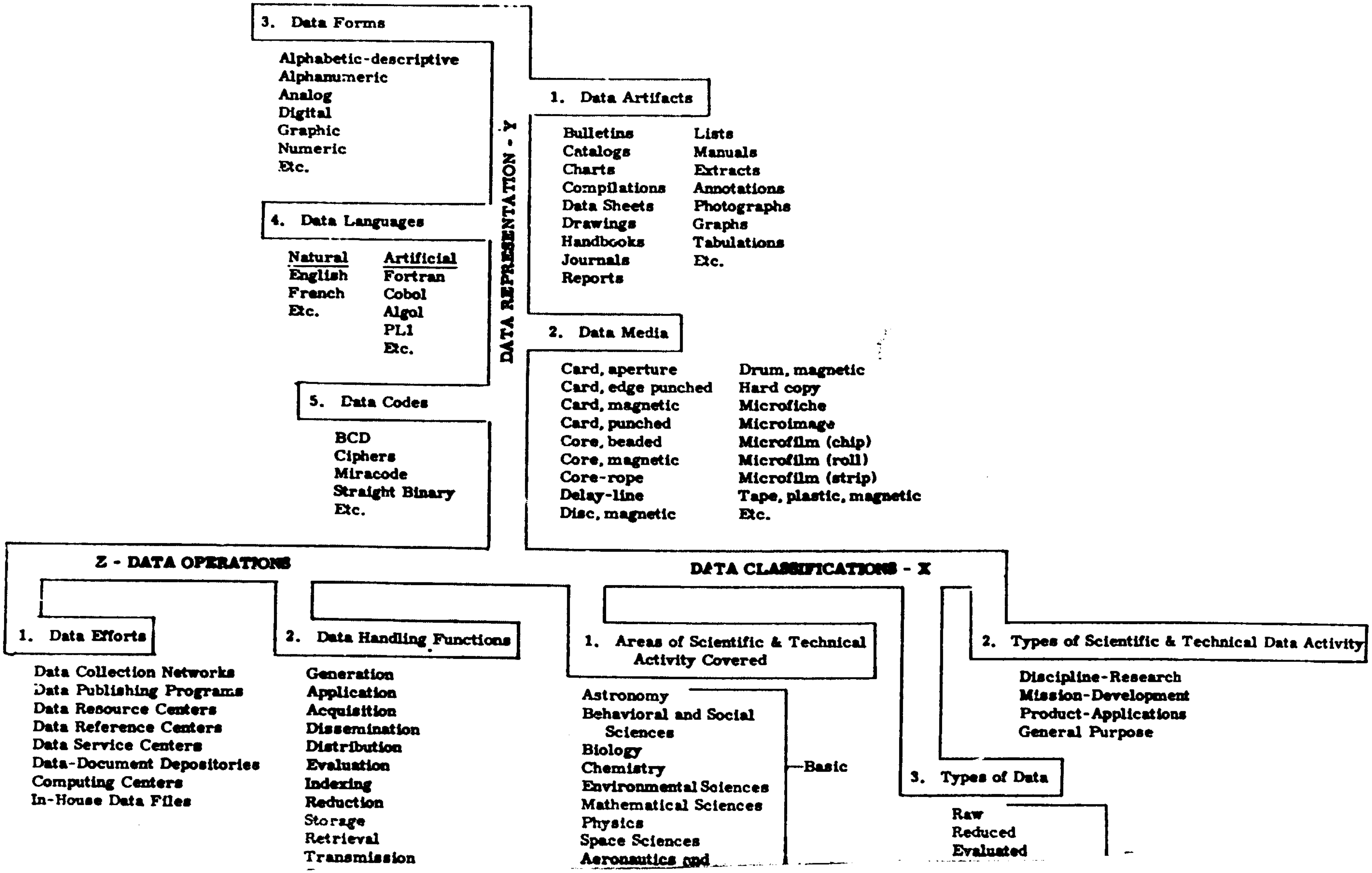


The coordinates or elements of the X axis, Data Classification, constitute description or identification of the data type, class or category included in or handled by the system. The coordinates or elements of the Y axis, Data Representation, describe the data media, forms, formats and languages used by and included in the system. The coordinates or elements of the Z axis, Data Operations, describe operations performed on and/or use made of the data.

Each of these main axes may be sub-divided into numerous other orthogonal coordinate axes. Figure III-C-3 displays a partial array around these three main axes of the multitude of system elements or components currently used or potentially available for inclusion in the development of data handling systems. Application of this scheme to an existing data effort illustrates its utility for description of the structure of current data handling systems. For example, the Joint Army-Navy-Air Force Chemical Data Center at the Dow Chemical Company might be characterized, in terms of its system elements,

FIGURE III-C-3

ELEMENTS OF DATA HANDLING SYSTEMS



Z - DATA OPERATIONS

BCD
Ciphers
Miracode
Straight Binary
Etc.

1. Data Efforts

Data Collection Networks
Data Publishing Programs
Data Resource Centers
Data Reference Centers
Data Service Centers
Data-Document Depositories
Computing Centers
In-House Data Files

2. Data Handling Functions

Generation
Application
Acquisition
Dissemination
Distribution
Evaluation
Indexing
Reduction
Storage
Retrieval
Transmission
Computation
Location
Summarization
Etc.

DATA CLASSIFICATIONS - X

Card, punched	Microimage
Core, beaded	Microfilm (chip)
Core, magnetic	Microfilm (roll)
Core-rope	Microfilm (strip)
Delay-line	Tape, plastic, magnetic
Disc, magnetic	Etc.

1. Areas of Scientific & Technical Activity Covered

Basic:
Astronomy
Behavioral and Social Sciences
Biology
Chemistry
Environmental Sciences
Mathematical Sciences
Physics
Space Sciences
Aeronautics and Aerospace

Development:
Architectural Engineering
Chemical Engineering
Civil Engineering
Construction Technology
Electrical Engineering
Electronics
Energy Conversion
Materials Engineering
Mechanical Engineering
Nuclear Technology
Oceanography
Ordnance

Applications:
Pharmacology
Agriculture
Conservation
Food Technologies
Medicine
Machine Technology
Military Sciences
Transportation

2. Types of Scientific & Technical Data Activity

Discipline-Research
Mission-Development
Product-Applications
General Purpose

3. Types of Data

Basic:
Raw
Reduced
Evaluated
Reference
Equation of State

Development:
Design
Form and Fit
Maintainability
Performance
Reliability
Test

Applications:
Maintenance
Operating
Personnel
Product
Provisioning
Quality Assurance
Training

Etc.

B

as follows:

- X-1 - Aeronautics and Aerospace, Chemistry, Chemical Engineering, Energy Conversion, Ordnance
- X-2 - Mission-Developmental
- X-3 - Evaluated, Performance
- Y-1 - Compilations, Data Sheets, Extracts, Annotations
- Y-2 - Punched cards, Magnetic discs, Hardcopy, Magnetic tape
- Y-3 - Alphanumeric, Digital
- Y-4 - Fortran
- Y-5 - BCD
- Z-1 - Data Service Center
- Z-2 - Acquisition, Dissemination, Evaluation, Extraction, Storage, Retrieval, Computation.

Perhaps, as insight is gained concerning the structure and functions of data handling systems, scientists and technologists will be able to apply them more fully to accelerate development and use of scientific and technical knowledge. Shipman was concerned about past failures to fully employ computers in data processing functions other than mathematical calculations when he stated the almost obvious fact that:

> "Of the three principal categories of information involved in scientific and technical information--data, procedures and methods, and conceptual framework, theories and ideas--only the first, data, are presently readily responsive to available machine storage and retrieval."*

In contrast, de Solla Price was looking toward the future of data management and data handling systems when he stated:

> "In a forecast for the reasonable future one may therefore expect an approach to computer success with the 'taxonomicisable' data banks which are the chief problem for the technologist."**

---

* Shipman, Joseph, "The Mounting Crisis in Primary Literature." Engineering Societies and Their Literature Programs: Proceedings of a Critical Appraisal, Engineers Joint Council, New York, January 17-18, 1967, p. 28.

** de Solla Price, Derek J., "Communication in Science: The Ends--Philosophy and Forecast," reprinted from Ciba Foundation Symposium on Communication in Science: Documentation and Automation, London, 1967, p. 206.

It should be noted that preceding this forecast, de Solla Price explicitly described the structuring or data organization schemes which had to be developed before the full potential of computer-maintained encyclopedias of data could be developed. It is this knowledge structure which must be developed in parallel with data handling structures if data management is to progress from an intuitive art to a science. The importance of development of valid classification schemes or structures for scientific and technical data organization cannot be overemphasized, for the progress in this area will be the major technical determinant as to how effectively future data management and data handling systems will function.

D. Data Management and Data Handling Systems-- The Challenge of the Future

Lewis Mumford, in his book Technics and Civilization, states that "Behind all the great material inventions of the last century and a half was not merely a long internal development of technics: there was also a change of mind. Before the new industrial processes could take hold on a great scale, a reorientation of wishes, habits, ideas, and goals was necessary."

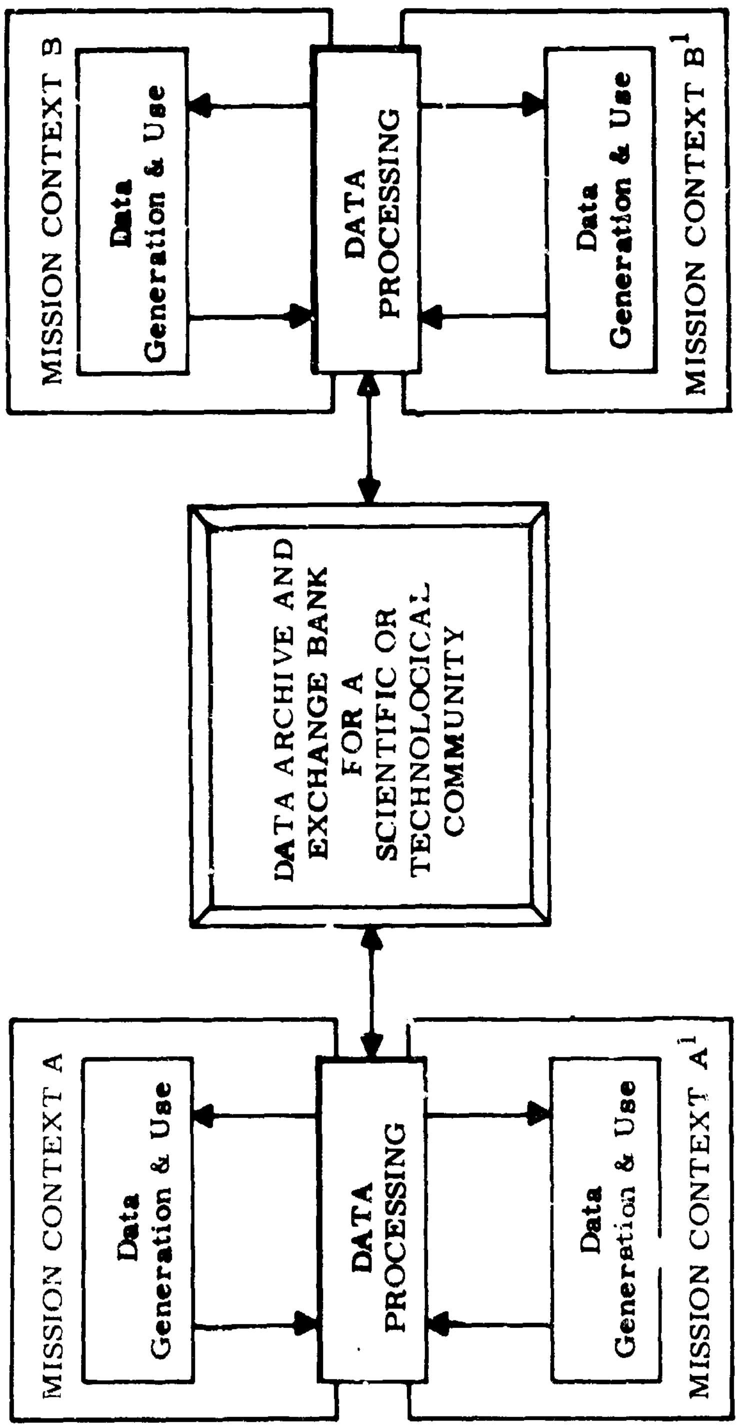
The development of data management and handling systems for the future will also require behind it dual forces of intellectual reorientation and technological advancement. Basic to data handling systems of the future will be computer technology. Utilization of its full potential challenges data management to reorient its viewpoint and technical capacity from that of document handling (i.e., manipulating and delivering documents which contain information) to one of scientific and technical data handling (i.e., manipulating and delivering the data or information itself). Implicit in such a reorientation is recognition of the potential of the computer to be (in the words of Marshall McLuhan) "an extension of the human central nervous system." In terms of the discussion of the previous section on data handling systems, this extension could constitute an immediately ready, accessible, and usable data bank and processing capability for the working scientist or technologist in conjunction with other working scientists. Figure III-D-1 portrays how the data processor or computer serving a given mission project or group might be connected to a common data-storage computer serving the archival and inter-mission communication requirement of several mission-oriented groups within related areas of science or technology. Such applications would extend computer usage significantly beyond its current usage, which is largely for iterative mathematical calculations.

This potential--one of coupling the creative capabilities of humans with each other and with the immense procedural, computational, and retrieval capacities of computers--is well expressed by the developers of such an "one-line intellectual community" pilot system at the Massachusetts Institute of Technology:

> "In short, it is now evident that much of the creative intellectual process involves moment-by-moment interplay

FIGURE III-D-1

AUTOMATED DATA HANDLING AS PROJECTED FOR FUTURE SCIENCE AND TECHNOLOGY



between heuristic guidance and execution of procedures, between what men do best and what computers do best. On the basis of that realization, it seems reasonable to project to a time when men who work mainly with their brains and whose products are mainly of information will think and study and investigate in direct and intimate interaction with extensively programmed computers and voluminous information bases. . . . The prospect is that, when several or many people work together within the context of an on-line, interactive community computer network, the superior facilities of that network for expressing ideas, preserving facts, modeling processes, and bringing two or more people together in close interaction with the same information and the same behavior--those superior facilities will so foster the growth and integration of knowledge that the incidence of major achievements will be markedly increased."*

The forces necessitating reorientation toward the development of an on-line intellectual community are not confined to technological requirements of the space program or the nation's defense. Martin Shubik of Yale University discusses the need simply as a consequence of the growing disparity between the total pool of knowledge and the amount of it an individual can afford to assimilate:

". . . Man lives in an environment about which his information is highly incomplete. Not only does he not know how to evaluate many of the alternatives facing him, he is not even aware of a considerable percentage of them. His perceptions are relatively limited; his powers of calculation and accuracy are less than those of a computer in many situations; his searching, data processing, and memory capacities are erratic. As the speed of transmission of stimuli and the volume of new stimuli increase, the limitations of the in-

---

*INTREX: Report of a Planning Conference on Information Transfer Experiments, eds. Carl F. J. Overhage and R. Joyce Harman, The M.I.T. Press, Cambridge, Massachusetts (1965), p. 26.

dividual become more marked relative to society as a whole. *Per se* there is no indication that individual genius or perceptions have changed in an important manner for better or worse in the last few centuries, but the numbers of humans, the size of the body of knowledge, and the complexity of society have grown larger by orders of magnitude. . . . "*

Daniel Bell of Columbia University, commenting in the same report, suggests that in response to this disparity the character of technology itself is shifting to exploit our growing resources of knowledge. He foresees our society becoming "more functionally organized, geared to knowledge and the mastery of complex bodies of learning."

> "Technology is not simply a machine," according to Bell, "but a systematic, disciplined approach to objectives, using a calculus of precision and measurement and a concept of system that are quite at variance with traditional and customary religious, aesthetic, and intuitive modes. Instead of a machine technology, we will have, increasingly, an '*intellectual technology*,' in which such techniques as simulation, model construction, linear programming, and operations research will be hitched to the computers and will become the new tools of decision-making."**

Thus, already, we see developing the required reorientation of "wishes, habits, ideas and goals" toward recognition and exploitation of the full potential of computers as an extension of man's central nervous system. The other requirement for the development of sophisticated national data-handling systems--that of technological capacity--is rapidly developing if not actually available today. Fortunately, the more than $20 billion expended in the U.S. on research and development during the past twenty-five years has not only produced a tremendous scientific and technical data resource; it has also developed the technology required to manage and handle this data resource.

---

*Shubik, Martin, "Information, Rationality, and Free Choice in a Future Democratic Society," *Daedalus* (Summer, 1967 - "Toward the Year 2000: Work in Progress"), p. 772.

** Bell, Daniel, "The Year 2000 - The Trajectory of an Idea," *Daedalus* (Summer, 1967 - "Toward the Year 2000: Work in Progress"), p. 643.

Figure III-D-2 illustrates the fantastic advance of computer capabilities over the past thirteen years, and projects the expected advances through 1975. Similar advances are being achieved in other areas of technology applicable to the handling of scientific and technical data. However, in addition to advanced hardware technologies, a third element will be required for the development of large, computerized data bases - that is, the tremendous labor of structuring, organizing, evaluating, and entering the data. Dr. Jean Weston addressed this element in a presentation concerning the future of drug information processing:

> "There have been a plethora of grandiose speculations on how wonderful the future will be in the field of biomedical information once the computer is properly harnessed and grinding away 24 hours a day. . . . All such pictures so painted--and fully capable of realization I might add--are such comforting and satisfying ones that I'm afraid there has been all too little practical contemplation or discussion of the massive amount of pure drudgery which must be carried out in the way of providing unified and mutually understandable and agreed upon information to go into the computer for its digestive and correlative activities to bring about this bright new day, especially on the part of some of the more starry-eyed speculators. Until the right kind of information is provided to the computer we forfeit its fantastic ability to deal with that information."*

While "drudgery" perhaps aptly applies to much of the labor required to develop large, automated data bases, the challenge appears exciting when the goal is viewed as a thorough taxonomy of an area of science or technology. De Solla Price emphasizes that "success here depends on the tight internal structure of science. . . . The prospects for any non-taxonomic indexing seem so gloomy that such automated Nirvana is unlikely to work for science." He suggests that "all further efforts to secure more sophistication in normal indexing be cut off forthwith." Further, he states, "it requires enormous intellectual effort to devise a

*Weston, Jean R., M. D., "The State of the Art and Its Future," keynote address presented at the Third Annual Meeting of the Drug Information Association, Philadelphia, Pa., May 24, 25, and 26, 1967, pp. 8-9.

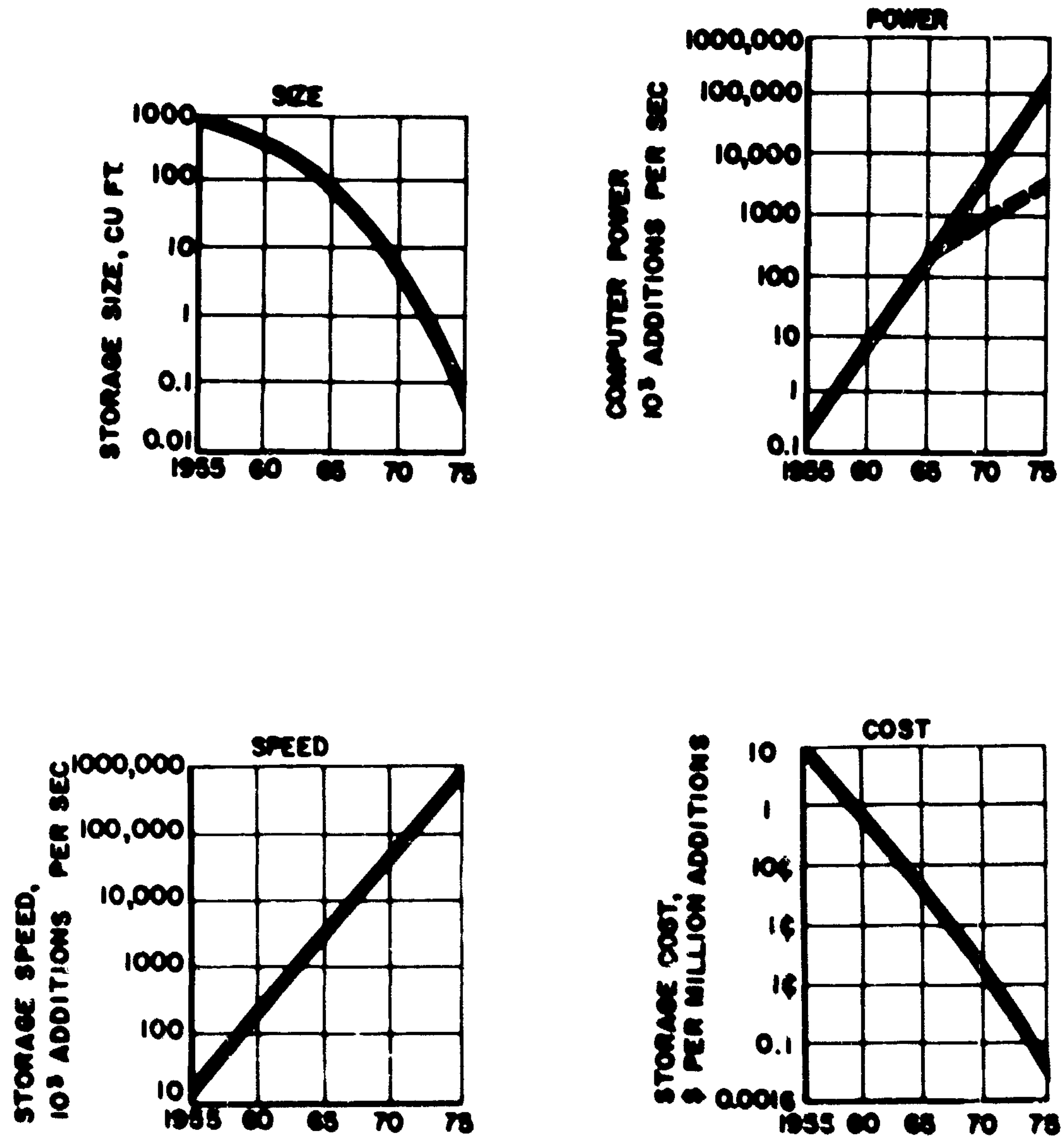


FIGURE III-D-2. GROWTH OF COMPUTER CAPABILITY IN THE U.S.

Source: Boehm, Barry W., "Keeping the Upper Hand in the Man-Computer Partnership," Astronautics & Aeronautics, April 1967, p 25 system for ordering such data. One has to know a lot of botany to be a Linnaeus; a successful Chemical Registry is worth a Nobel Prize in Chemistry."* The combination of this intellectual effort with fully exploited computer technology is the challenge of the future for data handling systems.

Many experts have found this challenge exciting. Commenting on the implications for the social sciences, Harold D. Lasswell of Yale University stated in the Saturday Review that "the computer revolution has suddenly removed age-old limitations on the processing of information, including the linkage of data with competing theories of explanation."** A graphic example of this limitation and how it has been removed by computers can be seen in the complexity of making a global weather forecast. Two leading scientists of the National Center for Atmospheric Research report that "it takes about a billion elementary numerical operations to compute a 24-hour weather forecast for the whole of the earth." They then go on to say that "today, we have computers capable of holding millions of pieces of data in readily accessible storage and of carrying over a million numerical operations per second. Thus, the time required to perform the billion operations for a 24-hour forecast is reduced to about 1,000 seconds or 17 minutes."*** The time dimension, until recently an insurmountable barrier to this application of our scientific knowledge of the atmosphere, has been reduced to a workable level.

Recently, another important prospect has come into view. J. C. R. Licklider, a leading information scientist, states that "during the last five years, man-computer interaction--'on-line' or 'interactive' information processing--has grown from a gleam in an eye to a major

---

*de Solla Price, Derek J., "Communication in Science: The Ends--Philosophy and Forecast," Ciba Foundation Symposium on Communication in Science: Documentation and Automation, eds. Anthony de Reuck and Julie Knight, J. &A. Churchill Ltd., London (1967), p. 209.

**Lasswell, Harold D., "Do We Need Social Observatories?" Saturday Review (August 5, 1967), pp. 49-52.

***Thompson, Philip D. and Roberts, Walter O., "Computing the Weather," The Christian Science Monitor (August 17, 1967).

sociotechnological movement."* The computer and its associated equipment is now providing man a tool that can perform accurately, rapidly, and (when sufficiently large volumes are involved) very economically the clerical operations of data comparison, logic procedures, and mathematical computation. These are operations a knowledge-user almost always must perform in translating resource data into terms significant to the context of his individual interest. Where the rigor of the resource knowledge permits such translations to be made by mathematical or similar logical processing, he can then use the computer so that it can be driven from his interest context and provide a printout in his context. He then has created a greatly enhanced facility for examining the data resource itself. With computer speeds what they are, the computer system's responsiveness may be rapid enough that the interrogator is, in effect, in a position to establish a computer-assisted "dialogue" with his data resource. In such a reactive relationship, he can pursue solutions to his problem based on the computer's response to each request for additional data or action taken. Each step in the process of finding the solution determines the next step until the desired goal is reached, or the problem-solving potential of the data resource is exhausted.

Many of the prospects for improved use of data assume the availability of large volumes of machine-readable data on which the computer can perform the desired functions. Although the number of automated data centers in the United States has increased from a small handful ten years ago to literally hundreds of formally organized centers today, an overwhelming percentage of potentially useful data is still inaccessible for extensive manipulation by computer.

With the development of these data centers as separate entities, pressure is building up to take action toward tying individual centers or systems together into a network or series of networks. Several plans have already been fostered to establish networks of this kind. In the proposed network link-ups, a technical problem exists as a result of different types of computers and their different programming instructions. These, obviously, must be compatible before data could be effectively transferred from one data center to another.

---

*Licklider, J. C. R., "A Crux in Scientific and Technical Communication," American Psychologist (November 1966), p. 1049.

The most critical problem, however, in creating working competencies in the computer-assisted mode is much deeper than establishing electronic transmission links between different data centers. There is the problem, even without an electronic link, of intellectual compatibility between the systems. For instance, are they using compatible data structures and languages? Not only does this mean that terms must have essentially the same or translatable meanings in each system--their data files must also have a high degree of compatibility in their conceptual structures. Structuring in many cases renders data useless unless it can be reformatted to meet the requirements of a different context. Computer codification languages and file structures impacting on the user must be modeled on the conceptual structures of the subject matter first and on the machine capability second. Codes designed to be easy for the machine can impose learning burdens the human user will not accept, even though the codes are both simple and rational. The perceived mechanics of the phenomena, and identified values in human affairs will constitute the most reassuring foundation for enduring codification commitments. The data specialist and the lexicographer must encourage the computer technologist to be faithful to the "language of origin" in pursuing his craft.

Solution of the problems concerning standardized vocabularies, terminology control, file structure development, and indexing criteria will control to a large degree the compatibility between future data systems. Solution of this problem comprises an intellectual challenge that must be met if some of the most exciting of the prospects held forth by the computer can be realized.

Throughout this study attention has been directed to the future, because the future can be influenced, if not invented, by contemporary actions During the conduct of this study, leading data system specialists and scientific and technical administrators were requested to project future events and trends which would impact on scientific and technical data handling systems. These projections, although reflecting individual opinion and not consistent as a set, are indicative of the events or trends which planners of future scientific and technical data systems must anticipate and accommodate in a total concept of data management and handling. The following list enumerates some of the projections:

1968-1973 Computerized evaluation of experimental data in the area of nuclear cross sections.

1968-1973	Implementation of extensive computer systems which process evaluated data files for more effective utilization of data in design and development activities.
1969-1974	Utilization of high-density, photochromic micro-image devices for digital file storage, capable of storing 400 pages per square inch with an access rate of 200 to 500 microseconds.
1969-1975	Full implementation of the United Engineering Information and Data System.
1970	Utilization of laser-luminescent display devices capable of good resolution.
1970-1972	Utilization of mass storage devices capable of storing 10^{12} bits of data.
1970-1972	Establishment of the position of "Vice President for Information" within commercial and industrial organizations.
1970-1972	Implementation of a data system for metals and materials properties.
1970-1974	Utilization of handwritten document readers for input, capable of reading a full character set at a rate of 50 characters per second.
1970-1980	Establishment of Federal Government data banks serving assorted agencies in the three branches of the Federal Government.
1972	Establishment of a four-year update lecture/workshop for practitioners to impart theory and to arrest or retard personnel obsolescence.
1972-1976	Printing of essentially all publications by electronic composition or photo composition.

1972-1977	Implementation of state government data centers, supporting state functions such as planning, urban and rural development, industrial relocation, unemployment, etc.
1972-1977	Full implementation of a National Chemical Information System.
1973	Availability of computer storage on a public utility basis.
1973	Direct access to large data banks on service and fee basis.
1973	Discovery of a way to build modular software which is subsettable and based on an individual's needs and machine size, but which has almost the performance of a custom code.
1973	Extensive use of computers in "on-line" data-acquisition systems and "interactive" modes of operation, where the user and computer form a complementary unit.
1973	Implementation of a National Physics Data Center.
1973	Implementation of a working national data and document network in the field of medicine.
1973	Implementation of well developed mechanisms and data banks in machine-readable form in several of the existing information services such as Chemical Abstracts.
1974	Establishment of revolutionary types of educational courses in mathematics and systems logic.
1975	On-line recall and remote terminal display of the most frequently needed handbook data simultaneously with the presentation of the problem being attacked.

1975	Implementation of community-operated and community-owned computer centers, available to small industry, school, public libraries, etc.
1975	Satellite communication of scientific data via television circuitry between established national systems in the allied nations.
1975	Utilization of small personal computers for personal files.
1975	Telecommunication of scientific papers or data prior to publication, with option of inputting to a storage and retrieval system.
1975	Creation of an integrated index for environmental data, embracing the contents of data centers for weather data, oceanographic data, space and aeronomy data, seismological data, geodetic data, gravity data, geomagnetic data, etc.
1975	Expansion and vitalization of the National Referral Center, with more effective interconnections with other information services.
1975	Implementation of educational computer networks on a nation-wide scale.
1975	Replacement of institutionalized data centers with national data/information centers in areas where institutionalized centers duplicate, overlap, or fail to provide services.
1975	Utilization of flat displays with a resolution of 0.1 mm. or better, in sizes of one square meter or better.
1975	Implementation of multi-processing computers with a throughput equivalent to one nanosecond per operation (one billion operations per second per system).
1975	Designation of specific institutions in national data/information systems as agents responsible for providing services, at standard fees, to all qualified users in the economy.

1975-1980	Utilization of teaching machines for almost all subjects, including formal schooling and adult education (in centers).
1978	Implementation of an Electrical and Electronic Engineering Data Center.
1980	Utilization of three-dimensional, colored display devices.
1980	Joint operation of EDUCOM, NSRDS, NCIS, and NPIS.
1980	Utilization of mass storage devices capable of storing 10^{18} bits of data.
1983	Implementation of a National Social Sciences Data Center.
1984	Domestic access to data in libraries and information centers.
1985	Replacement of conventional hard cover books with newer information media.
1985	Widespread utilization of voice recognition devices.
1985	Replacement of libraries in the conventional sense with computerized, filmed stores of information, with film reader facilities for reading rooms, browsing, etc.

Effective transition from current data management and handling practices to the future portended by the preceding or similar trends and events will necessitate acceptance of responsibilities, assignment of priorities and allocation of resources. As stated by Dr. Weston*, "We must work harmoniously and cooperatively with those presently in the field--and with any newcomers who show interest, who have ideas and who will also work cooperatively."

*Weston, Jean R., M.D., "The State of the Art and Its Future," keynote address presented at the Third Annual Meeting of the Drug Information Association, Philadelphia, Pa., May 24, 25 and 26, 1967, p. 10.

E. Data System Development-- A Question of Who and When

Scientists and technologists universally depend upon data as a source of information, and treat it as one of the more important archival records. Data is well suited to the aggregation, summation, and manipulation processes that become increasingly relevant as the information population grows. However, authors desire to publish more data than the economics of the present technical documentation system permits, and scientists and technologists normally can monitor only a fraction of the existing, relevant, and available data resource. For these and other reasons, it seems obvious that future large-scale data systems are inevitable.

The techniques and mechanisms now available for large-scale data systems appear to promise much. Their capacities for manipulating data parcels are substantial. As data collections grow larger, these new tools can accommodate increases in usage and file storage with relative effectiveness and economy. It is no secret that the computer, modern reprography, and telecommunication methods can be combined to provide an impressive capability. The present scale and complexity of computer-based activity is rapidly bringing it to a point of familiarity as a working tool of the scientist and technologist. Some of this activity (the INTREX and MAC projects at MIT are examples) have specific relevance to operations associated with national-network linkages, whether of information generally or data specifically.

What does not exist yet is sufficient commitment to data systems among institutions with important national roles associated with technical information. In varying degrees, all of the key institutions seem to require additional knowledge, support for innovative change, and motivation before such commitments can be expected. This deficiency is not related to data systems alone. It has been noted in many instances, including the Licklider Report, that the Federal Government is achieving only partial success in persuading the scientific community to cooperate in integrating public and private services into a unified system for scientific and technical communication. The Licklider Report concluded:

> "That the field [information science] is not yet well enough defined to justify an attempt to design a national system at this time [1965]. One must first develop principles with respect to centralization and

distribution of functions and must understand better the 'real' needs of generators and users of scientific and technical information"*

Since 1965, considerable progress has been made toward national information systems, especially in selected fields such as chemistry and air pollution. However, as noted in the Licklider Report, major deterrents, such as the inadequate knowledge of user requirements and the reluctance of organizations and individuals to enter cooperative arrangements to alleviate common information problems, still need serious study prior to the launching of operational scientific and technical data systems or networks. Care appears warranted, particularly where large, new data systems, conceived as vehicles accommodating a national aggregation of usages and users, are contemplated. Such systems cannot but impact, favorably or unfavorably, on many of the institutions (defined as narrowly as organizations, or as broadly as cultural traditions) now associated with technical activity.

A few specific examples will illustrate how many pages, now virtually blank, remain to be sketched out before designs and investments in large, new operational data systems that are national in concept can be committed on a responsibly informed basis. For instance, what does a data system do to (or do for) the prevailing document system that now carries the primary burden of formal communication and reference data transmission in the technical community? In a data-rich, well codified, "mature" science such as chemistry, for example, would a chemical data subsystem provide a major economic relief for the professional and trade-press document publishers? Would it reduce reporting costs for authoring organizations, and perhaps be self supporting through filing charges for the archival accountability requirement of the author? Would it retire some document classes such as the data handbook and the new data journal? Would complexes of document systems evolve in which the document system would support the data system with judgmental information that had been proved difficult or uneconomic to codify into the data system? Or would the data system support the document system and be reached through the document indexes, with the data system providing an amplified data base and perhaps a computational competence for the document system? In the generality, the answers to all these questions might well be "yes" from the system-operator's

---

*J. C. R. Licklider, et al., Report of the Office of Science and Technology Ad Hoc Panel on Scientific and Technical Communications. 8 February 1965.

perspective. From the viewpoint of existing active institutions, an individualized pattern optimal for specific documents and documentation traditions would probably be the answer.

What are the transitional stresses in going from traditional information-processing mechanics, which orient around the typewriter and the printing press, to computer-based methods? Under what circumstances does the national interest justify national investment (i. e., a Federal subsidy) in the change-over costs? At what level of codification, volume of accumulated material, and existing or potential usage is there enough of a data base to justify converting the data population of a field for computer handling? At what point does the benefit of more rigorous tools for technical-mission supervisors justify the cost of creating data systems which the "bench" technologist can be made accountable for using? Who should assume the cost burden associated with culture transitions (e. g., of usage traditions)? Who should assume the intellectual burden of data-codification language development; who should assume the burden of maintaining it; and who should assume the economic burden of each? Information-husbanding institutions such as the scientific society and technical trade association would appear to have primary accountability for the intellectual burdens of their field. Cost accountability, at least to the level of underwriting philosophy, appears generally a Federal burden. From the viewpoint among existing institutional activity, however, there will be sectors where the institutions will identify economic or ownership-strategy arguments for underwriting system-development or system-conversion efforts. Entrepreneurial institutions will exist or be formed that are willing to risk investments based on new processing modes. Some such ventures may draw in part from existing information or data activities that are publicly supported. They may aspire to occupy a mechanized-level niche left vacant for any of a number of reasons by the institution historically associated with the subject.

Prospects such as the latter raise the question of the equitable rights and obligations inherent to the generator, the owner, and the user of data that become part of a national data system or network. What policies will preserve genuine independence of information-husbanding institutions such as the scientific societies that may operate systems developed or sustained through Federal underwriting? How can usage of such systems be managed so as to preserve equitable opportunity for private-venture data services? The burden of resolving questions of this nature appears to run to those institutions through which an adequately informed, reasonably disinterested citizenry can speak.

These examples suggest strongly that large new data systems or networks, conceived as vehicles accommodating a national aggregation of usages and users, are not likely to reach the operational level unless at least three facilitating activities have functioned over a period of time. They are:

- Economically sheltered conditions under which major data-owning institutions can experiment with the functional merits of data systems or services that are potential companions or successors to existing data-containing operational activities of the institutions;

- Research activity broadly addressed to the state-of-art aspects of national data systems;

- A forum through which policy can evolve from the grass-root case problems and opportunities identified or experienced as specific undertakings are suggested or implemented.

In answer to the question of "when", it would appear that the sooner these three facilitating actions occur, the sooner we will see the appearance of operational national systems. Answers to the question "who" would appear to come first from sectors where the transition stresses are minimal and few or no issues of equitability are involved. The huge body of technical information generated for agencies in the Executive Department comes to mind. The scientific society's classical role, and its intellectual and communication resources, appear to merit particular examination for extension into the data sub-regime. Industrial and other economic institutions that generate and utilize data seem to offer intriguing potential for simple buy-and-sell relationships with "non-sensitive" data systems they both supply and draw from.

It is obvious that there is some threshold level below which a data system possesses insufficient content to attract a meaningful patronage. However, from the view taken here that institutional factors are the controlling limitation, this does not appear as a really serious constraint. Assuming that constraint, the most probable expectation is for the evolutionary emergence of expediently useful data services, promoted through the research and support resources of a Federally-funded national data program activity. Over a period of operational experience, networks of functionally compatible data services should emerge in turn, along with advantageous mergers in processing mechanics.

This view of "who and when" lacks the dramatic visualization of a Polaris weapon system or the first Earth satellite. It also provides no single marshalling point for a deterministic, system-manager type of technical design prescription that simplifies day-to-day guidance toward goals, resources, and deadlines... and also can help considerably in securing support for a national data program. It is, however, both sensitive and responsive to many individual problems and opportunities that are extensively and sometimes dramatically visible in the numerous sectors of the scientific and technical community. We believe this view of the way future national data systems will develop is realistic, and that a well-conceived national data program activity can materially speed their emergence. It is our judgment that a supportive program framed to utilize the broad mix of existing institutional competencies and motivations will prove to possess enduring viability as a facilitator of the process generally. We believe that support of the plan proposed elsewhere in this report will result,by 1975, in operational data activities that are significant to the context of this study.

IV. THE PROBLEM AS DEFINED BY CURRENT ISSUES

The overall problem to be resolved through improved management of the national scientific and technical data resource may be defined through consideration of the current issues or subproblems of which the problem is comprised. Nearly one hundred such issues or sub-problems were selected, categorized, and evaluated in the course of this study of national scientific and technical data activities. The issues were identified by examination of previous studies, workshop discussions, mail questionnaires, and personal interviews. They were screened and grouped into five categories of significance in planning future national data programs. The five categories used were concerned with the following aspects of data activities:

A. Data Management & Handling System Requirements;
B. Data Packaging;
C. Data Handling Equipment;
D Personnel Capabilities; and
E Institutional Roles.

The issues were grouped according to these categories and sent to panels of individuals who were experts in the respective areas. Five panels, composed of over 300 experts, were asked to rate the relevance, importance, and amenability to resolution of the issues and to give comments concerning potential recommendations that might lead to their resolution. A specimen set of survey instruments and a panel member response to one of the issue evaluation questionnaires are shown in Exhibit IV-1. The responses from the panels, as well as the evaluations of the project staff, served as the basis for compiling the five sets of current issues that comprise this section. The issues are enumerated and briefly discussed. The discussions are illustrative of current viewpoints and do not necessarily constitute a complete or balanced review of each issue. However, in their entirety they indicate the complexity and limited level of knowledge which currently exists concerning requirements for data management on a national scale, and the data handling systems which would be responsive to these requirements.

A. Data Management & Handling System Requirements

The broad questions concerning underlying rationales connected with the need for a national data system, methods of its possible implementation in whole or in part, and a broad indication of alternate structures that might be possible were the subject of the first evaluation panel. As in the planning of most major programs, it is essential to answer questions such as these, and to consider the vital role they play in the initial, or preliminary planning stage, of the program's life cycle. In some cases, there was an evaluation by the panel of issues concerning the role of data systems in the context of current scientific and technical activities. In other questions, evolutionary and final structure problems are considered, and in still others, the parallel evolution of several systems prior to integration is the subject of evaluation.

The evaluation response to this set of issues indicated a consensus that all of the subproblems are difficult, but not impossible to solve. All panelists seemed to agree that dissemination aspects of the data handling systems pose more difficult problems than other aspects, but generally there was good agreement that specialized data systems will be major vehicles for the compilation, storage, and "referral" of data. Several panelists expressed some caution against establishing new data centers or systems without thorough study as to need, interrelation, and compatibility. Additionally, few panelists seemed to feel that input of data into data handling systems offered any new or major problems.

Several panelists, in addressing their comments toward networks, indicated that one node point of the network should be able to communicate with another node point insofar as indexes are concerned, and thereafter, the user would request hard copy products. Here, however, the almost classical problem of failing to distinguish data and data packages obscured the issue. Some panelists suggested products, such as handbooks or reports, as data items to be delivered, while others suggested the data content of packages, such as a single identifiable number, or small sets of such numbers as the key items for transmission from one location to another. In such cases, of course, the problems of commonality and compatibility of highly structured referral and search techniques, let alone hardware of any type, again are the key problems.

In general, the Panel members formed two groups--one group unwilling to foresee any substantial change in systems used for management and handling of data and the other group anxious to assume that highly automated, comprehensive systems will soon exist to serve all of science and technology. The proper perspective obviously exists between these two extremes.

Radical changes in data handling systems probably will not occur quickly simply because system designers do not know precisely the data service needs of scientists and technologists. In addition, even when service requirements are well defined the system designer does not yet know how to effectively match data handling equipments and methods to data service requirements. Quite likely new data handling systems will be introduced in an evolutionary fashion with the computer first used as an aid for structuring, storing, and formatting data for distribution in conventional forms. Knowledge gained from such experience will then be applied to implementation of more highly automated systems including query processing capabilities. Similarly computer techniques which are already widely used in design and other data manipulation operations at the work station of the scientist or technologist will continue to be refined and expanded in application. It, therefore, does not appear unreasonable to anticipate a future merger of data handling systems to serve all of the scientist's or technologist's data handling needs--both archival and day-to-day manipulation. At least this possibility provides a future frame of reference which possibly can guide data handling system development efforts.

1. How Can the Purpose and Goals of a National Data Management Program be Removed from the Realm of Conjecture and be Made Tangible Enough to Enlist the Active Cooperation of Those Organizations Most Likely to Help Implement, and Benefit From, its Establishment?

The problem implies that major steps must be taken to give a detailed demonstration of the utility of a national data system; a description of the envisioned system, and a detailed plan indicating exactly how existing data systems would be integrated into the national system, including specifics of how their continued vitality is both maintained and enhanced.

There seemed to be agreement among the Panel members that a central organization is required to formulate and coordinate national data program development. In addition, there was general agreement that the organization should not be located in nor wholly directed by the Federal

Government. However, there was wide divergence of opinion as to how much and what type of control this office should have over any program.

2. Since Optimum Design (and Operation) of National Data Program Depends on Knowledge of the Quantity, Quality, and Obsolescence Rate of the Data in Existing Systems, What Effort Should be Made to Define these Characteristics.

The Panel felt that this type of information is essential to the implementation of a national data program and that the census should be addressed to two areas: (1) the quantity, adequacy, obsolescence, and other qualities of existing data; and (2) systems and artifacts in which data or factual information are recorded, packaged, stored, and disseminated. The first census requirement includes not only a study of particular data files, but also of data users. Such a census would facilitate efforts to avoid "unplanned" or "undirected" growth in evolving data programs. Also, it could be directed to evolve a National Index of Scientific and Technical Data.

3. How Can Unmet Data Management and Data Handling Requirements be Ascertained; and at the National Level, Should Emphasis be Placed Upon Meeting the Inter-Communities Requirements or Intra-Community Requirements?

These requirements can probably be best ascertained by and within specific communities such as those formed by memberships of professional societies, trade associations, mission-oriented agencies, and individual firms or programs. However, some coordinating and review process is probably required to assure that the total effort is conducted in a manner which minimizes the total effort expended and assures that all communities vital to the national scientific and technological posture are included in the determination of requirements.

4. How Can the Required Functions and Scopes of National Data Handling Systems be Determined?

This study has indicated that the current data service requirements of scientists and engineers are still largely undefined. More particularly, effective methods are not available to predict scientists' and engineers' data requirements in the future, especially their range of needs as they shift from task to task, and the diversity of data needs so implied by such movement. There is a need to utilize existing and new prototype data systems to test user response and to make relevant measures of system and service effectiveness. Initial tests should include study of systems that are mission oriented and those which are discipline oriented.

One study element would be the analysis of several mission oriented systems as programs force them to interface more and more with each other, and the resulting increases in the effectiveness of the programs. Another study element required is the analysis of two other mission oriented systems that are now being forced to interface where one involves purely scientific data and the other purely technical data. The ability to retrieve both types of data from a single system has long been acknowledged to be a major problem.

5. What Should Determine Whether Centralized Data Systems, Decentralized Data Networks, Coordinated Data Exchange Programs, or Data Source Referral Centers Should be Included in the Development of National Data Systems?

This study has indicated that data management is for the most part performed by scientists and engineers who use the data, and that centralized data efforts serve only as support elements for their technical activity. It further indicates that centralized services and systems will not mature for some time.

The real problem is to make initial examination of the effectiveness of possible alternatives and combinations of system formats for particular classes of data users and the demands of existing and projected systems. The problem, therefore, may not be so much the ability to pre-structure a system, but to define the data management requirement so well that the structure which finally does emerge is optimal from the standpoint of practicality and use. Factors to be considered in determining system format will include: motivating forces between users, operators, managers, and agencies who must "pay" for the systems; and the developing modes of scientific and technical activity.

6. Should Operation of Data Systems be Totally Separate from That of Document Systems; Should the Two Perform Complementary Functions, or Should They be Totally Integrated?

This issue strikes directly to the heart of three interrelated problems: (1) What constitutes data in contrast to information? (2) Are data being lost, or being made less easily amenable to direct data retrieval techniques, by document handling systems (which otherwise serve excellent information purposes)? and (3) Is one system growing, or being encouraged to grow, too rapidly at the expense of the other?

There seem to be two answers to this problem, although the outcome of both answers essentially is the same. namely, the operations of existing data and document handling systems should be conducted so as to complement and supplement one another. There appears to be validity to the opinion that document handling systems have been, and are, developing far more rapidly than data systems. The fault, however, appears to be more one of sheer necessity, i. e., merely keeping pace with the information produced, including the data implied by the information content of a document. Thus, the data management function in many cases is confined to indexing of the general level of a document's data content, and thus, in the majority of cases, the extraction and indexing of specific data content of the document are bypassed. A logical and quite useful supplementary activity for document handling operators would be to include an adequate indexing of the data content within each document. Such indexed information, of course, would facilitate identification of data for subsequent extraction and incorporation into data systems.

From a second point of view, it is important to note that increasingly large quantities of useful data are not being published, but instead are being sought, and incorporated directly into data systems, such as NSSDC In fact, the publication step is being completely bypassed and, instead, the data are transferred directly from the point of measurement into the data system. Yet these data, as a conglomerate, on occasion also represent information--especially when properly indexed or otherwise codified Present data systems will also have to work in concert with the document handling systems. The very long term result will probably be an integrated system. On the other hand, purely from a practical operating standpoint, the current best practice seems not to force total integration of the systems, but rather have each establish those practices which will produce some form of minimax interface on the operating level where ease of transfer from one system to another currently is so greatly in demand

7. <u>If a National Data System is Established, How Can the Differences Between Basic Research, Technological Development, and Application Operations Activities be Taken into Account, Especially if the Systems are Structured by Discipline or Technical Field?</u>

The question actually poses at least two basic problems, both of which are related to file entry and the search technique used by the user. The recognition that there are basic differences in requirements and approach by basic research scientists (as contrasted to a development engineer or an applications technologist) also implies that each type of search essentially employs a different technique. Thus, if a file is structured according to one type of search technique, so also is the original structuring of the file entry, and one entry structure often is not amenable to another type of search. This is further complicated by differences in data requirements of each discipline. These differences impose additional variations on the various types of search techniques which, in turn also affect the file entry structures. Therefore, while redundancy of entry to a proper degree is advantageous, it also can be imposed beyond reasonable limits.

The beginnings of an answer, therefore, would seem to lie in the avoidance of a major emphasis on discipline, and turning instead to structuring file entry by related sets of data such as properties and related sets of substances and items. This solution may or may not imply an additional structuring in the automation to accommodate the three broad search techniques implied by basic research, developmental, and applications activities. If these goals can be achieved, large populations of researchers from diverse disciplines might be served as well, if not better, by a coordinated or integrated system. Such structuring might further imply more intensive work on correlated files and file inversion techniques to enhance cross-referencing and cross-matching abilities at many levels of file entry. The ultimate goal, of course, will be the national system capability to accept an inquiry from a user even when the inquiry is directed to the wrong system component, with switching responses to accommodate his needs.

8. <u>To What Extent Should Data Services be Rendered by a Referral Center or Network, Rather than a Data Retrieval and Dissemination Network or System?</u>

There seems little question that it currently is easier to develop systems which only provide the user with the locations of data rather than one that also delivers the data. It appears sensible, therefore, to emphasize that referral centers (or networks) currently offer a logical, first stepping-stone in the transition from our presently uncoordinated data activity to the more highly integrated data dissemination system of the future. Note, moreover, that even after such advanced retrieval-dissemination systems are created, a mechanism similar to the basic referral network will still be required to direct a user inquiry to the correct locations where the data required for adequate response are stored. In fact, one of the first components of any advanced system must be an adequate inventory of data on hand at each center, and such an inventory, of course, is the basic substance of any meaningful referral system.

To facilitate utilization of existing programs, it seems logical that the existing National Referral Center at the Library of Congress could be supplemented with specialized referral centers in specific areas of science and technology (e.g., engineering materials). In turn, it then would be a major responsibility for each specialized data center to maintain all the indexes of scientific and technical data for the field it serves.

In connection with the concept of a system that will evolve and grow with time, it further seems logical to assign the responsibility to these specialized referral centers to determine which of its data will first be put on-line for automatic retrieval and dissemination. A good, and possibly ruling, set of criteria in all likelihood will have to be the time dependency of certain select sets of data (blueprint changes, disease incidence), or the volume of demand for other sets that are not necessarily so time dependent (table of physical constants).

9. <u>What Can be Done to Assure Coordination of the Efforts of Equipment and Software Suppliers with Data System Requirements</u>?

There is no doubt that equipment developments are moving rapidly in the information systems field and, to some, it may seem that equipment developments are controlling the extent and nature of the automated data systems development. It is obvious, however, that major equipment vendors generally design their products in response to predicted demand, and therefore, that equipment vendors design major units to serve as general a purpose as possible. It follows that these present day equipments are inherently flexible and adaptable to a wide range of demands from different systems.

Therefore, the most important activity for data system designers and managers is to continually define the requirements for future data systems. Thus, equipment manufacturers will respond and meet those requirements. In addition, the computer equipment and associated industries will provide more effective general purpose programming languages that are currently unavailable and will be needed in the future. The efforts of data system designers to define, document and publicize their current and future equipment and software needs are critical, and programs should be provided to promote the required effort.

10. Is the Management of Scientific and Technical Data Amenable to the Systems Approach, Especially in Leading to Improved Means of Communicating and Using Scientific and Technical Data?

This question needed a more definitive term than "systems approach". Some panel members considered systems from a purely equipment or mechanistic standpoint (i.e., input, processor, and output), while others viewed systems as a logical ordering of knowledge or a well-integrated unit that serves a useful purpose. Similarly, each of these "systems" concepts may be evaluated from several standpoints: utility, effectiveness, cost, benefits, productivity, and amenity to management. The issue, therefore, boils down to the question of how critical is a national system, now and in the future, to improved effectiveness of the data-handling aspect of data management. There have been few if any, quantitative evaluations of the present data (as opposed to document) management requirements in specific fields of science and technology. Application of systems approaches to data management will require such evaluations.

11. What Evaluation Processes Should be Included in a National Data Program; Are Available Processes Adequate to Perform Efficient Data Evaluations?

There are two aspects to this problem of evaluating and re-using data. The first is concerned with the validation of so-called basic scientific data by the community-at-large. This validation process usually requires a complex and sophisticated intellectual comparison process within the immediate framework of the particular data generators and users. The current methods to disseminate, compare, evaluate, and finally, incorporate a piece of data into a validated scientific and technical data bank are not generally efficient or effective. Certain systematic methods, such as those implemented by the National Standard Reference Data System, could be employed in other data sectors to get such data into useful storage earlier, and with more systematic and extensive qualifications than have been used before.

The second aspect of the problem is concerned with data that an individual user might need to evaluate. For example, data from certain experimental runs may need to be compared, either with each other, or with a reference base, to discover anomalies or discrepancies. This typical procedure might be redirected to the validation of the base data reference, or some other related datum. For this type of validation problem, the user community needs much more advanced and flexible programming tools by which it can perform the comparisons and validations. The continued questioning and re-appraisal of data by independent investigators appears to be the essential safeguard to scientific evaluation--which might otherwise freeze acceptance of "official" judgments. Thus, use of data banks for comparisons may also provide a means to "keep score" while a fundamental process of science _per se_ goes on.

An important question related data evaluation is what types of data should national systems collect, store, or make available to users of the system. Knowledge concerning data use, practices, and needs of scientists and engineers is inadequate to answer this question at the present time. There is a general consensus, however, that scientists most often prefer raw, but highly qualified, data; engineers prefer mostly refined data; and technicians almost always rely on highly refined data (evaluation of a new raw data set, however, is almost always done in a step-wise, reverse order). Thus, despite the studies that obviously should be conducted on this problem, a national program probably will always have to contend with a multiplicity and diversity of data assemblages. This realization may also provide a key design clue to national system configurations--one facet must provide for a hierarchical structure of specialized vs. more generalized (refined data) centers. And, since time and cost of the data refinement process will always be a key variable, studies directed to eliciting more precise information on this problem need to address themselves to use statistics of operating systems, experiments in controlled data-service environments, and laboratory modeling and simulation of data-servicing concepts.

12 Which Aspects of the Scientific and Technical Data Programs Should be Centralized? Which Require Coordination? What are the Factors that Determine these Requirements?

These questions lead to a key system design question that must be answered early in the planning stage of a national program. Currently, majority opinion seems highly inclined toward the development of several decentralized data systems with indexing and switching centers to link one with another. Concrete justifications can be advanced for the networking approach: (1) Existing banks, which naturally would be incorporated into a national system, are widely scattered. These banks also serve special purpose data service needs for nearby user communities, and possess existing competency for data handling operations. (2) The investment in existing data centers is too great to abandon. (3) From the standpoint of safety--against both natural and man-made disasters--a diversified network has a greater chance of survival either in whole or in part. (4) Current data transmission costs make large centralized data files economically unattractive. Whereas considerable thought has been given to the questions of centralization versus decentralization of system operations, the same question applied to programming and planning functions has not been seriously considered. Panel respondents indicated that such national program functions require considerable visibility but need not be organizationally consolidated in a central office.

13. What Should be the Inter-Relationship of National and International Data Programs; What National and International Arrangements Should be Fostered to Attain Adequate International Exchange of Data?

Over the past several years, a large number of international data programs have been initiated, many of which include participation of U.S. organizations. Consequently, the interface between national and international is a valid area of concern.

There are two overall problems that should be recognized before turning to the specifics of international data exchange. First, major geographic areas such as Europe do not now have a centralized data program. Thus, world-wide coordination of data exchange may require coordination of as many as 300 political units. Secondly, certain types of data are subject to either security or proprietary restrictions. Declassification and release, therefore, may always be something of a problem.

The main problem is not exchange of purely scientific data, which are already widely disseminated by traditional means, but technological data of greatly sought economic value. Among the recommendations made relevant to this problem area are: greater U.S. participation in CODATA activities; the exchange of data on an individual case basis; voluntary exchange by U.S. firms; greater participation in UNESCO efforts; and a searching review of Federal legislation as to whether or not it is maximally appropriate both as to exchange limitations and to reducing the U. S. technology gap in many areas.

14. <u>There are Several Problems Associated with the Management of Data Contained in National Systems and Other Formal Data Efforts: Who Shall Be Able to Address Each System? How Will Priorities be Determined in Servicing Data Requests? What Limits Might be Placed on the Scope of Data Requests?</u>

These problems can be solved only when user surveys are implemented to determine the need profiles of the scientific and technical communities, and the availability of potentially requested data, and ultimately, the cost/benefit ratios for operating data handling systems.

15. <u>Should Cost-Effectiveness be the Principal Criterion to Determine Whether Future Scientific and Technical Data Systems Will or Will Not be Developed?</u>

Recently, the concept of cost-effectiveness has gained wide acceptance as a basis for deciding whether or not a system would be developed or operated. It is vital, however, to realize that cost-effectiveness is primarily useful in deciding between the development of alternatives (i.e., improved, streamlined versions) of well-established operating systems. There appears to be little utility in applying the technique of cost-effectiveness to non-existent systems that will, when implemented, be one or more orders of magnitude advanced in conceptual development from those presently envisioned. Consequently, the emphasis should primarily be on the development of one or more systems, the determination of each system's effectiveness (not cost-effectiveness), and the operating improvements that thus can be derived from the iterative application of the system-effectiveness criteria. Later, if desired, and if a sufficiently long operating practice can be established as a basis, certain cost optimization techniques may then be applied in the selection of systems.

16. <u>How Can the Potential for the Rapid and Effective Use of Existing Data (i.e., The Direct Coupling to Daily Scientific Activities) be Demonstrated as Exploitable?</u>

Underlying questions associated with this issue involve two points of view: First, for users, data managers, and even data producers, the question <u>per se</u> generally is considered academic (i.e., more and more exploitation is obvious as more and more direct coupling is achieved). On the other hand, this is a pressing question for program managers, who must justify the implementation, or continuation of an effort such as a national data system.

It appears that a pilot test of the coupling effect could be created to prove the basic exploitation advantages that could be engendered. The test could utilize two groups of scientists or technologsts, with one group using coupled research and data handling equipment and the other, the control group, using traditional methods of data search. With careful monitoring and evaluation, the improvement in effectiveness thus made possible could be quantified and evaluated.

17. What Should be the Function of Informal Communications (Technical Meetings, Correspondence, Discussion) in Regard to National Data Systems, and How Can These Communications be Best Coordinated with Highly Structured Data Systems?

The communication functions, as implied by the question, are chiefly a means to convey information (as contrasted to data) usually of a conceptual nature. Such information is based upon data, but the data supplied is usually piecemeal and is retrievable in some form on demand. It appears, therefore, that little if any conflict actually exists between the objectives of data systems and informal communications.

18. In National Systems, What Should be the Place of Vendor, Equipment, and Product Data?

There seems little question that equipment, product, and vendor data form an essential part of any national technical data system. These data, sometimes highly redundant, generally are disseminated by vendor or equipment/product suppliers, often in the form of catalogs, data sheets, and advertisements. In toto, many more private dollars and man-hours of effort are expended on such activities than on scientific data generation. As a result, many legal, economic, and technical factors are involved with incorporation of such data into a national system.

The issue, therefore, apparently comes down to one of three possible approaches: (1) Have the national systems "accept" from vendors all data they send forward, and include it directly in the system; (2) Have the national systems reject all such data and, instead, let the producers of such data support their own separate data network; and (3) Have the national systems create a special index and abstract file of all available catalogs issued by the vendors and product/equipment suppliers.

19. Should Retrospective Data be Incorporated into National Systems, and, If So, What Criteria Should be Used for its Selection?

The rate of obsolescence of data varies considerably among various types. While many users of scientific and technical data may have little interest in the historical background of their "current" data, for others there may not only be interest, but a genuine necessity, for introspective review on many occasions. In other cases, of course, there is an entire class of data that, while not "current", could never be regenerated (e.g., weather and epidemiological records or engineering drawing changes). The latter class of data usually has a use both as current operational data and as a basic trend record when used beyond its moment of currency.

The best answer to the general question appears to be that some decision needs to be made initially for each type of data when it originally is recorded in the system. This decision would involve factors such as: back-logged data; the cost of maintaining retrospective files; the requirement for such a large implied accumulation; and whether or not backlogging would flood the national system beyond economical or physical bounds. In summary, the advisability of retrospective storage of data may depend upon three main factors: the cost and physical capacity of the system; the requirements of scientists and engineers; and the nature of the data itself.

B. Data Packaging

The existence of many variations in data packaging modes is one of the principal issues in establishing a national data system. The difference between the data form and packaging format requirements for data generators, disseminators and users is the main problem in this area. These differences in requirements and the associated issues impose formidable criteria for the design of a national data system.

One major aspect of data packaging problems, as evaluated by the panel, concerned the approach that should be used to best resolve them. Decisions concerning formatting and packaging of scientific and technical data fall largely into the domain of the working scientist and engineer, who are reluctant to change their usual operational modes, despite recognized benefits for themselves, as well as the national technical community at large. The significance of this point is emphasized by the importance ratings assigned by this panel to questions concerning: packaging aspects of automated-system operation, centralization of data files, and transition between media forms. The evaluative responses of this panel are summarized in the following series of issues, which are listed in their order of rated importance.

1. To What Degree Should Automation be Applied to the Total Data Handling System, and What are the Packaging Implications of this Decision?

The systems approach must be applied to this design decision to facilitate optimal accounting for the data packaging requirements of the many user and generator communities. The total system, when it is implemented, must match with a master design that facilitates automation in each sector to a degree which is technically and economically viable.

The greatest challenges will be design of the man-machine interface, and the development of programs and languages to transform data forms and formats so they satisfy both data user and data generator requirements. The essential design tasks must be performed for each sector of the system individually, with less attention to the overall system requirements. Hence, the degree to which automation will be applied on an overall system level will evolve from the design decisions relevant to the subsystems.

2. In View of the Disagreement Concerning Use of Current Capabilities to Build and Manipulate Large, Computerized Data Files, What are the Chief Factors Behind this Controversy, and What are their Implications for a National Data System?

The problem is that there is actually little relevant knowledge available concerning the operation of large-scale scientific and technical data files per se. There is, however, useful information in existence from related fields, such as command and control systems, as well as document handling systems. Such related experience does not confirm the actual applicability of equipment and software to solving the unique problems associated with scientific and technical data systems. This observation applies particularly to a large-scale data system, and thus, it will apply with even greater ramifications to any proposed national data system. It may, therefore, be necessary for the Federal Government to support demonstration programs to develop large-scale, multi-discipline data files that would yield on-line access to various combinations of public, proprietary, and private data.

3. How Can the Capability to Shift from Established Media Forms to Machine Processable Forms be Improved, and What Should be Done to Encourage Acceptance of these New Forms?

Improvements in shifting from established to machine-processable media probably will be in direct ratio to the availability of large-scale systems to more and more users, and especially the practicability of remote console facilities (and/or microform readers) to serve the needs of data users. Any lag in the shift to machine-processable media arises more from the present unavailability of such remote console equipment than from non-realization of benefits of doing so. Any method that could quantitatively demonstrate how the burden of the present data glut could be eased should find ready acceptance. Therefore, demonstration projects of this nature would be useful, and Federal sponsorship of such demonstrations might well be in order. If conducted as controlled experiments, the documented results would not only provide the requisite quantification, but additional guidelines for educational and training programs

4. <u>Should any Effort be Made to Improve Overall Effectiveness and Timeliness of the Published Handbook Format, or Should Emphasis Now be Directed to Other Means of Disseminating Data?</u>

Until remote access to central data files becomes an effective reality, publications such as handbooks will continue to serve as an important means of disseminating data. The question of handbook effectiveness revolves around the question of timeliness and the economic viability of frequently updating hardcopy data packages. To place the problem in perspective, the major requirement is for both the handbook publisher and the developer of automated systems to evolve techniques for continuous updating of data contained in media. Thus, publishers should be encouraged to continue their efforts to develop automated updating procedures, such as computer storage of the handbook to facilitate automated updating of handbooks. This degree of automation would be a first step in adapting this traditional medium to future automated systems.

5. <u>Is the Standardization of Data, its Format, its Quality, and Other Characteristics for Advanced Mathematical Modeling and Analysis of Complex Systems Beneficial or Detrimental to Scientific and Technical Work?</u>

Current, non-computer oriented procedures to review and evaluate scientific and technical data are not comprehensive in incorporating valid data, and advanced modeling and analysis procedures for complex computer systems generally tend to be more comprehensive in their data requirements and inclusions. Thus, a distinct general gain for data handling should be obtainable by adoption of the standardization techniques employed in response to the demands of advanced procedures.

However, experience indicates that standardization can be detrimental if it is imposed too rigidly or prematurely. There is a growing body of expertise in this area, and some of it is reflected in the data standardization techniques employed in the forementioned advanced analyses. Organizations implementing the standards which receive wide acceptance at present recognize an appropriate level at which they are beneficial, however restrictive. The optimal level of standardization will increase as greater systematization evolves in data management.

6. Is the Current Computer Programming Effort for Scientific and Technical Data File Management Adequate to Support the Development of a National Data System?

By comparison with the highly sophisticated computer programs that have been developed for scientific and technical computations, the current effort in this area seems largely inadequate and uncoordinated. However, it is important to note that a national data system is a secondary development of computer usage, and that computers have been developed to perform computations, rather than data storage and retrieval. The fact that a national data system increasingly becomes attractive to automatically manage masses of data changes the picture. There is still only a beginning awareness of the problems involved in attaining this "secondary" usage capability. As those problems become more clearly defined, and a definite demand can be demonstrated for their solution, steps will be taken to provide adequate programs to facilitate such usage. Problem definition, therefore, represents the real requirement. Therefore, it seems vital that scientists and engineers, not just programmers, contribute to the development and testing of new schemes and programs to solve this problem.

7. What are the Implications for Scientific and Technical Data Packaging if the United States Should Shift to the Metric System in the Near Future?

Once accomplished, the shift will permit increased international standardization in future systems. However, instantaneous conversion on a national basis would be costly and practically impossible. The obvious solution is a gradual effort; for example, in the engineering field, a five-year period to prepare for adoption of the metric system. Maintenance for certain types of long life-cycle (low obsolescence) data might require even longer periods. Perhaps the best way, from the national data system viewpoint, would be to begin to input all new data in metric form, and rely upon adequate software to convert to the English system where practical realities have not yet permitted the changes in usage.

8. From a National Systems Viewpoint, What Effort Should be Made to Coordinate and Improve Data Packaging Efforts Related to the Graphic Display of Drawings, Standards, and Specifications?

The problem is that a given data item frequently has utility only in a limited context, such as one defined by a specific industry or program. Consequently, data management and packaging using this mode should first be directed to optimization efforts that serve each industry or program. An important point is that data contained in drawings currently has no standard encoding technique. Thus, national level efforts should be directed at the development of encoding techniques that are applicable in many use contexts, especially those associated with subsequent manipulation of the graphics package across discipline or program boundaries. A development program of this sort would be quite important, since it would produce the equivalent of a national asset. At the same time, a parallel program effort should identify, extract, and package for national network distribution that small portion of graphic data existing in each industry or program that does, or will, have widespread utility.

9. From the Standpoint of National Data System Development Criteria, What are the Potential Trade-Offs and Relevant Cost/Benefit Factors Associated with Data Screening (Pre-Processing, Organization, and Transportation)?

The problem appears to identify objectives for which a file is both created and used. Therefore, in the sciences both screened and unscreened data files will be required (i.e., those that contain raw working data and those that contain highly evaluated and refined data). Reference files containing highly refined and evaluated data will be in increasing demand by all types of users, and such files probably should be developed first by a national data system. Ultimately, working files containing new measurements should be incorporated and handled to retain as much of the original data content as possible. To promote a form of cross-linking between working and reference files, it might be desirable to develop a data "extract" or "sample" from the working files, and to use these to serve the same purpose that abstracts serve for document files.

10. Do Present Programming Languages Restrict the Ability of the Scientist or Engineer to Communicate, and Use, Computers, and If So, How can this Problem be Alleviated?

There seems little doubt that communication is restricted, primarily because of the language multiplicity barrier. There is some good justification to expect that this problem will largely be alleviated as time goes on, since one or two languages are expected to achieve dominance in the long run. They should, moreover, also evolve more and more towards a "natural" language use that is almost as flexible for the programmer as it is for the scientist. Additionally, an ability to handle computers probably will be part of the normal education of most scientists and engineers.

11. Since Computers Can Now Compose Type for Documents, What Additional Advances are Required to Utilize this Technique for Data Dissemination, and What are the Scientific and Economic Implications for National Data Systems?

Most of the major technological advances required to apply the composition technique to data systems exist, and usage is lagging behind the technology; NBS and other organizations have shown the feasibility of automated publication of tables and similar reference data. As more, less expensive equipment becomes available to a variety of new users, especially those with small volume operations, usage undoubtedly will catch up to the technology; and as that occurs, the composition capability will become more important, since it can also be locked in with automated methods to screen and evaluate highly redundant raw data.

12. How Can Data Not Included in a Publication and Retained by an Author (and Often Available from his Files by Direct Request) be Effectively Packaged and Made More Widely Available to Potential Users?

Behind this question lies a root problem for the entire data management field: What constitutes useful data?, and should all raw data being generated daily by scientists and engineers be made available? It is obvious that there is an economic cut-off point for selection of data to be included in a national system. For example, if it is expected that only five individuals may want to scan certain data within ten years, it may not pay either to publish this data or to introduce it into some automated system. For such data, it may be more viable to publish a data abstract, and expect the data generator to keep the raw data in his files for infrequent retrieval.

Currently, there is no hard and fast rule upon which to make judgments concerning the incorporation or rejection of data in a national system. Each set of data probably must be judged upon its own merits by both users and data system operators, especially in the light of costs and existing system capabilities. If judged "not feasible" for incorporation into an existing system, the data might nevertheless be abstracted to identify its existence and location. It might also be placed into special "off-the-shelf" microfilm or magnetic tape files at a central bank.

13. What Action is Required to Make Data Processing Programs Developed by One Federal Agency Available to Other Agencies or Non-Government Groups?

Unless there is a direct security involvement (and there should be few of these), any program developed and supported by Federal funds should be considered in the public domain and freely available. Thus, for the present, directories of existing programs should be developed, maintained, and advertised. These directories might identify the organization which developed each listed program, the programming language which was used, and the data processing function that it performs. Where required, funds should be made available to provide better documentation of existing programs so that they may be more easily employed by other organizations. Program documentation is one area where some basic standards should be developed and enforced. (Actually, a number of attempts currently are under way to create directories, but these are mostly for computational routines. Some of these efforts are: NASA, COSMIC; Statistics, SPEC; NBS, TIE; International Computer Programs, ICD Quarterly; and Brooklyn Polytechnic, International Directory of Computer Programs.)

14. Are Existing Data Packaging Methods Adaptive to Possible New Requirements that May Result from Technological Advances Such as Micro-Electronics or Computer-Aided Design?

A fundamental aspect of data packaging which is often overlooked is that all of the data in question are somehow ordered and therefore are amenable to automated manipulation and use. In a major sense, therefore, any technological breakthrough in this area still fundamentally relies on this orderliness of data. The problem, therefore, becomes primarily one of compatibility, and great hope lies in the breakthrough now evolving in optical reading capability for printed data, since it essentially converts written characters into digital pulses. The major impact of such a breakthrough will be to convert some existing data packages into ordered, recoverable signals of an essentially digital nature. Such digital signals might be considered as the key form for all data packaging, since virtually limitless equipment capability can be postulated to process the digital signal or to transmit it by many modes.

15. Will On-Line, Real-Time Data Processing and Computer Services Be Extended to Include Access to Data Archives, and, If So, What Data File Construction and Packaging Requirements Will Result?

On-line, real-time digital access to both files and computational processes may ultimately be the keystone operation within certain of the scientific and technical communities. The major desire will be to extend the capability to search all data files for the particular data set required by the user, which may or may not be compatible with straight-forward computational routines. A key goal will be adequate software to permit comprehensive search of all data archives. Another goal will be to transform data into digital form so that the data is amenable to manipulation within the system. Several equipment, file, and software configurations must be tested. One configuration that will require much testing will be a distributed system, in which computing capabilities and working files are located in a large number of remote locations, thus permitting the user to simultaneously tap remote data files from several banks.

C. Data Handling Equipment

Equipment state-of-the-art does not currently constitute the major constraint on national data system development planning. However, use of available technology in future data systems appears highly desirable if not in fact the enabling element of future data management and handling systems. The impact and direction of computer and other equipment developments on plans for a national data system is the subject of this evaluation panel. Among the issues evaluated and rated as highly important with regard to equipment aspects of national system development were those related to criteria for determining degree of subsystem automation, the need for developments in graphic data dissemination and display, and the pacing requirement for improved peripheral equipment. Other issues rated as being of lesser importance include the need for microfilm media and large-screen data display equipment. The issues evaluated by this panel are listed in the following pages in the order of their rated importance.

1. What Criteria Should Be Used to Determine Whether or Not to Design Specific Facets of the National Data System Operations for Performance By Manual or Automated Methods?

The concept of a future national data system almost implies a great deal of automation. However, this study indicates that the input process for certain data may not always be amenable to automated techniques. In each case, therefore, studies should be undertaken with a heavy emphasis on technical and economic feasibility of automation. Two criteria will determine if an operation should be manual: does the operation have too much intellectual content to be reduced to feasible logic and scanning flows; and is there sufficient data volume and query rate to justify an automated procedure.

2. What is the Probability of Significant Innovations in the Field Now Served by Facsimile Transmission? Could Such Innovations Substantially Enhance the Feasibility or Effectiveness of a National Data System?

The probability of a significant innovation is great, and a highly probable trend is the replacement of facsimile by digital TV. Perhaps the key innovation required is a still more significant development of advanced multiplexing equipment. This would lead to greater and more efficient use of current and projected long-distance data communication channels. Another area where a significant development could occur is the combined bandwidth increase that may come from communication satellites. Naturally, any breakthrough of this nature will be an enhancement for a national data system, since picture copy, and direct delivery as hardcopy, will become easier, cheaper, and not impose a heavy time- bandwidth demand on the communication lines.

3. Will Currently Available Switching Equipment Economically and Effectively Meet the Functional Operating Requirements of the National Data System?

Assuming that the configuration of a national data system will consist of several data banks interconnected by a switching network to direct inquiries to the appropriate data bank, a problem is foreseen concerning the inter-switching requirement between data-bank computers. Furthermore, a much more advanced time-sharing ability must also be devised for sharing between an individual computer and many more terminals. Thus, a new generation of switching gear will probably be required. However, such requirements are probably several years in the future and such equipments will probably be available when the requirement is fully defined.

4. Which Items of Equipment are Most Critical to the Implementation of Large-Scale Systems Created for the Storage and Dissemination of Scientific and Technical Data?

The following is a summary of the equipments viewed as critical to national system development:

Storage Equipment:

- Access mechanisms;
- Associative memories;
- Increased storage capacity for texts and graphics, or improved means to condense them--or at least improved cost-performance ratios for existing large memories;
- Means to digitize text;
- Large-scale read-only memories (photo);
- Large-scale read-write memories; and
- Memories that are reliable after frequent use.

Communication Equipment:

- Improved image transmission and reproduction;

- Greater bandwidth;
- Inter-terminal linkage; and
- Cost reduction.

Input/Interfaces/Processor Equipment:

- Data acceptance for permanent storage from remote terminals;
- Improved, very large scale, and voluminous switching devices;
- Optical readers;
- Linking between terminals and peripheral (or "sister") computers; and
- Greater capacity to handle a large number of terminals in a real-time mode.

An additional equipment requirement will be remote terminal equipment to facilitate:

- The ability to make requests simply;
- Browsing;
- Swift and easy hard copy delivery;
- Links (and "translations") between computers and terminals;
- Direct data insertion into central storage; and
- Reduced terminal costs, especially for graphics.

5. What Microform Media Developments are Most Needed to Facilitate Operation of Large-Scale Scientific and Technical Data Storage and Dissemination Systems?

Current capabilities seem, for the most part, satisfactory. However, the consensus is that the requirements for larger storage capability, more rapid access to data, and improved optical resolution are growing at an increasing pace. One approach to this problem would be to refine present techniques to achieve an additional compaction by order of magnitude of one thousand or more, and at the same time, greatly increase access. Electron beam writing (on film) at 30,000 characters per second may afford these capabilities. The second overall approach, of course, would be a changeover to digitization of data stored in microform, which has the advantage of amenability to erasable storage.

6. What Research and Development is Under Way to Produce Less Costly Data Display Hardware, Especially Large Screen and Remote Displays, and Is There a Need for Funding Beyond What Equipment Suppliers are Willing to Provide for this Activity?

From the standpoint of a national data system, the requirement for large screen displays is considered insignificant. In general, there is long-term satisfaction with a cathode ray tube display, and the associated costs. On the other hand, in the field of graphics, there is dissatisfaction with the complexity and cost of the circuitry and storage required to display the graphics using CRT.

There are two areas of needed development: one is the creation of a low-cost copier for a CRT display. The second is that, in certain fields, particularly that of engineering drawings, there is an increased demand for three-dimensional displays. In this area, some non-proprietary form of government-sponsored funding might be advisable if greater impetus is desired.

7. Should Priority Be Given to the Development of Larger Memory and Hardware Logic, or Programming Languages and File Structures?

From the point of view of a national data system, file structure and programming languages are the important problems. Although present equipment is sufficient for a national system, specific developments in hardware could greatly affect system economy. Examples are the development of associative memory processors and rapid multi-access to data. There are two schools of thought concerning memory equipment. One school tends to feel that, with modular abilities, the achievement of 10^{12} bit capacities may be sufficient for some time. The other school not only raises some question as to modular efficiency, but it also feels that the intense research now going on should not be dropped until some form of solid state capacity approaching 10^{20} is achieved.

8. Can Business and Scientific Languages and Hardware Be Adapted for Large-Scale Scientific and Technical Data Banks, or Are Specialized Languages and Hardware Needed?

The equipment is not a problem. The scientific and technical data computer essentially will need some arithmetic capability, a fast sort and compare logic and a large random-access, inverted file memory. Present equipment can do this acceptably. A simple user query language is not available at present. Its development will depend on gaining experience with users. Language development is basically a problem in human engineering.

9. Is It Reasonable to Expand the Idea of a Computer "Utility" to the Concept of an Information or Data Utility? If So, Will Hybrid Configurations Be Required, with One Module Designed for a Computing Capability and Another Designed for Data Storage and Retrieval?

The concept of a data utility appears more reasonable than an information utility. The information utility will require a much larger order of transformation and associative processes than a data utility. Vast amounts of data per se are "locked up" in current documents and document handling systems. The traditional modes of scientific and technical information flow are based on the use of hard copy; and many activities associated with storage and retrieval of scientific and technical information involve intellectual processes that seem too complex to economically program, except in those instances where extensive repetitive operations are involved. Data is more orderly, repetitive, and valuable per byte. A data system is easily compatible with the mathematical calculator function needed to manipulate it, but because the use pattern is dissimilar, modular design seems to be indicated.

10. What Technological Advances, or Types of System Implementations, Are Required to Reduce the Cost of Data-Handling Equipment and Thus, Assure the Availability of Future Data Systems to Small-Scale Users?

In the special cases where the user must have his own complete set of equipment, there seems little doubt that small, compact, low-cost systems will eventually be mass-produced using integrated or macromolecular circuitry. Within a broader context that includes the individual user, there is almost unanimous agreement that presently emerging time-share systems are the answer, both with respect to need and as to reduced costs. Costs, for example, may be based upon a nominal charge for greatly improved, mass-produced consoles remotely installed at users' immediate locations. These consoles would link to a central data system (that may process, or refine, the data employed by the user). If large numbers of such consoles become operative, the pro-rated charges per console should easily provide a means to pay for, and support, the central equipment. Therefore, when high-volume demand is achieved, cost considerations will be less of a problem.

11. Are Available Telecommunication Equipments and Channels Adequate to Meet Data Communication Needs? Can Such Equipment Be Effectively Used Under Existing Communication Regulations?

In the foreseeable future, the increasing traffic loads are expected to swamp both long-line and satellite communication links as presently structured. There is, however, a sufficient amount of development work in progress that should greatly increase the capacities of existing linkages (e. g., increased bandwidth, signal compression, digitalization, automatic switching, etc.). For the next several years, the high cost of data transmission by land-based common-carrier channels may severely restrict frequent, long-distance transfer of large volumes of data. Therefore, until substantially more satellite communication facilities are available, data transmission costs may restrict the structure of nation-wide data systems. In fact, even computer processable data may be transferred more effectively by physical rather than electronic channels. In the future, however, communication satellites are expected to alleviate some of the constraints as to which system configurations will be economically viable. While costs for satellite communication may now seem high, most considered opinions reflect the attitude that these costs will continue to come down as technical improvements continue and traffic volume increases.

D. Personnel Capabilities

The enhancement of data management processes through the development of a national data system will to a large degree be controlled by the extent to which personnel capabilities are developed to enable use of the system. The skills, knowledge, and attitudes of the scientific community must be adjusted to the new potentials and environment created through the establishment of the system, and there are several inherent issues associated with the required modifications of the community's working patterns. The issues were the subject of evaluation by the panel on personnel capabilities.

The principal area of concern considered by this panel was the educational and training requirements needed to elevate the community's performance level to that necessitated by a national data system. Among the issues evaluated by the panel were the role of the universities, the apathy of the community to government-imposed standards, the required skills for use and operation of a national data system, and the availability of educational resources. The issues are presented in the following pages in the order of their importance, as rated by the panel.

1. How Can Universities and Other Educational and Training Institutions Instruct Scientists and Engineers in the Use of Modern Data Management Practices?

This problem has two aspects; obtaining the proper instructional resources, and then creating the proper milieu for the requisite exposure. For many universities, little or no automated facilities currently are available, and this is a problem that must be overcome. When suitable equipment is obtainable, the optimal approach appears to be the introduction of a model data system to beginning freshmen. Where this has occurred (especially in science and engineering), freshmen seem to use the computer as "readily and easily as the previous generation used slide rules". The students then are expected to turn, where applicable, to the computer as an aid throughout their entire careers at the university. Such practices actually are in effect at a few universities today for both real-time and batch processing systems. In addition to familiarity with automated data handling capabilities, the students gain a knowledge of modern data forms, and the procedures employed to manipulate the data. A future development will be student use of banks of "standard reference" data, including search facilitating indexes and censuses, and the advanced techniques that can be employed in such searches. With the advent of these latter abilities, many educators indicate that a formal course in data management would be advantageous "to round out the student's knowledge and capability".

A parallel problem is concerned with identification of the department within a university which should be responsible for the course curricula. There is a consensus that such formal course work might be part of a management sciences curriculum to be established with the cooperation of all substantive science and engineering schools in each university.

2. Many Professionals Usually View Data Standardization as a Tertiary Bureaucratic Function, Even Though They Admit that a Minimal Standardization May Be Vital for the Future Viability of a National Data System. What Educational and Public Relation Efforts Might Counteract Professional Apathy Toward the Problem?

In organizations where management can recognize, or be convinced of the value of standards, programs like the "zero defect" campaigns would be helpful. Otherwise, minimal standards can be established by those who recognize the need and, as long as all those who have a genuine interest are admitted to the discourse, the necessary standards will be established and enforced ipso facto (probably at the request of government agencies, societies, and trade associations). More than likely, standards will be readily accepted by scientists who have happily left the standardization effort to other interested individuals. Currently, the question of standardization may remain of minor importance to even substantive experts -- until such time as they individually conclude that a real pay-off accrues through the use of a particular approach. When this occurs, standardization efforts probably will change from a tedious to a virtually automatic effort.

3. Which Particular Data Management Skills are in Most Demand Today, and Would Large-Scale National Data Systems Compete for These Same Skills?

The major skills currently required are those of the system analyst, system designer, machine operator, programmer, indexer, and others with a library background. Large-scale national systems undoubtedly will rely heavily on these same skills and, if so, such systems will also rely on, and compete for, existing skills. However, existing and projected training, especially when coupled with the education university students will be receiving, does seem adequate to meet future demands. The greatest demand, both today and in the future, will be for system designer/analysts, who essentially are functional line managers with the capability to adequately comprehend disciplines, methods, and techniques in a given field.

4. With More and More Data Activities being Implemented by State Agencies, Federal Agencies, and Professional Societies, What Can Be Done to Assess the Manpower Capabilities these Groups Will Require, So That Adequate Educational Programs Can Be Planned and Implemented?

   A national survey, sponsored either by a Federal agency or an adequately funded professional society, will be required to establish a planning base.

5. In View of the Wide Range of Scientific and Technical Activities, How Should Educational and Vocational Training Programs in Data Systems be Designed to Meet this Variety, and How Much Scientific Subject Matter should be Provided?

   In general, the basic training received in most rigorous scientific disciplines is of a common nature. The emphasis, therefore, should be reversed from that implied by the question. The educational or vocational program in the scientific data systems field should build upon as much basic scientific background as can be provided, and not the reverse. When a data system orientation has been added to a basic scientific background, it then will be possible for the individual to enter at will, and be reasonably at home in, a variety of other scientific disciplines. A career built in this fashion will always be adaptable to the changing scene, whereas a career built by the other method will always be in danger of being outdated.

6. Would Greater Programmatic Effectiveness Result if an Educational Program Which Consists of an Undergraduate Degree in Science or Engineering and a Graduate Degree in Business or Technology Administration, were Created for Scientific Data System Specialists?

   Prior efforts by industry and government in allied fields of technical administration indicate positive results of such programs. Similar results should also accrue in the field of data system management, particularly for those who will eventually be responsible for the primary management and operation of the large-scale systems of the future. The National Science Foundation is currently supporting programs of this nature at the Georgia Institute of Technology and Lehigh University. Most educators recognize that the step-wise effort must concentrate on an undergraduate science or engineering degree as the base discipline, and not the reverse.

7. What Program Should be Established to Motivate Data Users to be More Active in Searching and Acquisition of Relevant Data in Existing and Future Data Resources Prior to Generation of Redundant Data?

In basic research, such as is published in primary journals, the user is forced to be quite active in data retrieval because he must generate non-redundant data to assure publication of his results. However, the data user only has a certain amount of time for data retrieval, and where such retrieval is difficult and time consuming, it may actually be cheaper to regenerate the data he seeks. Furthermore, the user often is forced to be passive with regard to data search, because of educational deficiencies, inflexible methods, and inadequate index sources. The answer is to provide greater flexibility via vastly improved programming software. The user must then be trained in such uses; thereafter, time savings and ordinary competitive requirements will spur his active manipulation and search of data sources.

One panel member had given this problem extensive thought. His comments are as follows:

- Treat the transfer process as a four-terminal T-network. Take the user as he is, and define his characteristics as load impedance. Fix the image transfer constant (ϕ) with major effort on (allocated to) the primary information generator (sender impedance) and the data system (third independent property of the network). This concentrates effort on the two properties most readily accessible;
- Do not treat the user as the acquirer, or evaluator, of data transfer systems. The user is a system component; his evaluation can never be other than subjective;
- Do not seek to evaluate data utility from a user viewpoint. There are no known mathematical concepts on which to base the structure of a calculus for data utility. (This is a conclusion after four months of intensive research);

- Apply the concept that the user is required to be regenerative in varying degrees - defined by purpose, environment, and age of data. Consider that the data is a cue to certain intellectual processes of the user, which (processes) adapt the data to the immediate purpose. Train the user accordingly. This lowers the load impedance (Point 1); and
- Train users to data system basics, design, management, and use. Have the user acquire knowledge essential to supplementation of data, and analyze these user characteristics. In summary, make an effort to design user orientation to within practical limits, but concentrate primarily on user characteristics for design application. Don't depend upon making major changes in the user; design the system to fit him.

8. To Better Exploit Evolving Data Systems, What Type of Education Could be Provided for Engineers and Scientists to Encourage Them to Enhance their Performance through Increased Interaction with Data Archives?

Formal training programs that develop basic skills, that provide information as to what is available from data systems, and that induce motivation to use them will produce the best solution.
If the national system is to meet the needs of the engineer/scientist community, scientists and engineers must also be motivated to deliver their personally generated data to the system. One problem is that the community is inclined to disfavor input of raw data for use elsewhere unless they are sure of the effects of this effort. Another problem associated with such close coupling of scientific efforts and data systems is that certain concepts, or developments, seem to have resulted simply because their creators had limited data to begin with, and might easily have been discouraged from formulating their alternate solutions, had existing formalized data been available to them.

However, the coupling process again hinges upon the future development of adequate data manipulation programs. At least one source advances the opinion that between 70 and 1[illegible] common data elements form the "fundamental body" of all data in a given field, and constitute the structural building blocks for all data items in that field.

9. Should an Educational Program be Started to Encourage Use of Less Familiar Data Packages, which are Often Discredited as Less Reliable than Traditional Data Media (like Handbooks), Even Though They are as Reliable?

Most scientists and engineers do not use less traditional sources primarily because they are not familiar with them, and it is questionable if an individual scientist can afford the time to be constantly checking all new sources of data. Therefore, the best way to appraise scientists as to the existence of new data sources appears to be by all forms of advertisement in all likely areas.

The existence of a national data system, of course, would change the situation radically. Such a system would index (and cross-index) the data it had accepted into all its banks. The scientist-user would then discover the existence of such data at the time he had the greatest need for it. Thereafter, the scientist's use of the data would be strictly dependent upon his subjective judgment as to the reliability of the data and its applicability to a specific problem. Much of this can be quickly determined if the data in question is qualified as to source, and how it is obtained and evaluated. A data bank that expects to be accepted on a par with high reliability traditional sources must have the ability to qualify the data it incorporates, and to employ standardized techniques in order to quickly reflect the qualification and level of reliability of any data group. If so, then any data source will quickly be accepted; and widely used if the data thereafter prove valid in application.

E. Institutional Roles

Implementation of plans for a national data system must be based on the utilization of existing or future institutions and organizations in government and private sectors. Therefore, roles which these institutions must play in the startup and operation of a national data system affect the policies and programs of the system to a large degree.

Associated with the institutional roles is a set of issues that greatly influence decisions concerning their definition. For example, the problems associated with identifying the organizations that should develop and apply standards, and how they should be implemented, are of enormous significance in determining the packaging and management requirements of the national data system.

The respondents who evaluated the issues in this panel rated the issues as most important which were related to data standards, the assignments of financial responsibility, the need for national level activity, coordination requirements, and the language interface between institutions. The evaluated issues follow in the order of their rated importance.

1. In View of the Probable Need for Standardized Data and for Data-Handling Methods, What Institution Should Determine Which Activities Should Be Standardized, and How Should These Decisions Be Made?

The Federal Government has the largest involvement in scientific research, and thus in the resultant data, and has the greatest financial burden in support of data system activities and research. It might be recommended, therefore, that Government at least take action to assure a present modicum of standardization in the handling of basic scientific data, especially for those automated methods that are broadly applicable in more than one field of research. This is a minimum action recommendation, since it is clearly recognized that the field is highly dynamic and that premature overstandardization might easily stifle research. This limited Government involvement should allow for thorough discussion and consultation with all elements from information handling agencies, institutions, and industries.

Conversely, and in contrast to the recommendation concerning data handling, data standardization itself must be approached with a full appreciation of all the technical implications. There is only one group capable of handling that task, the scientists using a particular set of data themselves. Thus, actual standardization of any set of data must be a cooperative effort by the user scientists, engineers, and technologists from government, industry, and the universities. All three of the latter sources will probably have to bear a share of the cost of the effort, especially to promote discussion and interaction among the scientists responsible for evolving the standards.

2. Should Users Pay for Total System Development and Operation Costs, or Should the Federal Government Underwrite either the Development or Operation of National Systems, and If So, Which Systems?

Economic viability is an obvious prerequisite for the implementation of any large-scale scientific and technical data system. In view of the scope and complexity of such a system, it seems both necessary and sound for the Federal Government to support at least the development and implementation of such systems, especially those where Federal agencies will also be major users. It also seems sound for high-volume demand systems to be ultimately self-supporting through the sale of their services to both the government and private sectors of the U.S. economy.

A particular problem arises, insofar as systems may not have a high-volume demand but may supply data that materially support technological progress. In such cases, a strong position might be established for not only developmental support, but also for operational support, either in whole or in part.

Another particular problem relates to the access to, and costs of, educational institutions that must educate the future users of such systems. While they may provide a high-volume demand for such systems, should their costs be equal to industry-government users which, in the end, will reap the benefits of such teaching efforts?

None of the above precludes the possibility that the data systems in question could, or should, be managed by non-governmental groups. Probably many different types of "experiments" will be required to be undertaken in order to determine what sectors of the economy would be willing to pay the operational costs for the data provided.

3. <u>If DATA MANAGEMENT Is Defined as the Scientist's Husbandry of Data, and DATA SYSTEM MANAGEMENT as the Assembly and Operation of Equipment, Materials, and Procedures to Facilitate Data Management, What Level of Attention Should Be Directed to Coordinating Functions?</u>

Data management must be performed by the scientist and engineer for the foreseeable future. On the other hand, it can be envisioned that, at some date, data system management could be well developed and capable of handling very large, complex sets of data. Then it might be required that data systems directly couple their activities with the scientific and technical efforts. Prerequisite to this, there is a requirement for sophisticated systems, tools, techniques, and common nomenclature for data system management. In addition, there is the problem that neither the government nor any other institutional entity can control data management in a large scientific community. On the other hand, much might be done in the way of research and standardization in the data systems management field by the government and private sectors. An important concept that should not be overlooked in activities of this nature is that data systems are not ends in themselves; they are tools to aid the scientist as data manager. On the other hand, the scientist should be willing to negotiate with the data system manager concerning the mode in which his data is to be best stored in a data bank, as long as it can be retrieved in the form the scientist desires.

4. <u>What Action Should Be Taken to Coordinate the Cooperative Data Activities of Government and Non-Government Organizations, and What Institution Should Perform this Coordination Function?</u>

A single agency, not necessarily governmental, could be designated as responsible for coordination efforts. Its focus could be on the three areas of scientific and technical data activity, i.e., basic research, engineering development, and technical applications.

Through its publications, its planning and evaluation staff, and its advisory panels, the agency would provide a continuously regenerative feedback function, the primary objective of which would be to assure that successful development and operating activities would gradually become widespread.

5. Assuming that Input Language Must Be Relevant to the Data Generator and Output Language Relevant to the Data User, Which of These Should Regulate the Language Created for a Data System, and What Institutional Entity Should Solve this Problem?

A long-term solution is for the machine processor to make the language transition between data generator and supplier. This will, however, require research and development, which should be supported by the Federal Government. In the meantime, however, the Government should encourage, by funding, research and development directed toward cooperative efforts by software suppliers to develop effective languages with broad applicability. Such applicability could lead to an alternate, or conjugate, long term solution--a language that provides for the user a direct communication capability while using terms common to his discipline, and with little or no recourse to a special language. As positive results evolve from such software developments, some standardization should be implemented.

6. Since It Is Becoming Increasingly Apparent that Digital Data Communication Traffic Will Shortly Exceed the Demand for Voice Channels, How Can the Special Requirements of Scientific and Technical Data Systems Be Best Communicated to Regulatory, Legislative, and Advisory Groups; and Who Should Define the Special Issues that Arise?

Current advisory activities within the Executive Office of the President indicate that far-reaching decisions on communication regulations can be expected in the near future. If so, the implications for scientific and technical data systems must be quickly determined if they are to receive proper consideration prior to the implementation of new communication policies and regulations. In this regard, there appear to be two major considerations: first, the justification for special digital channels for scientific and technical data; and, second, an agreement as to what will or will not constitute private data, including techniques to keep that set as small and compact as possible, along with procedures to expedite declassification techniques. Consequently, it may be the responsibility of the COSATI Ad Hoc Study Group (on Legal Aspects Involved in National Information Systems) to explore, jointly with government and non-government groups, the implications of all proposed regulations via a continuing dialog with the FCC.

7. Who Should Have the Responsibility to Assemble and Convey Information on Available Data Files, the Availability of New Data, and the Existence of New Data Generators?

A logical step in the development of national data systems appears to be the inventorying of existing data resources and creation of indexes to facilitate access to this resource. Such an undertaking if pursued on a crash basis would be very costly; however, the task could be subdivided and pursued by individual scientific and technological communities. The development of an inventory appears to be a prerequisite step before data management and data handling requirements can be identified. Also development of indexes to data could serve as important service tools prior to the implementation of more ambitious data systems. For example, a data referral service could be based on the index of data existing within a community of science or technology. Such services would constitute a vital supplement to the National Referral Center. It also appears that a need exists for a service similar to the Science Information Exchange which would inform data centers and other organizations interested in a given class of data as to which research and development projects plan to generate such data. Such services might substantially reduce the lag time between generation and subsequent use of the data by another scientific or technological organization.

Responsibilities for these activities logically fall to a designated organization within each scientific or technological community so that they will be carried on in close association with the work of that community. Such decentralization would appear much more effective than a single national service center. Decentralization of these activities would appear to be almost a necessity if the responsibility for development of national data systems and programs is to be decentralized.

8. Who Should Be Trained by Educational Programs Designed to Create Greater Understanding and Ability in Modern Data and Information Systems, and Where Should These Programs Be Conducted?

System users, system operators, and system managers should be trained in such programs. Probably the greatest continuing training need will be for system users. As to where these training programs should be conducted, the long-term answer obviously is in the secondary education system. Currently, however, it is necessary to recognize that the discipline, and thus its designed education programs, is in a period of transition. Consequently, many types of training means should be relied upon at present, including on-the-job training for users and operators. The more formal aspects of such training can be directed toward both system managers and operators. This would include pilot efforts to develop formal programs which, except for a few specialized instances, are sketchy or non-existent at the present time. Such efforts can be in the form of night school courses, supplemental graduate work, or in-house training.

9. Since Most Large-Scale System Developments Appear to Encounter Similar Problems, It Might Be Fruitful to Provide Developers Access to Evolving Know-How in Other Systems. Who Should Operate the Required Information Exchange Effort?

Various governmental centers now collect and disseminate pertinent information on data system development. Some of these are: DDC, CFSTI, NASA, AEC, NBS, NSF, and BoB. The requirement, therefore, may more truly be to designate one of these as the central service for such information exchange. Such a service might then work closely with professional societies and trade associations, especially to encourage the establishment of panels for data system professionals, publications, and meetings in order to communicate on-going developments to all interested professionals.

10. What Role Should the Federal Government, and Other Organizations or Individuals, Play in the Screening and Review of Data to Reduce the Input of Erroneous and Invalid Data into Data Systems?

A basic solution is to ensure adequate training for individuals handling data (i. e., those who initially record the data, those who process it and physically introduce it into the file, and those who manipulate the files in search of answers) In the case of certain types of data, utilization of up-to-date techniques is probably all that is required. Perhaps a feedback mechanism to a central data bank could record individual user reactions, as well as suggested additions to the files. Then, as errors are discovered, or new data added, the input by the data bank operators would be governed by opinions of the data users.

Types of data which quickly become obsolete will require continual screening and evaluation. This will be a costly operation. What may be needed are automated programs that up-date or refine data that have become obsolete in one file and transfer them to other files where they are pertinent. This probably implies hierarchical levels of storage.

11. In View of the Need to Re-Evaluate and Re-Align the Data Now in Many Data Archives, Who Should Fund, Manage, or Perform the Data Evaluation and Re-Organization throughout the Various Areas of Scientific and Technical Activity?

Since this question actually has many parts, it is useful to begin to evaluate it in steps. The first question that arises concerns the real need to re-evaluate existing data. Only a small percent of archival data is in useful and usable form--hence, the need for computer routines to process it into the useful form desired by the user.

It may be more viable to provide machine-searchable indexes, particularly for re-structured (re-programmed) files that permit more ready access and acquisition. The latter problem must be left to data managers, handlers, and programmers at the individual data centers. The former problem, that of a central index of all available files, is one that will require some form of national coordination. It may, moreover, be one of the critical problems of a national data system. Trade associations, professional societies, and major data centers could serve as coordinating points for this effort. Quite possibly, the Federal Government, through agencies such as the National Science Foundation, will have to lend financial support to the undertaking.

12. Since Data System Management Capabilities Are Not Keeping Pace with Equipment Developments, What Organizational Entity Could Monitor Both Areas to Promote Better Coordination?

As a major buyer and user of equipment, the Government undoubtedly should assume a predominant role in promoting coordination. If so, the capabilities of offices such as the Center for Computer Science and Technology, National Bureau of Standards, should be utilized. In this regard, one or two aspects might be worth particular emphasis.

First, orders for new equipment developments should be studied in regard to determining existing and future equipment capabilities. New equipment developments may then be rendered more compatible with system requirements.

Second, certain preliminary standards might be imposed upon the programming of data files. The explosion in programming languages almost seems to be at a faster rate than data generation itself. Yet, many languages are mere off-shoots of root ones, and tend to decay as greater sophistication is developed in the root language. Thus, some control over certain software aspects will also remove a burden of "constant revision".

13. Who Should Establish Policies, or Participate in Activities, Concerning International Systems of Scientific and Technical Data?

The question of which Federal agency does not seem to be as much of a problem as the multiplicity of government offices that are loosely allied with international unions among the scientific and technical disciplines. The real problem is more likely to be that these responsible offices have not been able to keep pace with the growing level and importance of international data activities; nor have they been able to keep pace with on-going data activities in the United States. Many international data activities currently involve multi-nation scientific efforts. These data constitute the base upon which the first international data systems will be operated. It would seem, therefore, that some rapid means must be found to apply the best of our national skills to the formulation of positions that comprehensively consider the many factors that will be critical to international data systems. Because specific fields of science and technology are parochial (even though they are international), it may be profitable to let international scientific unions continue to work on particular international data problems. Later, enough know-how will be available to permit the scientist and engineer to move easily from discipline to discipline in search of his answers within an international data system (perhaps, for example, from software achievements that permit versatile entry via a common language orientation).

14. Given the Desirability to Include Vendor Proprietary Data in National Systems Directed Toward Developmental and Application Activities, What Organization Should Regulate Equal Opportunity to Data as well as Data Quality and Reliability?

The regulation of vendor data input into large data systems probably will be almost impossible, except from some broad policy standpoint The more probable outcome will be that government system operators will have to accept almost all data from vendors, and that in all likelihood they will be swamped by voluntary vendor contributions. Private system operators may be able to be more selective. The best that the data system operators probably can do is to assure that a minimum level of data qualification is provided. Thereafter, the quality and reliability of such data probably will have to be controlled within the economic constraints of the market, and vendors that do not enforce quality standards upon their data probably will not survive. In view of the mass of vendor data that might be introduced to such systems, the Federal Government in all likelihood will have to provide initial support in the form of limited financial subsidies, especially in the implementation phase, and particularly if it wishes to avoid systems required to accept all vendor data.

V. RECOMMENDATIONS

A. Basic Assumptions

The specific recommendations,presented on the following pages, for study and implementation of national data system concepts are based on certain assumptions which are implicitly stated throughout this report. Basic to all the recommendations is the assumption that the Federal Government has the responsibility to ensure effective management and utilization of the nation's rapidly growing resource of scientific and technical data. This responsibility involves more than making significant scientific and technical documents available to potential users; merely providing document sources does not assure that data are effectively communicated or conserved for future use. This assertion concerning Federal responsibility subsumes the view that scientific and technical information is a vital national resource: a resource to be utilized in the most effective manner by all professions, industries and agencies; and one that must be maintained in the best possible working order if its potential and optimal benefits are to be exploited. Moreover, it is assumed that scientific and technological progress will suffer if the corpus of scientific and technical data is not systematically and adequately maintained in a functional form. Progress is also inhibited if the means for communication of data are not continuously improved in order to meet the scientific and technological community's needs as expressed in contemporary requirements. Meeting the challenge and opportunity to construct new and effective means of treating data is one of the most crucial problems facing science and technology today. The challenge is based on the realization that the opportunity exists to build superior systems by utilizing today's new tools, techniques, and knowledge and, in so doing, greatly extend the utility of scientific and technical data. It is this challenge, more than the fear of being inundated by the flood of data that should prompt the search for new means to handle data more effectively.

It is also assumed that national systems for management and handling of scientific and technical data are now evolving from the efforts already in existence within many scientific communities and agency missions. The role of the Government is to focus on these present efforts, to coordinate them, and to provide data management policies on a broad national scale. The possibility of a highly centralized direction of national data systems is neither feasible nor desirable. What is needed is not a unilateral system created by a Government fiat, but the creation of order within the current process of national data systems development. Incentives must be provided for an orderly development, most specifically in the form of Federal funds made available to elements of the evolving systems to enable them to develop their potential effectively.

The development of an effective national data system should not conform to one monolithic blueprint. Rather, the national system will evolve to satisfy specific requirements of "real" communities in science and technology. Some data management and handling programs or systems will probably be subject oriented, some process oriented, some mission oriented, and some by a combination of these when they exist within a "real" community of scientists or technologists.

Since data management is in the midst of a significant transition and will continue to be so for some time to come, it is important that the transition be more fully understood as to its nature and importance. Greater characterization of the transition will help to enlist the resources required to guide it in the direction most beneficial to our national scientific and technological efforts.

National data systems will not constitute a new activity so much as an effort to get better organized and do a more effective job of data management in both the public and private sectors. For this effort to be successful, objectives must be articulated, priorities established, responsibilities assumed, and resources allocated.

It is assumed that advanced technological methods and equipment are essential to the concept of national data handling systems for the future. Therefore, the Federal Government must provide national data system policies flexible enough to allow for effective introduction of technological change, and provide financial support to assure timely application of appropriate equipment capabilities. Although new technologies in data handling tend to cost more initially than the methods and equipment they replace, the benefits gained in improved performance of scientists and technologists can be expected to offset the increased equipment cost. Valid cost-effectiveness ratios are difficult to obtain in this context . especially during the conceptual and developmental phase; consequently, such ratios should not be given over-riding consideration in decisions relative to introduction of new technologies to national data systems.

The following recommendations for study and implementation of national data system concepts are presented in the context of these assumptions. They also reflect a firm commitment to the idea that our nation is capable of controlling its future through conscious choices.

B. Policy Recommendations

To the present, there has been no overall Federal policy toward data handling systems and data management, nor has there been a focal point from which direction could be sought. There is, however, a critical need, today, for the formulation and promulgation of Federal policies because, as this preliminary study has demonstrated, there exists throughout the country much activity in the design and operation of data handling efforts. Moreover, rapid changes in technology, new tools and methods, promising new concepts in data transfer, and the necessity to further consider national systems make it incumbent upon the Federal Government to provide the initial emphasis and direction if an orderly and effective use of the national data resource is to be achieved.

The scope of this study did not include consideration of policies for other than the Federal Government. However, it was observed that non-governmental organizations, especially the professional societies and trade associations, need to re-examine their current policies relative to their responsibilities in the management and handling of scientific and technical data. In general, it appears that these organizations must become more aware of the needs of their members, and become active in considering what actions can be taken to meet these needs.

1. The Federal Government Should Encourage the Recognition of Scientific and Technical Data as a National Resource Susceptible to Systematic Management.

Recognition of data as a resource provides a valid perspective from which the management of data and the design and operation of data handling systems can be approached. More importantly, this perspective provides a fuller appreciation of all aspects of the problem of data handling. It brings into focus not only the access and communication aspects, but also those vitally important functions of conserving, maintaining, and refining the data. These latter functions are extremely important to the scientific and technical communities.

Data should also be recognized as a national resource of concern to the entire country. Because the United States has made a considerable commitment to the generation of data, it must also see to it that it is properly maintained and managed if its greatest utility is to be achieved. The establishment of a national index or inventory of scientific and technical data should be an integral part of the implementation of this policy.

2. The Federal Government Should Establish a Policy Position that Initiation of National Data System Development is Now Timely and Should be Undertaken as Part of a Broad National Scientific and Technical Data Program.

There are many pilot studies and programs for national data systems networks covering specific scientific and technical disciplines. There are other systems being proposed and planned on the state level independent of any consideration of the potential development of a national systems network. There are also plans for national document handling systems, but as yet, there is no plan for national data handling systems, nor is there a plan that includes both data and document handling systems.

The Federal Government is in the position to provide the leadership and direction that will assist these various efforts to converge toward a future national system or system of subsystems. This can be done by supporting a national program which would have as its primary objective improved management of the national data resource and a supplemental objective of development of national data handling systems.

3. The Federal Government Should Encourage the Evolution of a National Data Handling System Through a Program of Decentralized Planning and Development Efforts.

Such an approach is in contrast to the centrally planned and directed system. The reason for a decentralized approach to the development of a national system or set of interconnected systems is that this approach is more likely to generate true definitions of system requirements. In addition, an extremely fluid state exists in the fields of data management and data handling systems design. This condition is due primarily to the rapid change in technology and the still experimental nature of practically all large-scale data handling systems, as well as the nascent stage of data management. So many new approaches and methods are being tested in the field of data handling that, if given time, the more successful of the new approaches and methods will become evident, especially if they are tested in limited, local, but real data activities.

4. The Federal Government Should Create a Means for Coordinating Developments in the Design of Large-Scale Data Management and Data Handling Systems and Networks in the United States.

The purposes of such a policy are numerous:

(1) To provide a sorely needed focal point for nationwide data activities through which guidance, direction, and information can be sought;

(2) To support and develop those prospects that have promising applicability to national systems;

(3) To identify areas of duplication and areas that are not now being served; and

(4) To give overall direction to the evolution of a national system by identifying broad data management objectives and by developing broad planning concepts. In doing so, emphasis should be placed on the identification of effective management and systems approaches as they develop through the scientific and technical community. Because large-scale data handling systems are only now entering the concept-definition phase, it is important that the best approaches be identified now, even if this requires a delay in achieving cost optimization. As part of such a program, the Federal Government should assume the responsibility to identify, evaluate, and make available information concerning techniques, methods, and equipment applicable to the development of data management and data handling systems.

5. The Federal Government Should Stress the Development of Intra-Community Data Management and Data Handling Systems, Rather Than Inter-Community Programs and Systems.

Emphasis should be given to bringing about effective and viable data-handling systems on a nationwide, intra-community level. Once knowledge is gained as to which system configurations best answer the problems of various communities, this information can be applied to inter-community systems development. Too little is now known concerning the types of data to be exchanged between various communities or the extent of exchange that should take place. Further development on the intra-community level will help to provide the necessary information for this second phase of national systems development by identifying valid structuring concepts, efficient equipment configurations, etc.

6. The Federal Government Should Recognize the Different Types of Data Activities in Science and Technology and Establish Policies Commensurate with this Recognition.

This study has attempted to show that problems of data management and data handling differ significantly among discipline-research activities, mission-development activities, and applications-product activities. For each of these areas, the types of data, the needs for it, and its users differ. Specific developmental policies are therefore required for each of these areas. Also, each of these types of data activities must be provided a voice in determination of the goals, functions, and structures of national data systems.

7. The Federal Government Should Place Greater Stress on the Husbandry and Use of Existing Data.

The current Federal policy is oriented toward supporting the generation of new data over and above making use of existing data. This policy is evident in the Government's funding programs, where insufficient funds are made available to see that the data generated is fully utilized. The support policies of the Government should require that the data generated in Government programs is handled so as to conserve its potential utility and the data are made accessible for other uses. A policy should be promulgated whereby Federal agencies engaged in scientific and technical research and development would designate a minimum percentage of their total budget for data management and handling.

8. The Federal Government Should Acknowledge in its Policy Formulations Both the Difference and Interrelationship Between Data Management Programs and Data Handling Systems.

Data management includes those policies, procedures, and actions used for coordinating and directing efforts to determine data needs, generate data, and handle data in a manner which permits optimal use and conservation. An assembly of procedures, personnel, and equipment interacting to perform operations on data (recording, reduction, dissemination, etc.) constitutes a data handling system. Basic policy recognizing and giving due consideration to this distinction will materially assist the Government in its efforts toward development, management, and use of our national data resource. Such recognition will aid significantly in bringing together the diverse talents and skills, not only of systems designers, but also of scientists and technologists, required to formulate and implement effective data management and handling systems.

9. The Federal Government Should Support the Development of Programs and Data Systems Which Aid a Given Scientist or Technologist to Interact More Effectively With His Own Data.

Too frequently, data systems are viewed as the means for communicating data between scientists or technologists remotely situated either geographically or institutionally from one another. In fact, many of the data-related problems faced by the scientist or engineer involve the handling, evaluation, use, etc. of data at his work station - frequently, data which he generated himself. Therefore, national data systems should be viewed as extending to this level if they are to make major contributions to science and technology. Federal policy should include support of systems development efforts directed at the day-to-day working needs of the scientist or technologist, as well as systems directed to the less frequently encountered needs for remote communication of data.

10. The Federal Government Should Better Define Responsibilities for Policy Formulation and Coordination of International Data Activities.

The key requirement is not for redefinition of responsibilities for conduct and direction of U.S. involvement in international data efforts. The main problem is that the attention given by responsible offices has not kept pace with the growing level and importance of international data activities. Existing offices in the National Science Foundation, Department of Commerce, State Department, etc. need to be strengthened not only to permit them to better represent U.S. interests, but also to enable them to establish better communications and working relationships with on-going data activities in the U.S. Means must be found to apply the best of our national skills to formulate national positions which consider the many factors important in establishment of international data systems. Particular effort must be made to avoid unilateral actions by specialized scientific or technical communities. Currently, much international data activity involves multi-nation efforts to collect data on a worldwide basis. These data will constitute the data base which future international data systems must handle. Consequently, it is critical that such activities be planned and conducted on the most informed basis possible.

11. The Federal Government Should Adopt a Policy of Encouraging the Private Sector of the Economy to Develop Data Handling Systems and Innovative Data Management Techniques.

In doing so, the Federal Government should encourage professional societies and industry to develop data systems within their own communities. These systems should maintain and conserve the corpus of knowledge for those subject areas. In order to stimulate data system development in selected areas of the private sector, the Federal Government should support the initial planning and development efforts. As these systems advance to an operational status, the Federal Government should decrease its support to allow the economics of the marketplace to serve as a criterion of effectiveness.

Policy concerning support of data system development in non-government communities should recognize and acknowledge the realities of public and commercial interests in data activities. The Federal Government, therefore, should support to a greater extent discipline-research activities as opposed to applications-product activities.

12. The Federal Government Should Establish a Policy to Encourage the Accessibility of Scientific and Technical Data to as Many Potential Users as Possible.

Such a policy should not conflict with full recognition of the property rights of individuals or organizations. Rather, it would be promulgated with a specific delineation of private data (data which an individual or organization does not desire to disclose or release), proprietary data (data which the owner or possessor will release under prescribed conditions such as payment of a fee), and public data (data for which ownership and possession is in the public domain). Government support should be given to efforts for removal of the economic barriers which result in data being restricted when, in fact, the owner or holder has no objections to release of the data. In particular, the Federal Government should establish policies required to assure that data generated at Government expense are more accessible to other potential users.

13. The Federal Government Should Encourage Greater Recognition of Information or Data Handling Systems as an Integral Part of the Total Information Transfer Process.

In the past, concentration on increasing the effectiveness of document handling systems (including libraries) has overshadowed the efforts of handling data to the extent that data handling systems have hardly been recognized as part of the scientific and technical information management and transfer process. This narrow concept or picture of the function of information systems must be redrawn to include data handling systems and data management as a major part of the process of information transfer. It is vitally important that this more inclusive view of information systems be made widely known so that those who are considering doing something about their information problems will be aware of the various possibilities open to them.

Data handling systems go beyond the normal concept of document handling systems, in that they are more closely tied to the actual daily working environment of scientists and technologists. In this sense, they are more like the other tools used in the daily course of work and therefore, are not considered within the legitimate province of the scientific and technical information program of the organization, nor of the information systems design specialist.

C. Improvement of Existing Data Services and Systems

The following recommendations concern themselves with improving the existing methods of managing and handling data. Scientific and technical data are handled, packaged, and stored in many different ways and are transferred through a variety of media. When data are viewed as having a distinct quality unlike other information, a perspective is facilitated with regard to the methods employed to package, store, and manage it. This perspective helps to illuminate the shortcomings in the way data are handled today under existing methods, and also the possible means by which those methods might be improved.

Today, data are recorded, packaged, and transferred through a variety of media and formats. This study has attempted to analyze these various means to determine what steps can be taken to improve the existing services and systems. The recommendations cover six areas in which improvements can be made. The areas are the media employed, data automation, national systems, barriers in handling and managing data, education, and international data exchange.

1. Demonstration Projects Should Be Conducted to Explore New Media for Storing, Packaging, Formatting, and Transferring Data.

Such demonstration projects should be controlled experiments carefully conducted to gain knowledge about the effectiveness and potential of these new media. Documenting the results could be useful for education and training purposes, especially among data system operators and managers. A major intent of the demonstration projects would be to induce acceptance of the new media and formats (including microforms) by the intended users of the new methods. In order to reach a wide audience of potential users and to gain a wide acceptance, the demonstration projects should be held in a number of different areas throughout the United States in cooperation with, and with the support of, the private sector.

In concert with this technique to introduce new data packaging methods, additional studies should be undertaken to ascertain such important factors as the effect of computer-aided design on data requirements, technical feasibility, economic consequences, standardization and compatibility problems among the various data packages and data systems, and the adaptiveness of existing and new data-packaging methods to technological advances.

2. Study Should Be Encouraged on the Use of Computers and Other Data Packaging Techniques to Assist in the Publication of Data Handbooks and Similar Data Publications.

Scientific and technical handbooks and reference tools with a high degree of data content are an important and continuing means of disseminating data. Increasingly, however, these publications are becoming a less efficient method of keeping the scientist and engineer abreast of the latest data available. Publication time lag and the increased amount of scientific and technical activity makes obsolete an increasing proportion of the data content of these publications.

Greater emphasis should be placed on cutting down the time lag and the data collection and up-dating process through the use of computer techniques. Computers can be used for efficient up-dating and for revising handbooks. Prototype efforts using computers to prepare and maintain high data content publications, such as those planned by Project INTREX at M. I. T., should be supported. Knowledge about data structures and feasible transformations, which is gained from such prototype operations, will be useful in planning and developing future, more completely automated data systems Where computers now generate tables and graphs, efforts can be supported to couple these processes directly with that of producing a handbook.

Widespread use of source data automation, computer processing of data, and electronic or photocomposition processes will expedite data publication. Efforts at making these steps more efficient should be supported. Also, efforts should be encouraged to develop software techniques that will permit handbook data to be selected from computer storage for computational purposes Software offering this possibility to scientists and engineers will greatly increase the desirability of computer storage of selected handbook data.

Data handbook publishers should therefore be encouraged to participate actively in such studies and demonstrations.

3. The Federal Government Should Seek Greater Understanding of the Electronic Medium as a Means for the Husbandry and Transfer of Scientific and Technical Data.

Automation has been applied, mostly in isolated cases, to each of the functions (i. e., collection, reduction, analysis, evaluation, dissemination) involved in handling scientific and technical data. There is yet no case where all of the functions have been automated and pla· ·d within a total system utilizing in an optimal way the combined asset · man and electronic equipment. In no instance has the electronic medium been completely divorced from the medium of the printed page as a means for storing and communicating scientific and technical data. (Electronic printout of manipulated and retrieved data is obviously a necessary adjunct to such a system using the electronic medium.)

Those functions that are now automated within existing data handling systems should be studied and compared to ascertain the most successful methods now in operation. System concepts should prove fruitful in analyzing the successfully automated functions in terms of compatibility and possible integration.

4. <u>A Sense of Community and Participation Should Be Fostered among Data System Managers and Developers and among the Data Efforts of Science and Technology.</u>

The COSATI Committee has taken the first step by providing, through the present study, a preliminary directory of data activities in the United States. As a second step, a new referral center might be charged with maintaining cognizance of existing data activities in science and technology, and directing those seeking data to the proper data handling system. An added function of such a referral center should be to develop and maintain a directory of scientific computer programs. Third, a conference or series of conferences could be held among data handling systems managers and developers. One purpose of such a conference would be to acquaint the specialists with the data activities in the various fields of science and technology. Fourth, a data notification system could be established whereby the various data handling systems would be made aware of initiated and on-going data generating projects. Such an alerting system for data handling systems is now done on a limited scale by Science Information Exchange.

5. <u>One or a Series of Studies Should Be Undertaken to Determine the Interrelationships and Roles of the Different Types of Information Systems.</u>

A truly national information system would have as its components data handling systems, document handling systems, information analysis centers, libraries, management information systems, computer service networks, and possibly other types of systems. These elements provide distinctive services and perform varying roles. The problem to be resolved is how they can most effectively complement and assist each other and what should be the nature of the interface among them. Establishing a national system for one of them should not be detrimental to the service and role of one or all of the others. Initial studies could concern themselves with the possibility of augmenting current document indexing methods to include the identification of data within the document and notification for possible extraction and subsequent inclusion into an appropriate data handling system.

6. A Sustained Effort Should Be Mounted to Minimize the Barriers That Restrict Existing Data Handling Activities from Achieving Their Optimal Utility as Conservers and Communicators of Scientific and Technical Data.

Military, proprietary, and security restrictions act as barriers to the transfer and increased utility of a great deal of scientific and technical data. Existing barriers of this kind should be systematically identified and evaluated on an individual basis, especially in regard to their implications for the effectiveness of existing and future data systems. Periodic tests to determine whether existing barriers continue to be justified could lead to greater dissemination of data, and at the same time provide necessary guidelines for data systems managers. For instance, unclassified data contained in classified documents could be utilized by being transferred to non-sensitive data systems.

The increased number of data systems in science and technology throughout the country constitutes a growing argument and opportunity for a change-over from the English system of measurement to the metric system. Agreements can be established that new data systems will adopt, or at least include, the metric basis. For data systems in which computer conversions are available, the English-metric barrier can be eliminated by giving the user a free choice.

7. The Federal Government and the Professions Should Jointly Attack the Difficult Barrier Represented by Lack of Standardization of Data.

Data standardization is perhaps the most important technical challenge for those concerned with barriers to data usage. The great merits of standardization, in its application to system operations, are self-evident. The equally great hazards of ill-considered standardization consist of the creation of intellectual barriers through terminology or criteria that are incompatible among too many sectors because of the diversity of usages expected in each sector, or which are too rigid to grow with advancing technical knowledge and practice. Strong government support and organizing leadership appear appropriate, but the controlling technical guidance should be provided by the professions and other data-generating and data-using elements. The professional societies should take it upon themselves to support such programs in standardization, but must recognize the time that is required in such an activity, and the comparative lack of status associated to such a "labor of love." The Government might contract for some recommended work in standardization which, when finished, could be brought before a government-appointed committee for discussion and approval.

8. Effective Educational and Promotional Activities Should Be Recognized as Vitally Important and Funded for Seeking Optimal Usage of Operational Data Systems.

The educational effort should be carried on over a broad front, both by the operating systems themselves and by secondary-access systems such as the National Referral Center and the Science Information Exchange. The secondary services are particularly well placed to develop directories and to maintain network-type, organized displays of families of data systems, their scope, resources, and service policies. Operating services should be encouraged to experiment with demonstrations and cooperative arrangements that take direct advantage of their internal resources and service skills. The schools, professional and trade meetings, and major data-using institutions, both public and commercial, are all important candidates for these education efforts.

9. Contractors and Technical Units of the Federal Agencies Should Be Used to Test and Develop Data -Utilization Strategies in Technical-Effort Activities.

Economic shelter can be provided in Federal programs for tests of strategies in technical program management for data system usage. One example would be the requirement that a formal data search be conducted before authorization is given for a data-measurement action in a project activity. Tests of this nature should help to highlight data services and service features that are the most productive for identified requirements. Such tests will simultaneously advance the data-servicing art and perhaps develop techniques of technical management that are more knowledgeable concerning data system usage.

10. Data Systems Should Seek Means to Acquire and Make Usefully Available Qualifying and Critique Information Relevant to Data Items.

Reports on the experience of data contributors and prior users have the potential of deepening the significance of supplied data, and thereby strengthening the requestor's cognizance of the value of the data for his needs. Development of a workable means to acquire and provide such documentation may prove a break-through in the data-system art and in the use of data systems.

11. The Data Resources, Services, and Systems of Other Countries Should Be Studied for Their Potential Contributions to Our Data Systems.

It is evident that careful study of "exported" U. S. technical documents is an important means used by some countries to maintain their technical effectiveness. Non-U. S. activities of this nature should be surveyed to learn whether they generate data accumulations that constitute beneficial "imports" for U. S. systems and services. It is believed that most U. S. data services, particularly those that are research- and science-oriented, would benefit from strengthened data exchange relationships with technically compatible activities elsewhere in the world.

D. Development of Data Systems Capabilities

The national scientific and technical data system will provide services which to a large degree will be shaped by the development efforts of several professional communities. Among these are data-processing equipment manufacturers, software suppliers, the publishing industry, educational and training institutions, governmental agencies, and scientific and technical communities which generate and use data. The recommendations in this section deal with activities to assure that the capabilities of the national data system match with the requirements of the user communities. The activities include prototype test programs, coordination of software, media and equipment development, and generation of data standardization programs. These recommendations in this section are arranged in a sequence that facilitates logical thought development. The order does not imply rated importance.

1. The Federal Government Should Sponsor Demonstration Programs In Which Innovative Data and Media Would Be Employed.

Certain scientific and technical data have historically been packaged in specific media and formats. For example, the blueprint has been used for engineering drawings, product bulletins and catalogs for vendor data, and hardcopy handbooks for scientific reference data. The shifts, to date from the established forms and media to machine processable forms and microforms have been slower and less effective than they should have been. Demonstration projects should be implemented within a government programmatic and a non-government context. These demonstration projects should be conducted as controlled experiments with results documented for educational and training purposes.

2 Programs Must Be Implemented To Assure Coordination of the Efforts of Equipment and Software Suppliers With Data System Requirements.

There is increasing evidence that equipment developments are moving so rapidly in the information systems that they are controlling the structure of the automated data systems now being established. Therefore, scientific and technical data system designers and users must define their requirements more explicitly. These requirements cannot be effectively satisfied by equipment and program languages designed for business data processing or for mathematical computations, and equipment manufacturers cannot develop optimal equipment to meet ill-defined, non-standardized system specifications. However, by analyses and prototype testing, data system designers and users can systematically establish the functional characteristics of required equipment. In addition, data system designers must define, document, and publicize the current and future equipment market potential which exists in scientific and technical data systems. Equipment manufacturers and software firms have the basic capabilities required and can be expected to move quickly to meet economically valid equipment and programming requirements of scientific and technical data systems.

3. Future Data Service Systems Should Provide the User With Simultaneous Access to a Computing Capability and a File Containing the Data Required for Computations or Output Structuring.

Efficient means should be developed for providing effective access to both frequently used working files and to less frequently used reference files. Experimentation should be undertaken with different system configurations. One configuration to be tested should be co-location of the working files and computing capability with remote access to central reference data files. It is possible to expand this recommendation to include the concept of data utilities, where we may consider many such systems to be interconnected for associative searches and data exchanges. However, the concept of information utilities is only beginning to emerge, and there is an inherent danger associated with confusing data utilities with information utilities. Separation of these two utilities implies rather difficult transformation processes which will need further development.

4. Programs Must Be Established to Standardize Equipment, Data Form and Format, and Programming Languages.

Due to its large involvement in scientific research, the Federal Government has a substantial investment in the resultant data. Accordingly, the Government is increasingly assuming the financial burden of supporting the development of data systems to support research efforts. Consequently, the Government should take action to assure development and application of standardized methods of handling basic scientific data, especially those automated methods broadly applicable to data systems in more than one field of research. In contrast to data handling methods, standardization of data must be approached with a full appreciation of the technical implications. Therefore, scientists in specific areas of research must make the final determination of whether standardization of measurements and data is feasible or desirable. Whereas, Government-initiated standardization of data handling methods supporting research on a broad basis appears desirable, standardization of data handling methods supporting development or applications activities does not appear warranted except within specific Government development programs. Industry, through cooperative arrangements, should be encouraged to upgrade and standardize its developmental and applications data activities. In situations where it can be shown that standardization will contribute to a better integrated and stronger national scientific and technological competence, the Federal Government should, if required, subsidize standardization efforts. At a minimum, the Government should provide technical assistance.

In addition, as a major user of equipment in data systems, the Federal Government has a responsibility to select and use equipment in the most effective manner possible. Government practices materially influence the equipments developed and subsequently offered for non-Government use. Consequently, the capabilities of offices such as the Center for Computer Science and Technology of the National Bureau of Standards, which have been designated as responsible for government-wide technical review and coordination of data system development and procurement, should be augmented, in terms of funding and staffing, so that these duties can be performed adequately. Evaluative and planning documents developed in the course of Government systems implementation and operation should be made more freely available to non-Government organizations planning or developing data systems or equipments. In addition, organizations developing or considering the development of scientific and technical data systems should meet periodically, perhaps under Government coordination, to articulate and document their common equipment requirements.

5. A Program Should Be Started to Fund and Coordinate Development of Computer Programs Which Meet the Special Requirements of Scientific and Technical Data Systems.

Highly sophisticated computer programs have been developed for scientific and technical computations. However, there is a need for more effective programs to generate and control computer operations such as the rotation and translation of drawings and the interrogation of large data files. Computer programs must be developed for construction and manipulation of specific sub-systems of the national data system. It is vital that scientists and engineers, not just systems programmers, contribute to development and testing of the new programs, and that adequate user surveys are used to develop program design criteria.

6. A Directory of Data Processing Computer Programs Should Be Developed and Maintained.

Existing programs should be reviewed to determine their utility in other applications. Where required, funds should be made available to better document existing programs so they can be used by other organizations. In addition, standards of program documentation should be developed and enforced so that future programs will be properly documented. Government sponsorship will be required to initiate development of the directory, but the activity should eventually be self-sustaining as customers for the directory services provide support.

7. Several Research and Development Projects or Programs Should Be Used to Test the Applicability and Effectiveness of Automated Data System Concepts.

In these test projects, all operations involving data would be automated and incorporated into a system serving the project. For example, a typical scientist or technologist working on the project would have direct access to several data files which would be used not only to facilitate his own work, but also to communicate with other members of the project team. Information handling specialists would be planted in each project environment to aid in gathering and structuring of the data and the related data management techniques. Files directly accessible to the user should include the archival or reference files commonly found in data centers, as well as the frequently used work files often maintained either at the worker's desk or ... the computing center.

The operations of such projects should be carefully monitored and analyzed to identify data management methods and equipment applicable to other similar or larger scientific and technical program efforts. The program should begin with a partially automated system to include the services of some scientists and information scientists who can work with him to find out how he can best be served. This will be an education in both directions and will be much more beneficial than forcing a specific highly-automated system down the individual user's throat.

8. Programs Must Be Established to Develop Capabilities of Data Generators and Users Simultaneous To the Evolution of Data System Capabilities.

The objective of the data system is to serve the scientific and technical community, and this goal may be reached only if there are personnel development activities commensurate to system development activities. The skills, knowledge, and work attitudes must match the operating requirements of the system; moreover, these personnel capabilities must be prime considerations in the design of the system. On-the-job training programs, short courses, and workshops must be sponsored by both government and private organizations to develop the necessary data management skills, and to cultivate work attitudes that will foster the use of modern data handling systems. Furthermore, education curricula in engineering and the sciences must be modified to include instruction in the use of modern data management methods and equipment.

9. Professional Societies, Such As the American Society for Information Sciences, Should Be Encouraged to Establish Panels or Sub-Groups of Data System Professionals and to Undertake Development of Publications and Meetings to Communicate Developments in Scientific and Technical Data Management Systems.

Similarly, the scientific and engineering societies and trade associations should encourage the formation of sub-groups which would work toward becoming effective spokesmen, regarding scientific and technical data, for the interests of their profession or industry. The Federal Government should establish an information center to serve as a depository and dissemination agency for documents dealing with design, development, operation and management of scientific and technical data systems. The services of this center should be offered to non-Government, as well as Government offices. Such a center could be established by consolidating and augmenting some of the current information service activities of the NBS Research Information Center and Advisory Service on Information Processing, the NSF Office of Science Information Service, and the Bureau of the Budget Management Study File. As information concerning modern data management and handling systems is disseminated, the capabilities of the scientific community to use and help develop a national data system will be greatly enhanced.

E. Implementation of Future Data Programs and Systems

The promise of quantum increases in the utility of our national scientific and technical data resource provides the impetus for establishing future data programs and systems. The expectations are based on the realization, however limited at present, of the benefits of automating data handling, as well as the availability of data processing technology to implement real programs and systems. The recommendations of this pioneering study of national scientific and technical data management focus on these developmental actions, rather than the functional structure and organization of programs and systems. The approach is to define the objectives of the national systems program in terms of data management needs that must be met, and to recommend steps that should be taken to meet these needs.

Two points of paramount significance underlie this series of recommendations: first, the main problem to be solved is not the lack of technology but the lack of techniques to implement national programs and systems; and secondly, the implementation efforts are unlikely to evolve a monolithic system or program. The recommendations lead to the conclusion that considerable work is needed in the testing of several generations of prototype data systems and programs, with the concurrent development of techniques for managing and handling data. The recommendations that follow elaborate on some specific efforts that will enable attainment of these goals. The order in which the recommendations are listed gives some indication of their relative importance.

1. For the Foreseeable Future, Data Management Must Continue to be a Decentralized Process Directed by the Scientists, Engineers, and Administrators Responsible for Specific Scientific and Technical Endeavors.

However, as data system management methods and systems are developed and implemented, a capability will be created for management of larger and more complex sets of data. In the near future, efforts at the national level should be directed toward the development and test of systems or tools to facilitate better data management. Initially such tools or systems should be designed to facilitate currently definable data management functions, such as the location of data. As soon as data management functions are defined, data management requirements should be analyzed and articulated for workers at all levels from the bench scientist to the administrator of national-scope scientific and technical efforts. This should be done jointly by systems analysts and the workers involved in each level of activity.

There are several difficult aspects to implementing the recommended actions. Normalizing data management for the scientist will be difficult for disciplines that are not well-structured, subjective, and of significantly semantic conceptual content. However, highly technological and quantitatively well-developed fields will be approached first. In these fields developed data system management capability will automatically be transferred to organized data management. In these fields, it will be the responsibility of capable people in specific scientific and technical activities to adapt their work patterns so they are compatible with larger and possibly more complex contexts. It will be necessary for scientists and technolgists to play an active role in directing the software development efforts and in assuring compatibility of management requirements with system performance.

2. The Structures of Data Systems Must Evolve From Working-Level Responses to Real Needs. The Current Need is for Coordination and Financial Support of Systems Already Developing in This Fashion.

Prototype data systems should be tested which tie into a network several of the systems and services which the scientist or engineer now must use separately. The experimental system components should include automated recorders, computing equipment, automated archives of relevant data, archives of computer routines, reactive display consoles, and automated report generators.

The development of such prototype systems should be in response to adequately measured data requirements of specific user communities. Agencies sponsoring and coordinating prototype development efforts should monitor these activities to assure adequate system interfacing in the network development and attainment of defined data management objectives.

An initial step required in implementation of this recommendation will be to develop a coordinate system for describing current classification, media, form, format, languages, codes, and operations on data. The first need is to inventory existing information activities; the second, to review inventory data for duplication and gaps; and the third, to select related information activities to provide mutual support (i.e., initiate network planning).

3. Specific Agencies, Not Necessarily Governmental, Should Be Designated Responsible for Coordinating Data System Development Within Each of the Major Areas of Scientific and Technical Activity, i.e., Scientific Research and Technology Applications.

These agencies would assure that developing and operating data activities are gradually integrated into an effective national scientific and technical data system. They should be staffed with planning and evaluation personnel and should be advised by councils of leading scientists, technologists, and industry leaders from several mission- and discipline-oriented fields. Representation by several fields will assure satisfaction of their unique data management requirements.

Each coordinating agency should be provided with Federal funds not only to underwrite its operating expenses but also to disperse, in the form of matching fund grants, etc., to institutions developing or operating data efforts. Volume I, Section VI, elaborates on this recommendation.

4. The Professional Societies, Trade Associations, and Mission-Oriented Government Agencies Should be Encouraged to Identify and Service on a Joint Basis, the Data Needs of Their Community of Scientists and Technologists.

A centralized element of the Federal Government should establish forums and otherwise coordinate the programs of organizations tackling this problem. The government should assume special responsibility for identification and service of inter-disciplinary needs or those not served by other organizations. User need studies are costly and are only practical for discrete communities. Among the user need factors to be studied are the nature of the substantive work, the size and scope of technology involved, the education level and orientation of the community, the organization position of the users, and the motivational forces that drive data generation and search activities.

5. The Federal Government Should Acknowledge a Responsibility to Support, and Where Necessary, to Fund Efforts Directed Toward a Timely Development of Scientific and Technical Data Systems to Serve All Basic or Fundamental Research Activities in the U.S.

Where feasible, the operation of such systems should be by non-Government organizations such as trade associations, and should be at least partly self-supporting by sale of their services. Government funding of data systems to serve engineering and other developmental activities should be on a selective basis, with direct Federal funding and operation restricted to systems required for performance of specific missions assigned to Government agencies. Development of systems in other areas should be provided technical assistance and financial subsidies if it can be demonstrated that development of a given system will materially contribute to a stronger national scientific or technological capability. The government should, as it has in the case of scientific journals, support development and early operation of data systems where other funding modes are not possible. Eventually systems should become self-supporting, but up to now the start-up and operating costs have been so formidable that economic viability is not possible. However, systems should be carefully screened to avoid long-term or permanent Federal financial support when these costs can be sustained by the scientific community that will benefit from them. Two examples of systems which satisfy these criteria are:

- Multiphasic health screening facilities to gather stratified population data necessary for computing diagnostic probabilities.

- Data systems that can be used for testing Bayes theorem, discriminant functions, perception, likelihood ratio, and data collections for use in critical path planning of medical diagnosis and treatment when appropriate.

6. The Operations of Existing Data Systems and Document Systems Should Be Conducted So They Complement and Supplement One Another. Existing Document Handling Systems Should Augment Current Indexing of Conceptual Content of Documents to Include Adequate Indexing of the Data Content of Documents.

Such indexing would facilitate identification of data for extraction and incorporation in data systems. Increasingly large quantities of useful data are not being published; consequently, data systems must also acquire input data from other sources. Indexing of the data content of documents will facilitate direct search for data in the context of related information contained in the publication source. Multiple indexing of the data content will assure access by any one of several search routes.

In the future, it will even be desirable to by-pass publication of data and to transmit data from the point of measurement directly to the data system. The data system will thus perform many of the functions now served by publication (i.e., exposure for review and verification by colleagues, dissemination for use, and recording for archival or reference purposes). Therefore, data systems will in the future tend to supplant document systems, especially for archival purposes, for bench or console-level services to the technologist, and, to a lesser extent, for the scientist.

7. Existing Data Systems Should Be Used to Test User Response and to Make Other Measures of the Effectiveness of Specific System Operations and Service Concepts.

Existing data systems, such as the National Standard Reference Data System, National Oceanographic Data Center, and the National Space Sciences Data Center, etc., should be given support for developing and testing methods of identifying service needs, and for developing and testing means to measure the effectiveness of specific system operations in satisfying these needs. Additional systems should be implemented in other areas of science and technology, especially fields such as the biomedical sciences where broad-base attacks on scientific problems are required. This would permit development and testing of methods applicable to determining the data service requirements of the many diverse communities of scientists and engineers. Prototype systems should be implemented in typical work environments rather than in experimental information science laboratories.

Scrutiny of the existing systems should involve two quite different kinds of analysis. The first point of analysis is how the system relates to a well-defined body of users all by itself. The second point is that there should be careful thought given to the situation which is bound to occur when two systems with well-defined bodies of users discover or develop an interface which, in effect, makes them part of a larger system. The question here is, how do they relate to one another across the interface and how do they provide services to each other's users. The size of the individual systems will be an important point to take into account. Precautions that must be considered in the analysis include the danger of developing prototypes of use in analysis, but of little use in technical activities; and the possibility of results from poor systems distorting conclusions.

8. In the Area of Vendor Data, Initial Attention Should Be Directed At Upgrading of Data Activities in Individual Firms, Followed by Cooperative Efforts Within Trade Associations and Manufacturing Groups.

Increased effort should be directed to development of improved methods (e.g., computer controlled photocomposition of equipment catalogs.

automated design programs etc.) which can be applied to improvement of vendor data activities in a large number of industries. Motivation to develop hardware and software should be based on competition between manufacturers. However, incentives for private industrial and governmental utilization of the newly developed and improved methods are necessary. For example, hardware vendors with tens of thousands of employees are eager to help hospitals install automated health information systems and have yet to install them in their own plant facilities. Criteria of applicability and cost-effectiveness data are necessary to stimulate managers with limited intuitive judgment of new technologies, and government-sponsored studies may be necessary to develop these data.

9. <u>The Federal Government Should Support the Establishment of Demonstration Programs, Such As the Undertaking Proposed by the International Science Information Service, to Develop Large-Scale Multi-Discipline Data Files Which Would Provide On-Line Access to Public Data, Proprietary Data, and Private Data.</u>

Relatively little experience has been obtained from the operation of large-scale scientific and technical data files. Useful information has been provided by experience in related fields, such as command and control systems and document storage systems. Such experience has <u>not</u>, however, provided actual confirmation of the applicability of equipment and software to scientific and technical data systems. Therefore, demonstration programs are needed to develop data-file operation experience. Patent files offer one highly useful area on which to focus efforts.

In implementing such demonstration data files, stress should be placed on the compatibility of access with small-scale files in scientific and technical organizations. A pre-demonstration survey should be conducted prior to its establishment.

10. <u>Referral Centers Should Be Established to Identify the Location of Data Resources.</u>

Referral centers and networks offer a logical stepping-stone from our current uncoordinated data efforts to future, more highly integrated data systems. In fact even after highly integrated data systems are developed, a switching mechanism similar to a referral network will be required to direct inquiries to the location where the response data are available. The existing National Referral Center at the Library of Congress should be supplemented with specialized referral centers in specific areas of science and technology (e.g., engineering materials). Each specialized referral center should maintain indexes of scientific and technical data in the field served by the center. Ultimately, automated data location information can be used in directing queries to a national data system.

11. Future Data Service Systems Should Provide the User With Simultaneous Access to a Computing Capability and a File Containing the Data Required for Computations or Output Structuring.

Efficient means should be developed for providing effective access to both frequently used working files and to less frequently used reference files. Coordination of data file construction and manipulation must go hand in hand with a total system design in a computer system designed to make manipulation of stored data easy. Access to the system should be flexible so that both a sophisticated programmer and a scientist can have easy access to data and manipulate it.

Experimentation should be undertaken with prototype system configurations. Information should be stored in a manner appropriate for its degree of usage, such as: rarely used on aperture cards in microform, seldom used on tape, moderately used on disc, frequently used on drum, and quite frequently used in core. One configuration to be tested should be co-location of the working files and computing capability with remote access to a central reference data file.

12. Screening and Review Methodologies and Programs Should Be Established to Eliminate the Input of Erroneous and Invalid Data Into Data Systems.

Activities similar to those of the National Standard Reference Data System should be expanded to cover a broader range of data which are of use to the physical scientists conducting fundamental research. A similar system should be developed to cover basic data in the biological sciences. Critical reviews of data used primarily by the applied scientists and technologists should continue to be conducted on a decentralized basis by mission-oriented Government agencies, universities, and commercial firms. An assigned Federal agency should support development of tools to decrease the cost and increase the utility of data review and evaluation efforts.

Screening and review will require the efforts of technically qualified individuals and constant re-evaluation to facilitate elimination of obsolete data. Furthermore, user surveys should be used to validate evaluation criteria for both input and elimination processes.

13. Studies, Including Collection of Use Statistics in Operating Data Systems, Experiments With Workers Operating Within Controlled Data Service Environments, and Laboratory Modeling and Simulation of Data Servicing Concepts, Should Be Conducted to Acquire Knowledge Concerning Data Use Patterns in Science and Technology.

Results of such studies will indicate the degree of refinement, optimal output packaging mode, required input form, and format and other servicing requirements of future systems. The tests must be conducted using means which will resolve questions concerning actual or suggested systems and their effectiveness in providing data (in a defined field) in such a way that the users' questions are answered rapidly, accurately, and without bias or prejudice. Also, the tests must be made in a way to insure that the data system is completely passive; i.e., that the data system does not inadvertently influence research user or contributor opinions or conclusions other than through the supplying of fact.

F. Institutional Roles and Responsibilities

Priority is given in this section to recommendations concerning institutional roles that are both (1) of such breadth as to be significant in a national context, and (2) workable charters which we believe can be made vigorously viable in existing or achievable real-world environments.

There is a striking dearth of existing institutional activities that satisfy this dual criterion. Commercial publishing activities involving data, for example, are viable and nationally important, but necessarily ignore data-handling functions that do not promise profits. Professional institutions, as a class, are almost completely innocent of operational data-oriented activity. This is true despite the growing evidence that new tools for data servicing may break some of the near-disastrous bottlenecks traceable to the present publication systems of the professional societies. Many mission-oriented agencies tend to express relatively low-grade and self-serving purposes in their present data-management practices. Typically, one may find accumulated data directed toward archival storage with indexing that suggests the archive has been created primarily to prove the data have not been thrown away or lost. In a few instances (e.g., the Interservice Data Exchange Program), the indexing and the associated exchange activity result in a desirable continuing usage of the data, but the charters tend to be internally or defensively oriented (e.g., in the IDEP example, they are essentially a DoD-level aid for more effective selection and procurement of military commodities and components).

These observations are made not to criticize the institutions that are performing data management or handling, but to highlight the limitations of their charters as currently interpreted. Our comments are made primarily to emphasize two major points. The first is that currently, there is no central institution possessing (or at least expressing) a continuing obligation to look at data activities and needs in a national perspective The second is that there is, however, a significant corpus of potential institutional competency for accepting appropriate national charters concerning data, and bringing these charters to healthy operational viability.

It is evident from the previous recommendations that we have been impressed by the range and variety of individual data needs and usages disclosed in this study. Study findings suggest that the national interest in scientific and technical data will be served best in the near future through a primary emphasis on national programs that support, coordinate, and supplement existing activity, rather than effecting major mergers or creating revolutionary new systems. The recommendations that follow here are consistent with that viewpoint. We have suggested that functions required to develop an effective national technical data program activity be delegated to existing institutions that appear realistically compatible and qualified.

1. A Nationally Eminent, Non-Governmental Institution Should Create and Operate a National Advisory Council for Scientific and Technical Data.

The Council's charter should provide for two principal functions. It should be equipped to receive, evaluate, and, as it deems appropriate, advocate views from the scientific and technological community relevant to the broad national interest in data management and handling. Secondly, it should be charged with the duty of performing a broad oversight review, technical advisory, and policy formulation role for the key institutions associated with the national data program. In both these roles, the Council should be viewed as the ultimate public advisor concerning aspects of national data programs that affect the present and future strength of our national scientific and technological program.

It is recommended that the structure of the Council be essentially that of a consultative body. It should have a small permanent staff, which would function as a secretariat for a number of advisory panels. The panels should provide adequate representation for the experience and needs of scientific, technological and operational institutions concerned with scientific and technical data. The panels should include (but not necessarily be limited to) representation of:

- Discipline-research (scientific) data activities - professional societies, research specialists, etc.;
- Developmental-mission data activities - including relevant government agencies and their contractors, industry associations, etc.;
- Applications-product data activities - manufacturers, trade associations, public utilities, health institutions and practitioners, etc.;
- General-purpose technical data activities - including geophysical and other ambient-data activities, survey and monitoring organizations, the communities using their data products, etc.;

- Data-system technologies and activities - data-system managers and specialists, hardware and software institutions, information scientists, etc.

The National Council, its staff and panels would participate in development of a National Scientific and Technical Data System Development Plan. The National Academy of Sciences-Engineering would appear to be one institution possessing excellent potential to create and house the National Council on Scientific and Technical Data.

2. A Scientific Data Program Office Should Be Established in a Basic-Science Oriented Element of the Executive Branch of the Federal Government.

The activities of this office should fall into three principal areas, all of which would be supportive in character. The first of these would provide funding support to data-owning technical and professional institutions for studies, development, and (when appropriate) continuing operation by the institutions of scientific- as distinguished from technical-data efforts. Funding mechanisms such as matching grants are suggested, to induce the greatest feasible degree of planning initiative and operational identification by the participating scientific institutions.

The second activity would be directed toward strengthening of the data-usage potentials of existing information services of the scientific communities. Examples are the generation of indexes to the data content of existing documentation services, data-source referral services, and similar data-oriented ancillary products or activities.

The third activity would support grant and contract research addressed to advancement of the working effectiveness of the Scientific Data Program.

The administration of the Scientific Data Program calls for techniques generally associated with the offices and agencies supporting basic scientific research, except that institutions rather than individuals will normally be the entities funded. The National Science Foundation will be recognized as one agency possessing an appropriate existing charter and significant experience for the recommended assignment.

3. A Technical Data Program Office Should be Established in a Technical-Service-Oriented Element of the Executive Branch of the Federal Government.

The same three functional activities recommended for the Scientific Data Program Office are also recommended for this office, but technologically oriented organizations rather than scientific institutions would be the principal operating activities supported. The funding mechanisms for developing data-system potentials of national value would be directed toward commercially oriented technological institutions such as the trade associations. Special data services developed out of existing technological information resources (such as technical information generated in the programs of government agencies) would be developed by funding the data possessors, or operated through service contractors. The Office should have a broad dissemination service charter for data, comparable to the charter for technical document dissemination now vested in the Clearinghouse for Federal Scientific and Technical Information, but extending to non-federal, as well as federally generated data.

It is anticipated that even in its third function of supporting research directed to the advancement of the Technical-Data Program, the bulk of the actual research effort sponsored by the Technical-Data Program should be performed by the data-owning institutions or by research contractors. The requirements for direct research planning and supervision by the Program Office staff appear relatively modest.

The Department of Commerce has an existing basic charter to serve the technological community, and several operational activities that make it appear a well-qualified organization to consider for the recommended assignment.

4. A Data Systems Technology Program Office Should Be Established in an Element of the Executive Branch Possessing Expertise in the Technical Specialties Involved.

The establishment of this office is recommended to provide an institutional locus for leadership in development of techniques in the management and handling of scientific and technical data. The staff of this office would therefore have a more specific technical and programmatic accountability for national state-of-the-art levels (somewhat similar to ARPA's role in the military technology regime) than that associated with the other offices whose establishment we have recommended. The Data

Systems Technology Program Office would plan and conduct research, development, and demonstration projects. Some of these would be performed in-house, while others would be conducted under grant or contract. For example, a demonstration project by a commercial publisher might be commissioned, or a specialist organization employed to study an aspect of research of interest to the Office. The Program should be broad in its technical scope. It should include R&D on small-scale and local systems and methods aiding the individual scientist and technologist at his work-station, as well as work on larger, more complex, and more sophisticated data-system concepts and techniques.

This office should coordinate standardization efforts relative to equipments and handling methods used in scientific and technical data systems. To disseminate the knowledge generated through its R&D program, the Office should provide a suitable array of communication modes, ranging from an active conference and publication program to consultative and project services. A data systems technical information center could also be operated or sponsored by this agency.

The Department of Commerce currently carries several continuing technical assignments requiring the expertise called for by the program activity we have proposed. Should the Data Systems Technology Program Office be established in the National Bureau of Standards or other office of the Department of Commerce, we would anticipate some mutual reinforcement and benefit to result.

5. Mission-Oriented Agencies Should be Encouraged to Investigate and Develop Data-Husbanding Practices Contributing to Both the Institutional and National Levels.

The support programs described in Section VI of this report can provide funds for data-management studies and demonstration projects that will establish a more knowledgeable basis for optimal data practice in mission-oriented agencies. Such studies may reveal places where additional data-husbanding activity will be justifiable on a mission-effectiveness basis. The size and technological sophistication of the Federal mission agencies provides a fertile field for a wide range of such studies, including "data pool" interagency systems, identification of institution-owned data classes justifying joint support from outside using communities, and similar innovative approaches. COSATI, together with the National Advisory Council on Scientific and Technical Data, should encourage operational network tests extensive enough to generate experimental evidence of the total payouts potentially obtainable from data-handling systems operating on a national scale.

6. The Professional Societies Should Initiate Data Program and Systems Development Efforts and Related Studies, Pilot Demonstrations, etc. to Improve Data Management and Handling in the Communities Which They Serve.

The professional society is the primary steward of the languages of science and technology, including the data languages. It gives preferential treatment in its formal information services (which are now primarily the journals) to the more rigorously codified material generated by its community. It is thus almost uniquely equipped to undertake the leadership in basic studies and developments in its field leading to formal articulations of data languages, development and user testing of data indexes and compilations, and experiments with new, data-oriented information services that are not derived from intermediate publication formats. The potentialities for computer-based technical data service practices also deserve early and intensive study. Information service formats permitting the sale of information units (as contrasted to the present tradition of selling the "bundles" represented by publication formats) may prove the economic key to the society's capacity to maintain its traditional subject-comprehensive service charter without incurring financial disaster or becoming a de facto instrument of the Federal Government through major operational subsidies.

To the extent feasible, the data program development efforts of the societies should be integrated or coordinated with a National Scientific and Technical Data Program.

7. The Trade and Industry Associations Should Initiate Data Programs and Systems Planning and Development Efforts and Related Studies and Pilot Demonstrations, etc. to Improve Data Management and Handling in the Communities Which They Serve.

In serving technically specialized industrial communities, trade and industry associations perform roles comparable to those played by scientific and professional societies in their service to scientific professions. Effective data management in this sector is as important to the national interest as it is for data classes encountered in society activities. As noted in Recommendation 3 in this section, funding support from the Technical Data Program Office would be one substantive means to stimulate interest and provide economic shelter for innovative tests of methods for upgrading the husbandry of technical data. The trade associations are strategically located to contribute to Advisory Council panels, particularly on resolutions of issues involving proprietary/public interests in technical data systems. As an institutional class, they are probably one of the most appropriate focal points for coordinating technical data program developmental efforts and providing operational service to the industrial sector.

8. Commercial publishers, data-processing services, and the communications industry should be encouraged to advance the processing arts and establish organizations and facilities capable of meeting data-system service demands.

While the governmental and "societal" institutions must bear much of the burden of planning and development of viable data systems, commercial enterprise possesses the basic endowments for building and operating most of the "production plant" activities associated with data systems. The technological skills of the industry can be employed through contractual relations in pioneering work requiring the development of novel equipment, indexing, or processing arts. Experiments involving new publication formats (e. g., offering the data content of a proprietary handbook as a computer-based, reactive service, as well as in the traditional printed form) could be assisted through underwriting support from the Data Systems Technology Program Office (See Recommendation 4).

Commercial institutions generally appear the focal point for activities where cost and profitability criteria are influential measures or controllers of efficient performance, and where data activity provides sound opportunities for attracting private investment. Therefore, professional societies, trade associations, and Federal agencies should make maximum use of commercial services, not only as consultants on planning and development phases of data systems implementation, but especially as operators of segments or all of data systems when they are fully implemented.

9. The National Science Foundation and Other Organizations Currently Funding or Conducting Training and Educational Programs Should Consider the Special Educational Needs of Data System Designers, Operators, and Users.

Current programs directed to the education and training of information-science or library specialists should be re-oriented to also accommodate the needs of individuals oriented toward data management and data handling systems. Similarly, educational institutions and professional societies, such as the American Society for Information Sciences, should increase the number of institutes, seminars, conference sessions, etc. which deal with data management and data handling systems.

VI. IMPLEMENTATION OF RECOMMENDATIONS -- A TIME-PHASED PLAN

A series of recommendations concerning further study of scientific and technical data system(s) concepts were presented in previous sections. Most of these recommendations could be implemented independently; however, as noted previously, a major deficiency of current efforts toward data systems development is a lack of a means for coordination or a focal point for integrative actions. Too frequently, when data systems of the future are discussed it is in terms of fantasies of the year 2000 rather than the actions to be taken in 1969, 1970, etc. This section embodies our major recommendations in an integrated, time-phased plan. This plan identifies a network of actions which constitute a desirable sequencing of major steps and indicates some of the interdependencies among the recommendations for study and implementation of national data system(s) concepts.

A. Basic Considerations

As indicated in the introduction to this report, the Task Group has previously established three basic guidelines for effort directed to development of national scientific and technical information systems. These guidelines were:

- There should be no disruption of existing information channels;
- Account must be taken of widely differing capabilities of existing systems and the realities of funding long-established practices, rapid changes in information technology, and the differing needs of various segments of the user communities; and
- The Government cannot direct the private activities that form a major element of the national information capability--that it can only encourage them to join forces in a national system.

Also, the Office of Science and Technology had previously enumerated four desirable characteristics of national information systems. First, the systems would minimize the duplication of human effort both in the generation of data from research and development and in the handling of information resulting from this effort. Second, national information systems would require the establishment of certain standards for quality and form. Third, systems would not normally be operated by Federal departments and agencies, although exceptions would be required in some areas of science and technology. Fourth, the responsibility for national system(s) would be fixed in one Federal department to focus attention and effort on a specific set of objectives and activities.

Some of the basic questions examined in formulation of the recommended plan included:

- What are the principles which should guide national scientific and technical data system development efforts?
- How can the present situation be best illuminated and analyzed to relate present operations and capabilities to the overall objective of a more effective use of our national scientific and technical data resources?
- How can on-going efforts be promptly and effectively synthesized into a more unified and systematic total effort?
- What new or additional programs or systems will be required to either identify requirements or develop new means of serving existing needs?
- What should be the functional purposes of new programs and systems?
- What should be the relationship between the components, both new and old, of the total scientific and technical data program?
- What controls are required to assure that the national scientific and technical data program and related systems can be developed and operated effectively and efficiently?

The implementation plan proposed is based upon a belief that evolution of an effective scientific and technical data system must progress through the following stages:

(1) Development of an increased awareness and understanding of current scientific and technical data resources, data management and data handling capabilities, and data use factors by the many individuals

and organizations who must participate in the future planning, development, operation, and use of scientific and technical data systems.

(2) Application of systematic planning and evaluation methodologies to analyses of the data management and data handling system requirements for individual communities and between communities within science and technology.

(3) Coordinated application of improved data handling methods which

- satisfy the higher priority functional service requirements, especially those which are not currently being served,
- make optimal use of available personnel skills, equipment capabilities, and existing data resources, and
- are acceptably economical.

(4) Monitoring, evaluation, and refinement of data management programs and data handling systems to maintain a data system adequate to support national objectives for science and technology.

Efforts, to date, by the Task Group on National Systems have been directed almost exclusively to Stage 1 of the above sequence. The implementation plan presented in this section extends this effort and outlines actions required to move through the remaining stages of national data systems development. The plan is not highly prescriptive as to the configuration and functional structure of national data handling systems. Rather, primary emphasis is given to identification of actions which will evolve goals, competencies, and motivations which can be integrated into a comprehensive, yet decentralized program to achieve optimum utility from our national scientific and technical data resource. The recommended program should not, in fact cannot, be implemented on a crash basis; neither can its implementation be delayed if the U.S. intends to maintain its position of preeminence in science and technology. The plan is offered as a preliminary blueprint for establishment of a National Scientific and Technical Data Program. If the recommended plan is initiated in FY 1969, national scientific and technical data systems could be a functional reality as early as FY 1975.

B. General Plan

From the onset of this study, the search for effective means to develop a national data management capability has captured the attention, and remained the focal point of planning objectives. Study findings indicate that a fresh approach to organization, new working relationships, a working appreciation for the apparent and subtle differences between the data management needs of one community and another, and closer dialogue between the individuals and organizations involved are required for improved national data management. The plan recommended is intended to evolve a transcendence from the scale of individual or single-organization level data management to a broader, cooperative national program. It would be futile to attempt to escalate current efforts to a national data management program before the goals and objectives of such a program were firmly established and understood. An orderly and systematic step-by-step transition is necessary so that current programs are not disrupted and so that the proper stimulus can be established to evoke a responsive attitude among required participants. The plan introduces the recommended program in such a way that its development will be user-oriented, that the program will be responsive to change, and that timely modifications can be made during the transition from current data management practices to those established by the National Scientific and Technical Data Program.

Study indicated that an effective data program must not only include both government and non-government participants but should provide for interaction of these two major classes of participants at all functional levels within the planned program. The functional levels covered by the plan are:

- Centralized programming functions,
- Planning and coordinating functions, and
- Development and operating functions.

The centralized programming function consists of establishment of policies, definition of priorities, husbandry of legislative and budgetary needs, and overall review and evaluation of program effectiveness. This function would be coordinated by the Office of Science and Technology with consultation from the National Advisory Council for Scientific and Technical Data.

The coordinating and planning function consists of a systematic effort to involve a larger segment, both government and non-government, of the scientific and technological community in a cooperative planning effort directed to upgrading of existing data services and systems and to formulation of actions leading to improved future systems. The National Advisory Council for Scientific and Technical Data and its staff would coordinate the total planning effort with responsibility for coordination of detailed level planning assigned to two program offices -- one for scientific data activities and one for technical data activities. The National Advisory Council would maintain responsibility for integrating the planning efforts of the two program offices, other organizations such as the mission-oriented government agencies, and its own study results into a unified national program plan.

The development and operation function consists of implementation of programs and plans. Initially, this function involves local study and examination required to evolve national program needs. In subsequent phases of the program, this function involves the actual development and operation of data management and data handling systems. These functions would be conducted by designated agents within the various scientific and technological communities. These agents might be professional societies, trade associations, educational institutions or government agencies.

Figure VI-B-1* displays the major features of the proposed implementation plan. It contains four sequential and evolutionary phases, each of which is a prerequisite to succeeding phases and the ultimate objective. The following sections describe the plan and the sequence of steps involved in its execution. Description of the plan concentrates on the centralized programming and the planning and coordination functions because the plan, itself, is intended to evolve a further definition of the development and operation functions. Also, both the schematic and descriptions of the plan emphasize new programs and organizational responsibilities. However, it should be noted the continuation and improvement of current data handling operations is a vital part of the recommended National Scientific and Technical Data Program. The new programs recommended are not intended to supplant existing operations but to extend coordinated data management and data handling operations to additional areas of science and technology and to provide a means for coordination and improvement of existing programs and services. Panels of the National Advisory Council for Scientific and Technical

*Fold-out sheet at back of Report

Data provide a channel to assure that existing programs such as those of mission-oriented government agencies and agencies currently involved in collection of general purpose scientific data, have an effective voice in the development of national program plans. It is equally as important to note that the Panels also provide opportunity for inputs from commercial publishers and other non-government interests.

C. Phasing of the Plan

Figure VI-B-1* depicts the implementation plan as consisting of four phases, only three of which are detailed. The final phase, which is systems operation, obviously cannot be prescribed at this time.

Phase I - National Scientific and Technical Data Program Definition. This phase, extending for one year, can be viewed as consisting of two sub-phases. The first sub-phase essentially consists of establishment of the organizational structure required to implement the plan. The second sub-phase is devoted to a more explicit definition of the program plan.

The first sub-phase is initiated by review of the recommendations in this report by the Task Group, COSATI, and other advisory bodies to OST and the FCST. Assuming general approval of the recommendations, OST would initiate efforts to formally establish the National Scientific and Technical Data Program. This would involve coordination with affected Federal Agencies, the National Academy of Science - National Academy of Engineering, and exploration of the requirements for Executive and/or Congressional actions required to establish the Program. The two Federal agencies most affected, the Department of Commerce and the National Science Foundation, each are currently operating under specific legislation relative to facilitating the utilization of scientific and technical information by the non-government segments of science and technology. Pertinent legislation includes Public Laws 507 and 776, both passed by the 81st Congress, and Title IX of the National Defense Education Act of 1958. However, new legislation may be required either to establish specific authority for cost-sharing between these departments and non-government organizations or to establish a stronger justification for program funding.

Major organizational structuring steps to be taken during this sub-phase include establishment of a Scientific Data Program Office in the National Science Foundation, and a Technical Data Program Office in the Department of Commerce, creation of a Data Systems Technical Information Center to support the Program, and the organization of the National Advisory Council for Scientific and Technical Data. Figure VI-C-1

*Fold-out sheet at back of Report.

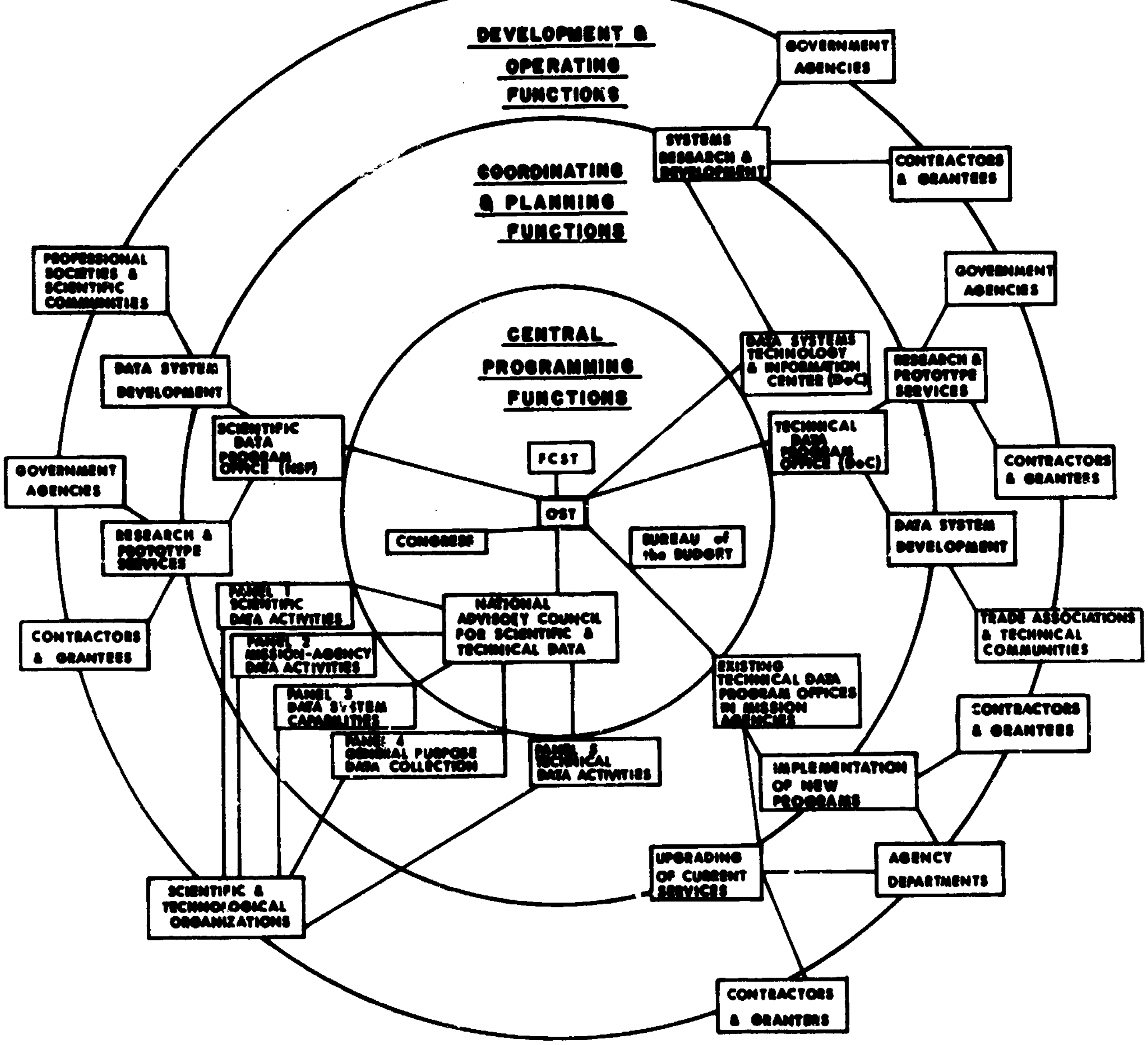


FIGURE VI-C-1 PRINCIPAL ORGANIZATIONAL and PROGRAMMATIC ELEMENTS of the NATIONAL SCIENTIFIC AND TECHNICAL DATA PROGRAM displays these organizational elements of the Plan and their relationship to other elements of the National Scientific and Technical Data Program.

Another important preparatory action will be the establishment of minimum budgets for data management activities within each of the Federal Agencies performing scientific or technical research and development.

The initial sub-phase would culminate with a White House Conference designed to inform the U.S. scientific and technological community concerning the Program and to enlist cooperation in its development.

The second sub-phase involves development of the National Scientific and Technical Data Program Plan. During this period, each of the major organizational elements of the Program would formulate and contribute inputs to the Plan. Both the Scientific Data Program Office and the Technical Data Program Office would further define their program objectives and establish procedures for selecting and establishing priorities among scientific and technical data activities to be included in their programs. Simultaneously, other Federal Agencies would identify the data management and handling projects within their respective agencies which would be coordinated with the National Program. These projects would be identified at the earliest possible date so that their interface with the National Program could be defined.

During this sub-phase, the National Council for Scientific and Technical Data would assemble a staff and establish contact with representatives of the various scientific and technological communities. Specialist panels would be selected and the panels would formulate inputs to the National Scientific and Technical Data Program Plan.

Phase I would be terminated by the joint issuance of the National Scientific and Technical Data Program Plan by the Office of Science and Technology and the National Advisory Council for Scientific and Technical Data.

Phase II - Formulation of the National Data System Development Plan. This phase of two years duration is vital to the proposed plan. It is during this period that requirements for data management and data handling systems will be critically reviewed both at the local and national level. Simultaneously, developmental and prototype tests will be conducted to ascertain the adequacy of equipments and methods to meet data management and handling requirements.

During this period, the Scientific Data Program Office and the Technical Data Program Office will be pursuing two activities. First, they will each select a limited number of scientific or technological communities to participate in Federally supported data system planning and development programs. Funds will be provided to a selected organization within the community for evaluation of data management requirements of the community and formulation of a data system development plan responsive to these requirements. The selected organization within each community will be required to obtain the cooperation and participation of other institutions in the community and will be required to follow general planning guidelines established by the Data Program Office. The latter requirement is intended to facilitate review of the plans and to permit study of similarities and differences of requirements from community to community. It is anticipated that an inventorying or indexing of the data resource of the community quite likely will be a part of the procedure employed to ascertain data management requirements. A second activity of each Program Office will be support of prototype efforts to develop methods or to initiate services which offer unusual potential for improving the management, dissemination, or use of data within specific scientific communities. These prototype tests would be selected not only to alleviate specific problems but also to identify those methods and services which should be considered for implementation on a broader scale. For example, the Technical Data Program Office might explore the utility to commercial food processing firms of a data resource referral service which provided access to data files created by the activities of the Department of Agriculture, DOD Quartermaster operations, etc. The Scientific Data Program Office might explore the feasibility of creating an index to the data content of the journals serving a given community of researchers. All of these prototype services would be carefully monitored to ascertain their potential application to national systems.

Whereas the Scientific Data Program Office and the Technical Data Program Office would direct their support largely to specific communities, a concurrent program centered within the Department of Commerce would support tests of new methods and services broadly applicable to the data management requirements of several scientific or technological communities. This program would emphasize the adaptation of computer and other technologies to data handling functions.

During this phase, non-government organizations would begin to play an increasing role in formulation of the National Data Program. Professional societies, trade associations and other appropriate organizations would apply to the Scientific Data Program Office or Technical Data Program Office for planning grants for their communities. Once selected as the designated organization and awarded a planning grant, the professional society, trade association, etc. would coordinate the evaluation and planning effort, making sure to provide for participation of all interested groups in the community. The designated agency would culminate its particpation in this phase of the program by submission of a data system development plan to the sponsoring Data Program Office. This development plan would be formulated specifically to meet the requirements of the community and might be highly complex or simple, and might require extensive or little change in current data management and handling practices in the community.

During this phase, the National Advisory Council for Scientific and Technical Data, its staff and panels would be studying inter-community data management and data handling requirements. In addition, these bodies would be formulating plans for integrating appropriate data management and data handling efforts into a national program. These efforts would terminate in the joint issuance with the Office of Science and Technology of a National Data Systems Development Plan. This plan would integrate findings from prototype tests of methods and services as well as the system development plans generated by individual scientific and technological communities. It would also integrate findings from operations and analyses conducted by the mission-oriented agencies of the Federal Government. Since these data handling operations are intimately associated with on-going research and development efforts and would have a longer operating history than any of those initiated and tested under sponsorship of this Program, their contribution to formulation of the National Systems Development Plan should be substantial.

Phase III - Development of National Systems. This phase of three years duration will test the feasibility of national data system development. During this period, several systems will be under development concurrently and will undoubtedly differ substantially as to structure and functional purposes. Each of the Data Program Offices will be contributing to support of development of full-scale data handling systems to serve specific communities of scientific or technological activity. A significant part of the activity during this phase of the Program will be devoted to development and testing of methods and facilities for serving the data handling needs of specific communities. These facilities may be centralized or decentralized depending on the needs of the community served. However, considerable attention will be given to questions of standardization to assure an optimization of compatibility between all systems being developed as part of the National Scientific and Technical Data Program. The National Advisory Council will study standardization requirements and make recommendations to promulgated through the Data Program Offices.

Whereas the development effort within specific scientific or technological communities will be largely evolutionary, the development efforts within Federal Agencies will be increasingly integrative especially in the area of general purpose data collection. Through its appropriate Panels, the National Advisory Council will continue to study and develop plans for consolidation of the data activities of Federal Agencies wherever it can be shown that such consolidation would result in more effective national use of the data resource.

The development phase will be terminated by a review of development and integration efforts. This review by OST and the National Advisory Council will precede the shift of systems from developmental to operational status.

Phase IV - Operation of National Systems. This phase of the Program will increasingly involve non-government organizations for it is hoped that many such organizations will voluntarily participate in the Program without requiring government support of development operations. This should become increasingly feasible as effective methods are developed and the benefits accruing from data system development are demonstrated. It should be noted, however, that it will probably be a considerable period before the Federal Government can terminate its support of data system development efforts. In fact, it can be anticipated that even after the first scientific and technological communities reach an operational status with their systems, other communities will not yet have initiated determination of data management requirements.

D. Special Implementation Considerations

1. Fiscal Factors: In the previous sections of this report, current technical data activity has been characterized, key problems and opportunities identified, informed judgments concerning national systems aspects marshalled, and specific policies and actions recommended. A national data program that embodies these recommendations has been articulated and a phased plan for its implementation has been structured. If this plan is placed in effect promptly and supported at the level recommended, it can be expected to yield operational data-system activity, nationally significant in its volume and character, by 1975.

Within the Federal R&D program activity,* the means exist to create virtually overnight a major shaping force for future national data systems. It can be implemented by adoption of our recommendation that each agency allocate a designated minimum percentage of its budget to husbandry of the data generated by the program. Based on the suggested "tithing" criteria of 10% for basic research programs, 15% for applied research, and 5% for developmental programs, an activity level in the Federal agencies of approximately $1.35 billion per year would thereby be identifiable as related to the National Scientific and Technical Data Program approximately three years after the Program was initiated.

It should be appreciated that this effort level is a relatively modest fraction of the estimates others have made of data activities in Federal programs. For illustration, activities associated with general-purpose data collection are estimated as over $400 million annually: DOD data-activity costs are variously estimated as between $2 and $3 billion annually. The program and reporting accountability suggested in this recommendation therefore should impose minimal burdens on programs containing any reasonable present level of data-husbanding activity.

An estimated $71.8 million of "new" money, or about 0.4% of the R&D budget, will be required initially to fund the national-level planning, development, and support offices of the Program. Fundings should be expected to increase appreciably over the initial 6-year period covered by the recommended implementation plan. It is expected that a $100 million budget level would be reached by the sixth year and would then recede to the initial budget level which would then be maintained for a number of years. A breakdown of the budgetary allocations considered appropriate for the second year of the plan follows:

	$ Million
National Advisory Council	$2.5
Scientific Data Program Office	
(50%) Data system development program (matching funds)	16.0
(30%) Special data services	9.6
(10%) Supporting research	3.2
(10%) Administration	3.2
(100%) Total	$32.0

*Federal Funds for Research, Development, and Other Scientific Activities, Fiscal Years 1966, 1967 and 1968, Volume XVI: National Science Foundation, NSF 67-19.

Technical Data Program Office	
(40%) Data system development program	$ 8.24
(40%) Special data clearinghouse service	8.24
(10%) Supporting research	2.06
(10%) Administration	2.06
(100%) Total	$20.60
Data Systems Technology Program Office	16.70
TOTAL	$71.80

These activity levels are believed sufficient to establish a program activity that has some reasonable chance of producing meaningful development actions at the national level. We think it particularly important that the Program, from the beginning, begin to function as a facilitator of advancement in data-management activity in major data-producing and data-using institutions. It will be noted that the "new-money" level proposed is intended to facilitate such developments nationally, and also that it totals only 5% of the Federal data activities to be identified and accounted for by the National Scientific and Technical Data Program.

As a matter of somewhat incidental interest, initial funding for the Scientific Data Program was established as 0.5% of the basic and applied research budget of the Federal Government. The Technical Data Program fund was established as 0.2% of the development budget, and the Data Systems Technology Program as 0.1% of the total Federal R&D budget.

Over the 6-year development cycle projected, the functional efforts of the National Program will fluctuate appreciably in scale and character. Figure VI-D-1 reflects a current estimate of the magnitude and pattern of these changes through to the point characterized by continuously operating systems.

As inferred previously, the initial year or Phase I of the Plan will require only a small amount of funding. Such funds can probably be obtained from existing agency budgets for FY 1969. The second year of the Plan, FY 1970, would represent the initial year for specific funding for the Program. From this initial level, the amount of funds expended in the Program would increase much more rapidly than the funds budgeted specifically for implementation of the Plan. Some of these funds would be contributed by non-government participants in the form of matching contributions to program costs. A much larger increase would result from identification and coordination of other Federal Agency data management and handling

PHASE DESIGNATION	PHASE I	PHASE II	PHASE III	PHASE IV
PROGRAM ELEMENT	PLANNING		IMPLEMENTATION	
	NATIONAL DATA PROGRAM DEFINITION	NATIONAL DATA SYSTEM(s) CONCEPT DEFINITION	NATIONAL DATA SYSTEM DEVELOPMENT	NATIONAL DATA SYSTEM OPERATION
TIME SEQUENCE	1 YEAR	2 YEARS	3 YEARS	CONTINUOUS
RELATIVE LEVEL of EFFORT by FUNCTION				
CENTRALIZED PROGRAMMING*	1	>1	1	<1
PLANNING & COORDINATION	2	20	60	20
IMPLEMENTING & OPERATING	0	10	100	>500

*Arbitrarily designated as 1

FIGURE VI-D-1

RELATIVE EFFORT LEVELS IN THE FOUR PHASES OF THE RECOMMENDED NATIONAL DATA PROGRAM activities with the new elements of the National Data Program to be established in accordance with this Plan. The latter aggregation of existing Federal data activities will account for much of the increase in the level of implementing and operating effort shown for Phase IV in Figure VI-D-1.

2. Technical Factors: The discussions and recommendations of this volume employ the simple term "data" for what is clearly evident from Volume II as extremely dynamic and diverse congeries of fact-matter within the national and international patterns of scientific and technological activity. This diversity must be dealt with realistically as the National Scientific and Technical Data Program is implemented. Broadly, it constitutes an underlying reality that cautions against approaches that can founder by being too inflexible.

Some of the technical and economic factors that significantly affect implementation concepts are:

- Physical growth of a specific data body calls for progressive change in data-management methods. Technologically desirable "housekeeping" of the body may call for special assistance to the data-handling institutions implementing the transition from one stage to the next.

- Intellectual growth of a subject field requires a continual housekeeping effort on terminology, a responsibility traditional to the professional society. In addition, a field will sometimes go through a revolutionary reform in its conceptual structures, as the field of chemistry did when the phlogiston theory was overthrown, or when key methodologies such as instrumental chemical analysis began to replace "wet" methods. The work required to reform data languages and structures, and to generate new data bodies can be massive under such circumstances, and at times will justify implementing aid. Technology-transfer program activity is a sphere of current importance that may generate need for translational dictionaries and similar tools highly relevant to the data program.

- Growth or diversification of the population using a data body, or a shift in its demand patterns, is another factor that can generate new implementation requirements for data service. For example, high-technology practices have created new demands and also provide potential new support for more rigorously validated data resources and more responsive services than have been traditionally demanded.

3. Sociological Factors: Sociological factors affecting implementation are probably more controlling than the technical and economic factors. Probably the most powerful one, fear that new data programs will compete with or weaken existing systems, is too familiar to need elaboration here. There are, however, some we can identify as specific and probably important within the scientific and technological community existing today:

- The community does not, in general, differentiate the data activity and the potentials for improved data activities and systems, from information activity as a whole. Data-oriented discussions, however, have readily awakened awareness and frequently strong enthusiasm. We have the sense of a strong latent recognition of the potentialities in breaking new ground with data programs. The essentially fallow ground now existing requires at least a modest educational cultivation. However, a thoughtfully framed program dropped into the present near-vacuum may thereby gain adherents much more rapidly than if it had to find its place among a diffuse array of pre-existing, inevitably competitive activities.

- Major data programs will produce more conscious recognition within the community of the potentialities for technological use of scientific data. They should have the effect of promoting science-based technological approaches, and of stimulating work within the community on better-codified terminology linkage between scientific and technological languages. In inplementing the data program, important benefits should result by drawing the community into the language-codification activities required to create adequate intellectual control of the data-handling systems that will be developed.

- The typical scientist and technologist has not been accustomed to view technical tools such as his data resources as also being a national resource. In part, this is the result of the phenomenological nature of most scientific and technological data, whose nationally significant attributes principally associate with their management rather than their substance.

4. <u>Priority and Incentive Factors</u>: The supply of development support funds provided for the Scientific Data Program and the Technical Data Program can be expected to prove much smaller than the demands made on them for data program development cost sharing. This situation should promote sound and vigilantly administered support commitments. It should also constitute an important educational exposure for Program specialists whose advice would normally be sought when major system priorities are under discussion.

A source of genuine expertise must be developed during the early years of the Program to prepare for the crucial decisions when operational system priorities must be weighed. The complexity of these considerations has become strikingly apparent in this exploratory study: as more is learned, we expect further complexity to become evident.

We also expect that sound rationales can be developed for identifying the significant value elements of a proposed data system development program. Figure VI-D-2 displays a coarse characterization of the data activities in the science-technology fields examined in this project study. They have been characterized by factors we believe are significant when evaluating the benefits expected to accrue from data system development efforts and expenditures.

Refinement of methods for establishing priorities and optimizing return on data system development expenditures constitutes a critical requirement if the National Scientific and Technical Data Program is to be implemented efficiently. At least initially, however, the selection of areas for program development can probably be made on the basis of less systematic and more intuitive bases. For example, the competence of the individuals or organizations applying for Federal support could easily become the over-riding consideration, for currently high competence in data system development is exceeding scarce. In order to attract the best available data system development competencies, it may be desirable to publicize the Program under a designation such as Project Dataphore. This title would emphasize the increased functional utility for data which is the main objective for the proposed Program and National System Development Plan.

FIGURE VI-D-2

FACTORS RELEVANT TO ESTABLISHMENT OF PRIORITIES FOR NATIONAL DATA SYSTEMS DEVELOPMENT

Area of Scientific and Technical Data Activity	Characteristics of Data Activities: Economic				Ownership			Source		Usage				Processing					Amenability to Systems Management
	commercial value	generation cost	national security relevance	GNP relevance	private	proprietary	public	acquired	originally generated	uniquely applied	routinely applied	number of users	degree of interdisciplinary usage	obsolescence rate	degree of knowledge structuring	volume	existing systems orientation	compatibility with ADP format	
Aerospace Science & Technology																			
system development	M	M	H	H	x	x			x	x		H	M	H	M	H	H	H	H
space science	L	H	M	L		x	x		x	x		M	L	L	M	H	H	M	H
Agriculture and Food Technology	M	M	M	H		x	x	x	x	x	x	H	L	M	M	H	L	M	M
Astronomy	L	L	L	L			x	x	x	x		L	L	L	H	H	L	H	H
Behavioral Sciences	L	L	L	L			x	x		x	x	L	M	L	L	H	L	M	H
Biomedical Sciences																			
biomedical research	L	H	L	L		x	[illegible]		x	x		M	L	L	M	H	L	L	L
clinical research	M	H	L	L	x	x	x		x	x		M	L	M	M	H	L	M	L
clinical practice	M	H	M	L	x	x	x	x		x	x	M	L	M	M	H	L	M	L
epidemiological research	M	M	M	L		x	x	x	x	x	x	M	M	L	H	H	M	H	H
Chemistry and Chemical Engineering																			
basic chemical research	L	M	M	L		x	x		x	x		H	H	L	H	H	M	M	M
applied chemical research	M	H	M	M	x	x		x	x	x		H	H	M	H	H	M	M	M
chemical engineering	H	H	M	H	x	x		x		x	x	L	M	M	M	M	L	M	M
Environmental and Geosciences																			
atmospheric sciences	M	H	M	M		x	x	x	x	x	x	M	H	L	M	H	M	M	H
geosciences	M	H	L	M		x	x	x	x	x	x	M	M	L	M	H	M	M	H
hydrologic sciences	M	M	L	M		x	x	x	x	x	x	L	L	L	M	H	L	M	M
Electronics and Electrical																			

Chemistry and Chemical Engineering																			
basic chemical research	L	M	M	L		x	x		x	x		H	H	L	H	H	M	M	M
applied chemical research	M	H	M	M	x	x		x	x	x		H	H	M	H	H	M	M	M
chemical engineering	H	H	M	H	x	x		x		x	x	L	M	M	M	M	L	M	M
Environmental and Geosciences																			
atmospheric sciences	M	H	M	M		x	x	x	x	x	x	M	H	L	M	H	M	M	H
geosciences	M	H	L	M		x	x	x	x	x	x	M	M	L	M	H	M	M	H
hydrologic sciences	M	M	L	M		x	x	x	x	x	x	L	L	L	M	H	L	M	M
Electronics and Electrical Engineering																			
basic physical electronics research	L	H	M	L		x	x		x	x		M	H	M	M	M	L	L	L
electronics technology (hardware)	M	M	M	M	x	x		x	x	x	x	M	M	H	M	M	L	M	M
electrical power engineering	H	M	M	H	x	x	x	x	x	x	x	H	L	M	M	M	L	M	M
communication systems engineering	H	M	H	H	x	x	x	x	x	x	x	H	L	H	M	M	M	H	M
Health and Sanitary Engineering																			
pollution control	L	M	L	L		x	x	x	x	x	x	M	M	M	M	M	L	H	H
sanitary systems engineering	M	M	L	M		x	x	x		x	x	L	L	L	M	L	L	M	M
Materials Science and Engineering																			
materials development	M	H	M	L	x	x		x	x	x	x	M	M	M	L	H	M	H	M
materials application	H	M	H	M	x	x	x	x		x	x	H	H	H	M	H	M	H	M
Mechanical Engineering																			
mechanical power engineering	M	M	M	M	x	x		x	x		x	M	L	L	M	M	L	M	M
mechanical and machine design	M	M	M	M	x	x		x	x		x	M	L	L	M	M	M	M	M
machinery and equipment operations	M	M	M	M	x	x		x			x	M	L	L	M	M	L	L	L

(continued on next page)

LEGEND: H=High
M=Medium
L=Low

B

FIGURE VI-D-2 (continued)

FACTORS RELEVANT TO ESTABLISHMENT OF PRIORITIES FOR NATIONAL DATA SYSTEMS DEVELOPMENT

Area of Scientific and Technical Data Activity	Characteristics of Data Activities																		Amenability to Systems Management
	Economic				Ownership			Source		Usage				Processing					
	commercial value	generation cost	national security relevance	GNP relevance	private	proprietary	public	acquired	originally generated	uniquely applied	routinely applied	number of users	degree of interdisciplinary usage	obsolescence rate	degree of knowledge structuring	volume	existing systems orientation	compatibility with ADP format	
Military Science and Technology																			
equipment operation	M	M	H	M		x	x	x			x	M	L	H	H	H	H	H	H
equipment development	M	H	H	M		x	x	x	x	x	x	M	L	H	H	H	H	H	H
military operations management	L	H	H	M		x	x	x			x	H	L	H	H	H	H	H	M
Nuclear Sciences and Technology																			
nuclear physics	L	H	M	L		x	x		x	x		L	L	L	M	H	L	L	L
nuclear technology	M	M	H	L	x	x	x	x	x	x		L	L	L	M	M	L	L	L
Oceanography and Marine Technology																			
oceanographic research	L	M	L	L		x	x		x	x	x	L	M	L	M	H	L	L	L
maritime operations	M	M	H	M	x	x		x	x		x	M	L	M	M	M	L	L	L
fishery operations	M	M	L	M	x	x	x	x	x		x	M	L	M	L	H	L	L	L
oil and mineral exploration and production	H	H	M	H	x	x		x	x		x	L	L	M	M	H	L	M	M
marine-systems development	M	M	H	M	x	x		x	x	x	x	L	M	M	M	M	M	M	M
Pharmacology and Toxicology																			
laboratory research	L	M	L	L		x	x		x	x		M	L	M	L	H	L	L	L
clinical research	M	M	L	L	x	x			x	x	x	M	L	M	L	H	L	L	L
drug application	H	M	M	M	x	x		x	x	x	x	H	L	M	M	H	L	L	M

oceanographic research	L	M	L	L		x	x		x	x	x	L	M	L	M	H	L	L	L
maritime operations	M	M	H	M	x	x		x	x		x	M	L	M	M	M	L	L	L
fishery operations	M	M	L	M	x	x	x	x	x		x	M	L	M	L	H	L	L	L
oil and mineral exploration and production	H	H	M	H	x	x		x	x		x	L	L	M	M	H	L	M	M
marine-systems development	M	M	H	M	x	x		x	x	x	x	L	M	M	M	M	M	M	M
Pharmacology and Toxicology																			
laboratory research	L	M	L	L		x	x		x	x		M	L	M	L	H	L	L	L
clinical research	M	M	L	L	x	x			x	x	x	M	L	M	L	H	L	L	L
drug application	H	M	M	M	x	x		x	x	x	x	H	L	M	M	H	L	L	M
Physics																			
basic research	L	H	L	L		x	x		x	x		H	H	L	H	H	L	L	L
applied research	M	H	M	L		x	x	x	x	x		H	M	L	H	H	L	L	L
Discipline-Research Data Activities	L	H	M	M		x	x		x	x		H	H	L	H	M	L	L	L
Mission-Developmental Data Activities	M	M	H	H	x	x		x		x		M	M	M	H	M	L	M	M
Application-Product Data Activities	H	L	M	H		x		x			x	H	M	M	H	H	M	H	H
General Purpose Data Activities	L	M	M	L			x	x			x	L	M	L	L	H	M	H	H

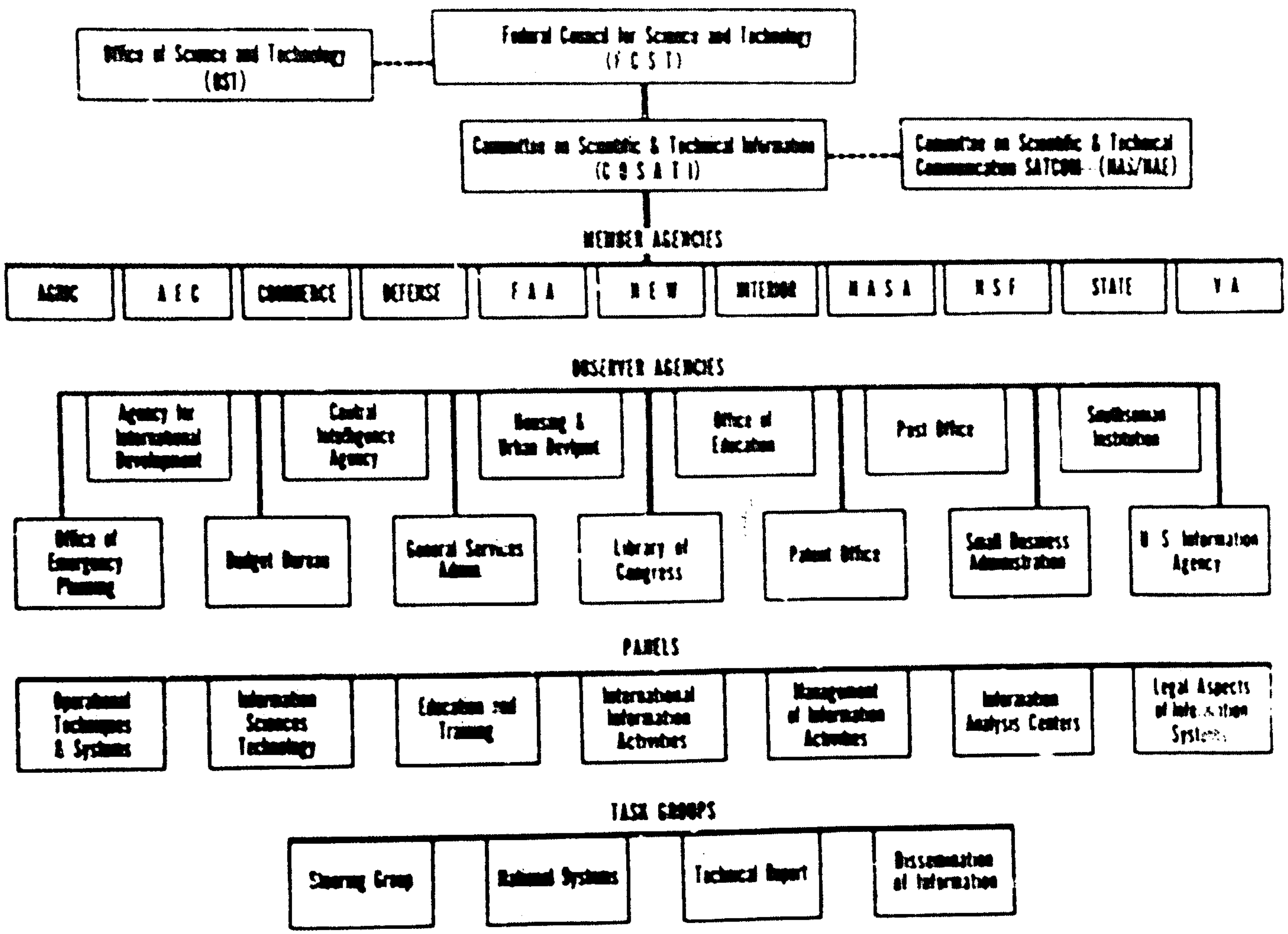
Legend: H=High
M=Medium
L=Low

B

EXHIBIT I-1

ORGANIZATIONAL STRUCTURE OF THE COMMITTEE ON SCIENTIFIC AND TECHNICAL INFORMATION

COSATI ORGANIZATION*



*Source "Progress of the United States Government in Scientific and Technical Communication," Committee on Scientific and Technical Information of the Federal Council for Science and Technology, Executive Office of the President. 1966.

EXHIBIT I-2

CHARTER FOR TASK GROUP
ON NATIONAL SYSTEM(S)
FOR SCIENTIFIC AND
TECHNICAL INFORMATION

FEDERAL COUNCIL FOR SCIENCE AND TECHNOLOGY

COMMITTEE ON SCIENTIFIC AND TECHNICAL INFORMATION

CHARTER

for
Task Group on National System(s)
for Scientific and Technical Information

GOALS AND OBJECTIVES

The Task Group will:

1. Undertake those investigations needed to (a) inventory and evaluate the resources (people, libraries and other services, equipment, materials and funds) currently being utilized in national and other domestic scientific and technical information activities, and (b) ascertain the information needs of users such as: scientists, engineers, managers, practitioners, and the technical public, as individuals and as groups, in and out of the government.

2. Based upon these and other findings, prepare recommendations and plans for the development of national information system(s) to include action for government agencies, suggestions for actions by the private sector, and steps to move from current to advanced information systems.

APPROACH AND SCOPE

The Task Group will undertake such studies as are necessary to provide requisite knowledge for its deliberations. Because of the varied interests and specialized knowledge of groups not directly represented on the Task Group, such as librarians, abstracting services, commercial publishers, and professional societies, it is the intent of the Task Group to call on representatives of such outside groups for information and suggestions. An illustrative list of problem areas includes:

1. Determine why and how the scientist, practitioner, manager, and the technical public assimilate and use technical information and identify trends that in prac-

tice and under certain environmental conditions may change these use patterns.

2. Examine the relationships between producers, processors, wholesalers, retailers, users, and systems of scientific and technical information. The study should seek to obtain such data as numbers of each type involved, size of operation, characteristics, trends, problems, economics, efficiency of effort, and education and training requirements. Both present and future aspects should be analyzed and evaluated.

3. Identify and evaluate a series of independent proposals for scientific and technical information systems presented in the last few years, considering for application those elements which appear to have immediate or future value for advanced information systems.

4. Analyze present and proposed national information systems which range from centralized to decentralized for costs, performance, resource requirements, impact, copyright and proprietary right problems, and methods of financing.

5. Examine other information systems in operation or under development of sufficient importance to the scientific and technical information community to warrant close coordination.

6. Consider the development of national information systems in relation to international scientific and technical information trends and patterns.

7. Review the state-of-the-art pertaining to equipments, facilities, techniques, organizations, as related to existing and potential national information system(s).

EXHIBIT I

PARTIAL GLOSSARY

FOR SCIENTIFIC AND TECHNICAL DATA ACTIVITIES

PARTIAL GLOSSARY FOR SCIENTIFIC AND TECHNICAL DATA ACTIVITIES

Many of the current difficulties in scientific and technical data management and system development efforts result from an inability to conduct precise and effective communications concerning the subject. Individuals, offices, and organizations frequently use a customized vocabulary with special connotations for each word. A means is needed to formulate an appropriate set of terms and definitions, and to encourage their acceptance and use throughout the scientific and technical community.

Although it was not an expressed objective of this study effort to formulate a scientific and technical data nomenclature, a partial vocabulary did evolve from the study. To a large extent, the terms had been previously conceived and documented by other individuals or groups. In addition, a limited number of terms and connotations were formulated to facilitate conduct and reporting of this study.

The following terms and definitions are representative of those informally developed during the conduct of the study. These and other definitions developed and used in the study are suggested as a focal point or basis for future development of a recommended nomenclature for scientific and technical data activities. Implementation of an effective nomenclature system would be an appropriate step toward a more systematic and effective management of scientific and technical data activities. The Department of Defense Technical Data and Standardization Glossary*, which incorporates many definitions from previous glossaries prepared by COSATI and other groups, was a primary resource in developing the following list.

* Technical Data and Standardization Glossary, Office of the Assistant Secretary (Installations and Logistics), Department of Defense, December, 1965. TD-2. 21 p.

PARTIAL GLOSSARY FOR SCIENTIFIC AND TECHNICAL DATA ACTIVITIES

APPLICATIONS DATA:
Data utilized in the production, operation, and maintenance of end items of equipment, material, products, and operating systems of all types.

APPLICATIONS-PRODUCT DATA ACTIVITIES:
The management and handling of data associated with the application of scientific and technical knowledge, material and/or techniques to the production, operation, and maintenance of end items of equipment, material products, and operating systems of all types.

CHARACTERISTICS (OF DATA):
Attributes which are germane to a given bit of data or are descriptive of a class of data. Typical characteristics of data include degree of refinement, (raw, reduced, evaluated, etc.), accuracy, precision, volume, rate of obsolescence, etc.

DATA:
Quantitative or qualitative representations of properties, characteristics, or attributes of objects, events, measurements, or observations.

DATA ACTIVITIES:
Any operations involving the management or handling of scientific or technical data. Data activities subsume both the formal and informal data efforts conducted to facilitate the generation, handling or use of scientific or technical data.

DATA-DOCUMENT:
A document which contains principally factual information or data, rather than conceptual information.

DATA EFFORT:
An organized activity which serves to facilitate the transfer of scientific and technical data from the generator to the user or from one to another of the intermediaries between the generator and the user. Functions of data efforts include collection, reduction, evaluation, transmission, extraction, storage, retrieval and dissemination of data.

DATA HANDLING:
The processing of data and its transmission from the source to the user. Data handling excludes the creation and use of data.

DATA HANDLING SYSTEM:
An assembly of procedures, personnel and equipment interacting to perform data operations such as recording, reduction, transmission, extraction, manipulation, storage, retrieval, formatting, and dissemination. Data handling operations are primarily conducted either to facilitate interaction and evaluation of

data by the generator of the data, or to facilitate the flow or transfer of the data from the generator to the user, either directly or indirectly.

DATA MANAGEMENT:
Those policies, procedures and actions used for coordinating and directing efforts to determine data needs, generate data, and handle data in a manner which permits optimal use and conservation. Data management is performed by all levels of participants in science and technology and is facilitated by data management programs and data handling systems.

DATA PROGRAM:
A plan or scheme of action designed for the accomplishment of a definite data management or data handling objective which is specific as to the time-phasing of the work to be done and the means proposed for its accomplishment.

DATA PROJECT:
Any identifiable study, task, component, system or program, directly applicable to data management and handling, such as data preparation, acquisition, storage, retrieval, reproduction, display, exchange, dissemination, utilization and system or program development or operation.

DATA RESOURCE:
The wealth represented by data; the aggregate supply or source of data.

DATA SCIENTIST:
A person informed in the field of data science who is capable of observing, measuring, and describing the behavior, properties, and flow of data; and who, through research, advances its understanding and use. The data scientist engages in data science *per se*; whereas a data specialist engages in data activities concerning a specialized subject.

DATA SPECIALIST:
A person primarily engaged in the management or handling of data in a particular field such as human engineering or solid-state physics. A librarian or documentalist, by way of contrast, devotes his efforts to document control and reference services.

DISCIPLINE RESEARCH DATA ACTIVITIES:
The management and handling of data associated with research which is primarily directed toward achieving greater knowledge or understanding of certain subject matter.

DOCUMENT:
A record of data, or of a concept, presented in any form from which information can be derived, e.g., a page containing data, a graphic representation; a tape recording, or a book.

FORMAL DATA EFFORT:
A data effort which has a recognizable structure, name, staffing or other definable attribute, and which functions to handle data by formal

means. Examples of formal data efforts include data collection networks, data-document depositories, etc.

FORMAT (OF DATA):
The mode of representation used in recording, storing, retrieval, transmitting or presenting data. Format defines the sequence of symbols or sets of symbols used to represent data.

GENERAL PURPOSE DATA ACTIVITIES:
The management or handling of data in a context which is neither an integral part of a scientific or technical effort nor serves as a direct support for specific scientific or technical efforts. Such activities frequently collect or organize data which subsequently finds application in scientific or technical efforts.

INFORMATION:
An elaboration of, description of, or extension of data. Knowledge communicated or received.

MISSION-DEVELOPMENTAL DATA ACTIVITIES:
The management and handling of data associated with the practical application of scientific or technical knowledge, material, and/or techniques directed toward a solution to an existent or anticipated technological requirement.

NATIONAL DATA RESOURCE:
The collective data wealth of a country, or its means of producing wealth in the form of data.

PRIVATE DATA:
Data for which the ownership rights are held by an individual or organization which prefers to retain the data in security, or strictly for private use.

PROPRIETARY DATA:
Data for which the ownership rights are held by an individual or organization who will sell or otherwise permit controlled use of the data. In many cases, proprietary data are copyrighted.

PUBLIC DATA:
Data which are in the public domain and can be used without consideration of ownership rights.

SCIENTIFIC DATA:
Data generated by research or other study employing the scientific method. Scientific data are normally generated by discipline-oriented research activity, but are applied in all phases of scientific and technological activity.

SCIENTIFIC DATA HANDLING SYSTEM:
A set of data handling operations by which data are processed and transferred in connection with discipline-oriented research activities. Scientific data handling systems facilitate the maintenance of the knowledge structure of a science or discipline. A major function of this system is validation of new measurements and establishment of relationships between new data and the pre-existing data base

TECHNICAL DATA:
Data generated as a result of the practical application of knowledge, material and/or techniques directed toward a solution to an existent or anticipated technological requirement.

TECHNICAL DATA HANDLING SYSTEM:
A set of data handling operations by which data are processed and transmitted from originator to the technologist user. It is by means of the technical data system that the coupling of basic research to development and application is achieved.

EXHIBIT I-4

SELECTED READING LIST

Science Communication
Washington, D. C. 200 07
COSATI Data Systems Study
Final Report - F44620-67-C-0022 30 April 1968

SELECTED READING LIST

During the past several years, relatively little study and examination has been directed specifically to scientific and technical data management and data handling systems. Rather, most existing study documents treat the broader questions of scientific and technical information systems. Consequently an individual desiring to familiarize himself with the current situation concerning scientific and technical data activities should read both the few documents specifically directed to data activities and a selected set of the documents which deal with the broader concept of information activities.

The following listing constitutes a set of documents which the reader could use to relate this Report to the views of leading data system specialists, findings of previous studies, current system development efforts, and future data system potentials. The listing consists of three sections--Part A, General; Part B, Descriptions of Current and Evolving Systems; and Part C, Data Systems--State of the Art.

A. General

1. Air Force/Industry Data Management Symposium Proceedings, Ballistics Systems Division, Norton Air Force Base, California. Conference held September 28, 29, and 30, 1965, Beverly Hills, California. AD 626 032.

2. American Institute of Physics, "Toward National Information Networks" -- a selection of reprints from Physics Today, Washington, D. C., Vol. 19, No. 1, January 1966, 15 pp.

3. American Library Association, The Library and Information Networks of the Future, Chicago, 1963. AD 401 347.

4. Auerbach Corporation, DoD User Needs Study Phase I, Vols. I and II, Philadelphia, Pa., May 14, 1965, Final Technical Report 1151-TR-3. AD 615 501 and AD 615 502.

5. Brock, Clifton, "The Quiet Crisis in Government Publishing," in College and Research Libraries, Vol. 26, No. 6, November 1965, pp. 477-489.

6. Cahn, Julius N., A System of Information Systems, U. S. Senate, Committee on Government Operations, Subcommittee on Reorganization and International Organizations, presented at Fourth Institute on Information Storage and Retrieval, American University, School of Government and Public Administration, Washington, D. C., February 12, 1962, 21 pp.

7. Committee on Scientific and Technical Information (COSATI) of the Federal Council for Science and Technology, The Copyright Law as it Relates to National Information Systems and National Programs--a Study by the Ad Hoc Task Group on Legal Aspects Involved in National Information Systems, Washington, D C, July 1967, PB 175 618.

8. Committee on Scientific and Technical Information (COSATI) of the Federal Council for Science and Technology, Progress of the United States Government in Scientific and Technical Information, Washington, D. C., 1966, 35 pp. PB 176 535.

9. Cuadra, Carlos A., Ed., Annual Review of Information Science and Technology, Vol. 1, New York, John Wiley and Sons, 1966.

10. DOD/NSIA Technical Information Symposium for Management Proceedings, Statler Hilton Hotel, Los Angeles, California, May 26-27, 1965.

11. de Solla Price, Derek J., "Communication in Science: The Ends--Philosophy and Forecast." Reprinted from Ciba Foundation Symposium on Communication in Science: Documentation and Automation, 1967, pp. 199-209. (Edited by Anthony de Reuck and Julie Knight. Published by J. & A. Churchill Ltd., London.)

12. Dunn, Edgar S., Jr., Review of Proposal for a National Data Center, Office of Statistical Standards, Bureau of the Budget, Executive Office of the President, Washington, D C., December 1965, Statistical Evaluation Report No. 6.

13. Graham, Warren R., Exploration of Oral/Informal Technical Communications Behavior, American Institutes for Research, Silver Spring, Maryland, March, 1967, 61 pp.

14. Hall, R. M. S., "The Development of the United Kingdom Data Program," Office for Scientific and Technical Information, State House, London, England, Journal of Chemical Documentation, Vol. 7, No. 1, February 1967, pp. 18-20.

15. Henderson, Madeline M., et al., Cooperation, Convertibility, and Compatibility Among Information Systems: A Literature Review, National Bureau of Standards, NBS Misc. Pub. 276, Washington, D. C., U.S. Government Printing Office, 1966.

16. Hornig, Donald F., "Communications and Civilization," Federal Council for Science and Technology, in IEEE Spectrum, May 1966, pp. 43-46. Address presented at the IEEE International Convention, New York, N.Y., March 23, 1966.

17. Hoshovsky, A. G., and Album, H. H., "Toward a National Technical Information System," Headquarters, Office of Aerospace Research, USAF, and Dept. of Mechanical Engineering, Stanford University, AFIT Program, in American Documentation, October 1965, pp. 313-322.

18. Huntoon, R D., The Measurement System of the United States, Institute for Basic Standards, National Bureau of Standards, Washington, D. C., 1966, 10 pp.

19. Kochen, Manfred, Ed., The Growth of Knowledge: Readings on Organization and Retrieval of Information, John Wiley and Sons, Inc., New York, 1967, 394 pp.

20. Knox, William T., "Toward National Information Networks: 1. The Government Makes Plans," Federal Council for Science and Technology, Committee on Scientific and Technical Information, in Physics Today, January 1966, pp. 39-44.

21. Knox, William T., "National Information Networks and Special Libraries," Office of Science and Technology, Executive Office of the President, Washington, D. C., in Special Libraries, November 1966, pp. 627-630.

22. Licklider, J. C. R., Libraries of the Future, Cambridge, Mass.: Massachusetts Institute of Technology Press, 1965.

23. Mikhaylov, A. I., et al., Organization of Scientific and Technical Information in the Communist World, Translation from the Russian, Aerospace Technology Division, Library of Congress, Washington, D.C., January 24, 1966, 88 pp. AD 627 802.

24. National Academy of Sciences-National Research Council, Communication Systems and Resources in the Behavioral Sciences, a Report by the Committee on Information in the Behavioral Sciences, Division of Behavioral Sciences, Washington, D.C., 1967, 67 pp. Publication 1575.

25. National Academy of Sciences-National Research Council, Division of Medical Sciences, "Communication Problems in Biomedical Research: Report of a Study," reprinted from Federation Proceedings, Vol. 23, No. 5, September-October 1964, pp. 1118-1176, and Vol. 23, No. 6, November-December 1964, pp. 1297-1331.

26. National Science Foundation, Scientific Information Activities of Federal Agencies, Washington, D.C., U.S. Government Printing Office (A series of reports published periodically under this title.)

27. North American Aviation, Inc., DoD User-Needs Study, Phase II: Flow of Scientific and Technical Information Within the Defense Industry, Final Report, Vols. I, II, and III, November 30, 1966, AD 647 111, AD 647 112, and AD 649 284.

28. Overhage, Carl F. J., and Harman, R. Joyce, Eds., INTREX - Report of a Planning Conference on Information Transfer Experiments, Massachusetts Institute of Technology, Cambridge, Mass., September 3, 1965, 276 pp.

29. President's Message on Communications Policy to the Congress of the United States. The White House, Washington, D.C., August 14, 1967.

30. President's Science Advisory Committee, Handling of Toxicological Information. Washington, D.C., The White House, June 1966, 21 pp.

31. President's Science Advisory Committee, Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information, The White House, Washington, D. C., January 1963 (The Weinberg Report)

32. President's Science Advisory Committee, Report of the Office of Science and Technology Ad Hoc Panel on Scientific and Technical Communications, The White House, Washington, D. C., February 1965, 21 pp. (The Licklider Report)

33. Rubinoff, M., Ed., Toward a National Information System, Second Annual National Colloquium on Information Retrieval. April 23-24, 1965, Philadelphia, Pa., Washington, D. C., Spartan Books, 1965.

34. Sackman, Harold, Computers, System Science, and Evolving Society: The Challenge of Man-Machine Digital Systems, New York, John Wiley and Sons, 1967, 637 pp.

35. Sullivan, Ralph H., and Dubester, Henry J., Data Collection Network of the U. S., Office of Science Information Service of the National Science Foundation, Washington, D. C., 1966, 18 pp.

36. Swanson, Rowena, Information System Networks. . . Let's Profit From What We Know, Office of Aerospace Research, Arlington, Virginia, June 1966.

37. System Development Corporation, Recommendations for National Document Handling Systems in Science and Technology, Appendix A--A Background Study--Volumes I and II, conducted for the Committee on Scientific and Technical Information, Federal Council for Science and Technology, November 1965, AD 624 560.

38. U. S. Congress, House of Representatives, Committee on Science and Astronautics, Government, Science, and International Policy, a Compilation of Papers Prepared for the Eighth Meeting of the Panel on Science and Technology, 90th Congress, 1st Sess., Washington, D. C., GPO, January 1967, 220 pp.

39. U. S. Congress, House of Representatives, Committee on Science and Astronautics, A Bill to Provide a Standard Reference Data System, Hearings before the Subcommittee on Science, Research, and Development, 89th Congress, 2nd Sess., HR 16897, Washington, D. C., GPO, June 1966, 181 pp

40. U. S. Congress, Senate, Committee on Government Operations, Report to the President on the Management of Automatic Data Processing in the Federal Government, 89th Congress, 1st Session, Senate Doc. No. 15, Washington, D. C., U. S. Government Printing Office, March 1965.

41. U. S. Congress, Senate, Committee on Government Operations, Summary of Activities Toward Interagency Coordination, Report No. 369, 89th Congress, 1st Session, Washington, D. C., U. S. Government Printing Office, June 1965. (The Humphrey Report)

42. U. S. Department of Commerce, Scientific and Technological Communication in the Government, Task Force Report to the President's Special Assistant for Science and Technology, Washington, D. C., April 1962, 81 pp. AD 299 545 (The Crawford Report)

43. Waddington, Guy, A World System of Evaluated Numerical Data for Science and Technology, Central Office, ICSU Committee on Data for Science and Technology, Washington, D C., Presented before the Division of Chemical Literature, Symposium on Compilations of Data on Chemical and Physical Properties of Substances, 152nd National Meeting of the ACS, New York, September 12, 1966, 14 pp.

44. Weisman, Herman M., "Needs of American Chemical Society Members for Property Data," Office of Standard Reference Data, National Bureau of Standards, Washington, D. C., presented at the Div. of Chemical Literature, Symposium on Compilations of Data on Chemical and Physical Properties of Substances, 152nd National Meeting of the ACS, New York, September 12, 1966, in Journal of Chemical Documentation, February 1967, pp. 9-14.

B. Description of Current and Evolving Systems

1. Belfour Stulen, Inc., Mechanical Properties Data Center Inventory Report 626, Data Storage Content of the Mechanical Properties Data File, Technical Information Systems Division, 2 pp., January 1967, AD 648 239.

2. Bisco, Ralph L., "Social Science Data Archives; Progress and Prospects", Council on Social Science Data Archives, Ann Arbor, Michigan, reprinted from Social Sciences Information sur les Sciences Sociales, Volume VI, 1 February 1967, pp. 39-74.

3. Bisco, Ralph L., and Glaser, William A., "Plans of the Council on Social Science Data Archives", reprinted from Social Sciences Information sur les Sciences Sociales, Volume V, 4 December 1966, pp. 71-96.

4. Brady, Edward L., "The National Standard Reference Data System", Office of Standard Reference Data, Institute of Basic Standards, National Bureau of Standards, Washington, D.C., printed in Journal of Chemical Documentation, Volume 7, 1 February 1967, pp. 6-9.

5. Dillon, E.L., and Nichols, C.W., "Handling of Statistical Well Data by Computer", reprinted from The Bulletin of the American Association of Petroleum Geologists, Volume 49, No. 9, September 1965, pp. 1520-1531.

6. Department of Defense, Department of Defense Engineering Data Retrieval System Plan, Joint Working Group for Planning Engineering Data Retrieval System, Dept. of Defense, April 1964, 71 pp.

7. Goldberg, Stanley A., Recommended Approaches to Design of the U.S. Army Engineering Data and Information System, Edgewood Arsenal, Maryland, December 1964, Technical Report No. 5, Report No. EDIS-2, AD 453 737.

8. Hackett, O. Milton, "National Water Data Program", Office of Water Data Coordination, U.S. Geological Survey, Department of the Interior, Washington, D.C. Reprinted from Journal of the American Water Works Association, July 1966, pp. 786-792.

9. Hager, George P., A Network of Chemical Information Handling, American Chemical Society: Division of Medicinal Chemistry and Division of Chemical Literature, September 1, 1964, 6 pp.

10. Interagency Data Exchange Program, IDEP-IV Program Summary: Scope, Objectives, Operation, July 1967, 14 pp.

11. Information Management, Incorporated, System Development Plan for a National Chemical Information System, A Report for the National Science Foundation, Burlington, Massachusetts, April 1967, 31 pp., AD 650 900.

12. Johnson, H. Thayne and Grigsby, Donald L., The Electronic Properties Information Center, Hughes Aircraft Company, November 1965, 80 pp. Contract AF 33(615)-1235.

13. Kahles, John F., Operation of the Air Force Machinability Data Center, 1965 Congress of the International Federation of Documentation (FID), Washington, D.C., 10-15 October 1965.

14. Klinger, Richard F., Annual Report of the Aerospace Materials Information Center, Air Force Materials Laboratory, Research and Technology Division, AFSC, WAFB, Ohio, February 1967, 43 pp., AFML-TR-67-#2.

15. Lemmon, Gene C., A Proposed Organization and System for Handling and Managing Air Force In-House Generated Test and Evaluation Technical Data and Information, Air Force Flight Test Center, Edwards Air Force Base, California, September 1966, 94 pp., AD 640 811.

16. Miller, James G., "EDUCOM: Interuniversity Communications Council", Pittsburgh, Pennsylvania, printed in Science, Volume 154, 28 October 1966, 6 pp.

17. National Aeronautics and Space Administration, PRINCE/APIC - The Parts Information Center, Marshall Space Flight Center, Huntsville, Alabama (1967), Brochure.

18. National Science Foundation, Nonconventional Scientific and Technical Information Systems in Current Use, December 1966, U.S. Government Printing Office, NSF 66-24.

19. National Advisory Committee on Research in the Geological Sciences, A National System for Storage and Retrieval of Geological Data in Canada, A Report by the Ad Hoc Committee on Storage and Retrieval of Geological Data in Canada, 1967, 175 pp.

20. Naval Fleet Missile Systems Analysis and Evaluation Group, Army, Navy, Air Force & NASA Failure Rate Data (FARADA) Program, Corona, California, Revised 1 June 1967, 20 pp.

21. Purdue University, Thermophysical Properties Research Center, Annual Report: 1966 for the Period January 1 to December 31, West Lafayette, Indiana (1967), 32 pp.

22. Schoenfeldt, Lyle F., The Project TALENT Data Bank, American Institutes for Research and University of Pittsburgh, Paper presented at the Third Technical Conference of the Council for Social Science Data Archives, Ann Arbor, Michigan, May 10-12, 1966, 20 pp.

23. Speight, Frank Y., and Cottrell, Norman E., The EJC Engineering information Program--1966-67: A Progress Report on the Role of Engineers Joint Council in Improving Dissemination of Engineering Information and Data, Engineers Joint Council, New York, New York, February 1967, 45 pp.

24. Speight, Frank Y., "Numerical Data Activities of Engineering Societies", Engineers Joint Council, New York, New York, presented before the Division of Chemical Literature Symposium on Compilations of Data on Chemical and Physical Properties of Substances, 152nd National Meeting of the ACS, New York, September 12, 1966, Journal of Chemical Documentation, February 1967, pp. 26-30

25. U.S. Army Materiel Command, Development of Coordinated Data Systems for Handling AMC Scientific and Technical Data and Information, Headquarters (Washington, D.C.), 1963, 6 pp.

26. U.S. Department of Labor, The BLS Information System - Background and Principles, Bureau of Labor Statistics, Washington, D.C., September 1967, 25 pp.

27. Van Meter, Clarence T., Lefkovitz, David, and Bedrosian, S.D., A Study for a Proposed Chemical Information and Data System, Institute for Cooperative Research, University of Pennsylvania, April 1965, 146 pp. Contract No. DA 18-035-AMC-288(A), AD 477 110.

28. Vette, J.I., The Operation of the National Space Science Data Center, National Aeronautics and Space Administration, Greenbelt, Maryland, October 23-27, 1967, 5 pp., AIAA Paper No. 67-946, presented at the American Institute of Aeronautics and Astronautics 4th Annual Meeting and Technical Display at Anaheim, California, October 23-27, 1967.

29. Weather Bureau, The National Meteorological Center, Washington, D.C., 21 pp.

C. Data Systems - State-of-the-Art

1. Agalides, Eugene and Swisher, Scott, "The Limitations in Biological and Medical Data Acquisition and Processing," General Dynamics/Electronics, Research Department, Rochester, New York; Medical School, University of Rochester, Rochester, New York, reprinted from Proceedings of the Conference on Data Acquisition and Processing in Biology and Medicine, New York, 1963, Pergamon Press, 1964, pp. 223-241.

2. Banzhaf, John F., III, "Copyrighted Computer Programs: Some Questions and Answers", Computer Program Library, New York, New York, in Computers and Automation, July 1965, pp. 22-26.

3. Baran, Paul, Communications, Computers and People, Rand Corporation, Santa Monica, California, November 1965, 20 pp., AD 624 431.

4. Baum, C., and Gorsuch, L., Editors, Proceedings of the Second Symposium on Computer-Centered Data Base Systems, System Development Corporation, Santa Monica, California, 1 December 1965, 303 pp., AD 625 417.

5. Bell, C. Gordon, Time Shared Computers. Carnegie Institute of Technology, Pittsburgh, Pennsylvania, May 15, 1967, 86 pp., SD-146, AD 655 380.

6. Berul, Lawrence, Information Storage and Retrieval: A State-of-the-Art Report, Auerbach Corporation, Philadelphia, Pa., September 14, 1964, 228 pp. AD 630089.

7. Bonn, T.H., "Mass Storage: A Broad Review," in Proceedings of the IEEE, Vol. 54, No. 12, December 1966, pp. 1861-1870.

8. Bourne, Charles P., Research on Computer Augmented Information Management, Stanford Research Institute, Menlo Park, Calif., November 1963, 49 pp. Tech. Documentary Rept. No. ESD-TDR-64-177, Contract No. AF 19 (628) 2914, AD 432 098.

9. Cahill, W.J., Thompson, D.W., Perkins, S.T., Howerton, R.J., A Computer-Oriented Neutron Data Storage and Retrieval System, Lawrence Radiation Laboratory, University of California, Livermore, Calif., September 23, 1966, 35 pp. TID 4500, UC 32, UCRL-50132.

10. Chase, J. David, "Searching for Physical Property Data," Celanese Chemical Co., Corpus Christi, Texas, as printed in Chemical Engineering, September 12, 1966, pp. 190-196.

11. Coggan, B.B., The Design of a Graphic Display System, University of California, Dept. of Engineering, Los Angeles, Calif., August 1967, 185 pp., Report No. 67-36, AD 658 314.

12. Committee on Scientific and Technical Information of the Federal Council for Science and Technology, Information Sciences Technology: First Report of Panel 2, September 1966, 9 pp.

13. Diebold, John, "What's Ahead in Information Technology," The Diebold Group, Inc., New York, N.Y., reprinted in Harvard Business Review, September-October 1965, pp. 76-82.

14. Drug Information Association, "Advances in Drug Information Processing," Proceedings of the Annual Meeting of the Drug Information Association, Volume 2, Chicago, 1966, 347 pp.

15. Edwards, Raymond A., "Time-shared Computers in Research," IBM Corporation, Data Processing Div., from Industrial Research, May 1966, pp. 63-72.

16. Fanwick, Charles, Trends in Computer Hardware, System Development Corporation, Santa Monica, Calif., March 17, 1966, 35 pp. SP-2393.

17. Fay, William T., and Hagan, Robert L., Computer-Based Geographic Coding for the 1970 Census, U.S. Department of Commerce, Bureau of the Census, Washington, D.C., October 1966, 13 pp.

18. Gold, Michael M., and Selwyn, Lee L., Toward Economical Remote Computer Access, Carnegie Institute of Technology and Massachusetts Institute of Technology, July 1967, 13 pp. AD 657 783.

19. Goldberg, Murray D., Cross Section Data Compiling - Present and Future, Brookhaven National Laboratory, Upton, New York, presented at American Physical Society, Jan. 29, 1966, and prepared for the Neutron Cross Section Technology Conference, Washington, D.C., March 22-24, 1966, 9 pp.

20. Goodman, Clark, Facsimile Transmittal of Technical Information, Houston Research Institute, Inc., Houston, Texas, May 31, 1965, 61 pp. Presented to National Science Foundation, Washington, D.C.

21. Grove, Alexander C., "International Standardization--Interface with the Future," from IEEE Spectrum, August 1966, pp. 91-101.

22. Hobbs, L.C., "Display Applications and Technology," in Proceedings of the IEEE, Vol. 54, No. 12, December 1966, pp. 1870-1884.

23. Jackson, Kingsbury T., A Study of the Application of Present and Future Methods of Automation, Retrieval and Portrayal to Department of Defense and NASA Engineering Documentation Systems and Centers, University of Alabama, AD 417 583, 221 pp., 1963.

24. Informatics, Inc., A Study of the Remote Use of Computers, Bethesda, Md., September 1966, prepared for National Bureau of Standards, Contract CST-313, PB 175664.

25. Jones, Arthur E., "Sharing Communication Networks," condensation of paper presented to the Science-Technology Division's Nuclear Science and Engineering Sections and the Metals/Materials Div. at the 56th Special Libraries Association Convention, June 8, 1965, Philadelphia, Pa., in Special Libraries, December 1965, pp. 705-808.

26. Kasher, Asa, Data-Retrieval by Computer: A Critical Survey, The Hebrew University, Jerusalem, January 1966, 73 pp., Contract N62558-4695, NR 049-130/6. Technical Report No. 22.

27. Landis, Daniel, Slivka, Robert M., and Jones, James M., et al., Evaluation of Large Scale Visual Displays, The Franklin Institute Research Laboratories, April 1967, RADC-TR-67-67, Final Report, AD 651 372.

28. Lawlor, Reed C., "Information Retrieval and Copyright Law Revision," prepared for presentation at the Third Technical Conference of the Council of the Social Science Data Archives, Ann Arbor, Michigan, May 10-12, 1966, Reprinted from Social Sciences Information sur les Sciences Sociales, Feb. 1967, pp. 75-85.

29. Licklider, J.C.R., The System System (AFOSR-1673) and Bridges over the Gulf between Man-Machine-System Research and Man-Machine-System Development (AFOSR-1127), Bolt, Beranek, and Newman, Inc., Cambridge, Mass., January 1962, 30 pp., AD 424 284.

30. Ludwig, George H., Advanced Space Information Systems, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Md., April 1967, 20 pp., N67-26587.

31. Markus, John, "State of the Art of Computers in Commercial Publishing," McGraw-Hill, Inc., New York, in American Documentation, April 1966, pp. 76-88.

32. McLaughlin, Curtis P., Computer Aids for Aerospace Design and Engineering: Innovation, Economics and Public Policy, The Rand Corporation, Santa Monica, California, July 1967, 53 pp., Contract F44620-67-C-0045, AD 657 009.

33. Merritt, Richard I., and Lane, Robert E., "The Training Functions of a Data Library," Yale University, Political Science Research Library, New Haven, Connecticut, 1966, Yale Papers in Political Science, No. 23.

34. National Science Foundation, Current Research and Development in Scientific Documentation, No. 14, 1966, U.S. Government Printing Office, NSF 66-17.

35. Nisenoff, N., "Hardware for Information Processing Systems: Today and in the Future," in Proceedings of the IEEE, Vol. 54, No. 12, December 1966, pp. 1820-1835.

36. Nolan, J.F., and Armenti, A.W., An Experimental On-Line Data Storage and Retrieval System, Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, Mass., Feb. 3, 1965, 36 pp., Tech. Rept. 377, AD 615 658.

37. Park, Ford, "The Printed Word," from International Science and Technology, Jan. 1967, pp. 24-44.

38. Peacock, Andrew C., et al., "Data Processing in Clinical Chemistry," National Institutes of Health, Bethesda, Md., Jan. 11, 1965, reprinted from Clinical Chemistry, Vol. 11, No. 5, May 1965, pp. 595-611.

39. Petersen, H.E., and Turn, R., System Implications of Information Privacy, The Rand Corporation, Santa Monica, Calif., April 1967, 43 pp. Prepared for presentation at the Spring Joint Computer Conference, Atlantic City, April 17-19, 1967. P-3504, AD 650 847.

40. Reed, Lawrence E., Advances in the Use of Computers for Handling Human Factors Task Data, Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, April 1967, 16 pp., AMRL-TR-67-16, AD 656 701.

41. Pomeroy, Paul W., Operation of a Long Period Seismograph Network Utilizing Magnetic Tape Recording and Analysis of Data Recorded on Magnetic Tape, Lamont Geological Observatory, Columbia University, Palisades, New York, July 1, 1967, Contract AF19(628)-5058, Project No. 8652, Final Report, AD 657 320.

42. Scherr, Allan Lee, An Analysis of Time-Shared Computer Systems, Massachusetts Institute of Technology, Project MAC, Cambridge, Mass., Doctorate Thesis, June 1965, 176 pp., MAC-TR-18, AD 470 715.

42. Shaw, Christopher J., Computer Software: Developments and Recent Trends in Programming and Operating Systems, System Development Corporation, Santa Monica, Calif., August 26, 1966, 18 pp., SP-2582/000/00, AD 640 655.

44. Stevens, Don, "A Survey of Optical Character Recognition," Philco Corp., Willow Grove, Pa., in The Electronic Engineer, December 1965, 5 pp.

45. United Aircraft Corporate Systems Center, Graphic Data Handling Techniques: Final Technical Report, Farmington, Conn., June 1967, 84 pp., Contract DA-44-009-AMC-1831(X), AD 659 807.

46. U.S. Congress, House of Representatives, Automatic Data Processing Equipment: Economy of Operations, Goddard Space Flight Center, Staff Study for the Subcommittee on Space Science and Applications of the Committee on Science and Astronautics, 89th Congress, First Session, 1965, 54 pp.

47. Ware, W.H., Future Computer Technology and its Impact, The Rand Corporation, Santa Monica, Calif., March 1966, 29 pp., AD 631 941.

48. Warshawsky, Erwin H., "Design by Computer," Mesa Scientific Corp., in Industrial Research, October 1965, pp. 55-60.

49. Webb, Glenn C., Data Transmission System Performance Prediction, Bunker-Ramo Corp., Canoga Park, Calif., Nov. 1966, 39 pp., Contract DA-36-039-AMC-00954(Y), AD 651 268.

50. Ziehe, T.W., Data Management: A Comparison of System Features, Tractor Incorporated, Austin, Texas, October 1967, 36 pp., AD 661 861.

Science Communication
Washington, D. C. 200 07
COSATI Data Systems Study
Final Report - F44620-67-C-0022

30 April 1968

EXHIBIT IV-1

SPECIMEN SURVEY INSTRUMENTS
AND RESPONSE

Science Communication, Inc.
Washington, D. C. 20007 Tel FEderal 3-1343
1081 Wisconsin Ave., N.W.

Dear :

We are asking your cooperation in a broad analysis and planning study covering scientific and technical data activities of national importance. Science Communication, Inc. is conducting this study for the Committee on Scientific and Technical Information (COSATI), of the Federal Council for Science and Technology. The Committee operates under the chairmanship of a staff member of the Office of Science and Technology, Executive Office of the President.

An important product of our study will be a time-phased plan for use by the COSATI Task Group on National Systems for Scientific and Technical Information. The purpose of the plan will be to formulate these policies and actions that will facilitate development of adequate national systems for scientific and technical data management. Through this plan, we hope to benefit the interchange of technological know-how and the conduct of research and development. Two desirable by-products of development of the plan will be (1) a clarification of the role that scientific and technical data, in various stages of refinement, play in the technical decision process; and (2) an assessment of the amount of attention devoted to data on the national level. The enclosed Statement of Work for our study indicates the scope and areas of emphasis which the time-phased plan must accommodate.

Prerequisite to our preparation of the time-phased plan is the selection of important issues or problems which must be resolved before national data systems can be realized. The generation of candidate issues occurred during a previous phase of our project when we conducted workshop discussions, visits to data centers, and mail surveys. We have selected those problems and issues which seem most valid and important, and grouped them in six categories. It seems imperative at this point that we verify and evaluate these problems by consulting experts with extensive knowledge of the subject areas involved. Therefore, we have chosen a panel of experts to evaluate each group of problems. The panels are as follows:

Science Communication
Washington, D. C. 200 07

Panel A-1 - Data management requirements
Panel A-2 - Data systems requirements
Panel B - Media, forms, software, and artifacts for packaging and transmittal of data
Panel C - Equipment for handling data
Panel D - Institutional roles in data management
Panel E - Educational and training requirements for future data handling

Because of your demonstrated knowledge of data systems requirements, we are asking you to help by participating in Panel A-2. It is requested that you participate as an individual rather than as a spokesman for any organization or group. Although we realize that your participation will involve a considerable expenditure of time and effort, we feel that your responses will greatly enhance the quality of our recommendations to the Committee.

We hope to find areas of consensus among panel members concerning the problems presented and possible resolutions. We also want to identify additional problems, and to acquire informed forecasts of trends and future developments which must be considered in planning data systems.

We hope that you will find the questionnaire thought provoking, and that the results of our study will prove beneficial to you and your colleagues in the near future. The following sheet provides a few instructions to guide you in completion of the questionnaire. If you should need any clarification or instruction, please write or call me (202-333-1343).

A primary purpose of our study is to develop a discourse with knowledgeable individuals and groups concerned with data management and data systems. We, therefore, request that you augment your response to the enclosed questionnaire with other comments or observations which you feel would contribute to achievement of the study purposes.

Sincerely,

SCIENCE COMMUNICATION, INC.

B.K. Farris

B. K. Farris
Vice President

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67093
Approval Expires March, 1963

INSTRUCTIONS FOR COMPLETION OF PROBLEM EVALUATION AND FORECASTING QUESTIONNAIRE

The purpose of this questionnaire is threefold: (1) to obtain your opinion regarding national data handling problems and recommendations for their resolution, (2) to identify other problems relevant to planning national data systems, and (3) to learn your views concerning future developments which will influence the development of national data systems. We need your opinions concerning the specific aspect of data systems identified on the attached questionnaire. Other panels will address themselves to the other aspects of data systems.

The following five-step procedure will move you quickly through the questionnaire and help you provide the responses we need.

1. Read each problem statement and feel free to edit, criticize, or rewrite it. If you run out of writing space, please use the back of the page.

2. Indicate in the spaces provided your judgment of the relevance of your experience to each problem statement, how important the problem is to the planning of national systems and the amenity of the problem to resolution within the present-day political and technological context.

3. Read the recommendation given for resolution of each problem. These recommendations do not necessarily represent the viewpoint of our Project sponsor or the position of Science Communication, Inc. We have conjectured them to evoke constructive criticisms, which we ask you to write down in the space following each recommendation. You may choose to make an alternate recommendation or modify or criticize the conjectured recommendation.

4. In the space provided on the page following our problem statements and recommendations write down any additional problems that you think pertain to that specific aspect of national data handling which your panel is considering. Then follow the same evaluation and recommendation procedure as indicated above.

5. Finally, on the last page of the questionnaire, answer the question concerning future developments that are relevant to plans for future national data systems. In providing your forecasts or predictions, please give projected dates for the possible developments (e.g., 1975, 1990, 2050). You may find it desirable to qualify your projections by identifying relevant assumptions.

PART I - STATEMENT OF WORK AND PERIOD OF PERFORMANCE*

A. RESEARCH

1. The Contractor shall furnish scientific effort during the period and at the level indicated in paragraph 2, together with all necessary related services, facilities, supplies, and materials, to conduct the following research:

a. Conduct studies which will embrace at least three forms of data pertinent to science and technology, as follows:

(1) Data acquired in the course of conducting experiments or examining natural phenomena, or in the course of performing tests according to prescribed procedures.

(2) Data which describe the characteristics or performance of a natural phenomenon, a material, a device or a component.

(3) Data which instruct, guide or aid skilled or semi-skilled persons in the proper use, maintenance or replacement of artifacts, or in techniques and procedures.

These data may be embodied in any physical format, from magnetic tape through standard reference texts and handbooks to programmed instruction or other manuals.

b. Establish how the various types and forms of data are acquired, stored, retrieved, packaged and disseminated for various specific types of users, why these packaging methods have been adopted, and what changes in storage, retrieval, packaging or dissemination of data are foreseen in the near future.

c. Place special emphasis on uses made of data by various functional groups (e. g., research, design, quality testing, product application, etc.) and the degree of processing or refinement of data needed for such functional groups.

d. Develop a preliminary census of data efforts in industry, the professions and government to guide the formation of national policy with respect to data collection, reduction, storage, retrieval, analysis and dissemination. This census, and a time-phased plan for the remainder of the study, will comprise the effort to be the first phase of the study, and the subject of this procurement.

*Air Force Office of Scientific Research Contract F44620-67-C-0022.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

ISSUE EVALUATION QUESTIONNAIRE

Panel A-2, Data System Requirements

PROBLEM STATEMENT: The current data service requirements of scientists and engineers are largely undefined. Additionally, effective methods are not available for predicting future data requirements. If the preceding is accurate, how can the required functions and scopes of national data systems, which are intended to service the needs of scientists and engineers, be determined?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☑	Easily Solvable	☐	Highly Relevant	☑
Secondary Importance	☐	Difficult	☑	Relevant	☐
Not Important	☐	Impossible	☐	Not Related	☐

POSSIBLE RECOMMENDATION: Prototype data systems should be used to test user response and to make other measures of the effectiveness of specific system operations and service concepts. Existing prototype data systems (National Standard Reference Data System, National Oceanographic Data Center, National Space Sciences Data Center, etc.) should be given support for developing and testing methods of identifying service needs, and for developing and testing means to measure the effectiveness of specific system operations in satisfying these needs. Additional prototype systems should be implemented in other areas of science and technology. This would permit development and testing of methods applicable to determining the data service requirements of the many diverse communities of scientists and engineers. Prototype systems should be implemented in typical work environments rather than in experimental information science laboratories.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

I would agree with the recommendation that prototype data systems should be used to test and evaluate user requirements. However, my orientation toward the solution to the problem is a little bit different. I would stress the point that prototype systems now in existence are of several kinds. The first kind are those which have been set up by administrative action in response to a rather broad range and loosely-defined sense that something should be done. The National Standard Reference Data System is an example of this. On the other hand, certain other systems or centers have been set up by a mission-oriented operation within a Federal agency or elsewhere (including privately-sponsored in-house information systems) where the authoritative body knew explicitly what it expected of the system and how it expected it to be done.

That is, there was a clear mission need related to a well-stated mission problem. A third set of system elements includes those which have been established within very narrow scientific or technological limits which contain a rather small but explicit set of users. Here the system has been set up in response to the clearly stated needs from a small body of users. In many cases, the people who are running such a small center or system are themselves users of the data. And they know exactly what their audience wants because they are members of their own audience. In the last category I include the JILA Information System at Boulder, Colorado.

All three of these kinds of systems can in fact be used to explore the requirements which would be placed on national systems.

I agree with the recommendation which has been proposed in general, but I would like to stress that scrutiny of the existing systems should involve two quite different kinds of analysis. The first is an analysis of how the system relates to a well-defined body of users all by itself. Such a test is particularly applicable to the second and third types of systems which I categorized.

The second point is that there should be careful thought and analysis given to the situation which is bound to occur when two systems with well-defined bodies of users discover or develop an interface which in effect makes them part of a larger system. The question here is, how do they relate to one another across the interface and how do they provide services to each other's users.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

PROBLEM STATEMENT: The data activities for each phase of scientific and technical activity (e. g., basic research, technological development, application operations, etc.) differ so much that it is difficult to envision a single system servicing all phases of scientific and technical activity even if its coverage was limited to a single discipline such as chemistry. If national data systems are established, how could these differences be handled, especially if the systems are structured by discipline or technical field?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☐	Easily Solvable	☐	Highly Relevant	☒
Secondary Importance	☒	Difficult	☒	Relevant	☐
Not Important	☐	Impossible	☐	No. Related	☐

POSSIBLE RECOMMENDATION: Large-scale or national data systems should not be structured on a discipline basis. They should be structured instead to cover related sets of properties (e. g., thermodynamic, electrical, etc.), whereas the systems to serve communities applying science and technology should be structured to cover related sets of substances or items. Initial priority should be given to the property-oriented data systems, which would be capable of serving the needs of large populations of researchers in different research fields.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

The problem statement here does recognize an important issue. In the Office of Standard Reference Data we were very seriously concerned for quite a long time with ways of structuring the subject matter content of data of the physical sciences. We had contract studies made on this subject, we attempted to analyze it ourselves, and we looked carefully into the previous analyses which had been made by other bodies. Of particular note is the study which was made by the Office of Critical Tables in attempting to develop a property list. I myself came to the conclusion that the problem of defining the universe in clear and unambiguous terms could not be solved. However, I think that the problem is susceptible to resolution because I don't think an exact closed solution is necessary. Along these lines I turn now to the recommendations which the questionnaire provides. I agree with the statement that large-scale or national data systems should not be structured on a discipline basis. However, I do not agree that they should be structured to cover related sets of properties uniquely. The reason I say this is that it is apt to lead into the trap of dividing up the universe into neat boxes and then finding that the boxes don't fit the universe. I agree that priority should be given to property-oriented data systems but I am not sure that even they are capable of serving the needs of large populations of researchers in different research fields.

I would propose the following alternate recommendation. I would suggest that the grand plan for structuring of large-scale or national data systems should not be superimposed from above, but instead the system should base itself on those specialized and knowledgeable data centers and information system activities which prove themselves capable of serving small or medium sized groups of users. The structure which the national system would place over these would be flexible, inter-connected, susceptible to change, and above all, redundant in its information paths. The only requirement I would place on a national system would be that it should be capable of accepting an inquiry from a user even when the inquiry was directed to the wrong component of the system, and somehow processing the inquiry to a communication point where the inquiry could be handled appropriately. There are several ways that this can be done. The first would appear to be merely to accept inquiries from anybody by any component of the system and shuffle them up to a headquarters activity for sorting unless the recipient himself was clearly aware that he was the right person to answer the inquiry. If the inquiry then reached the central point it could be switched down to an appropriate sub-unit. This means that somewhere in the system there is a fairly competent referral center. The second aspect of the solution is to provide easy access to the entire content of the system and trust to the inquirer to do his own retrieval. This merely means in the simplest case that you have large sets of books and indexes to the contents of the books. Then you make the books readily available and the person who wants to find the answer can go through until he finds the right book and the right answer. If he doesn't find an answer he can be fairly confident that the system does not contain the answer. This does not provide any mechanized interconnection of individual specific answers, but perhaps that can come later as sophistication builds up or perhaps the individual specific answer can be keyed in with a recognition of a place where competence in this area exists so that the inquirer, having found a specific answer, will know exactly where to go for a more complicated one.

Science Communication
Washington, D. C 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

PROBLEM STATEMENT: Each year, many more dollars and man-hours of effort are expended on equipment, product, and vendor service data activities than on scientific data activities. What should be the place of such data activities (equipment, product and vendor) in national data systems?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☒	Easily Solvable	☐	Highly Relevant	☐
Secondary Importance	☒ (scribbled out)	Difficult	☐	Relevant	☒
Not Important	☐	Impossible	☒	Not Related	☐

POSSIBLE RECOMMENDATION: Equipment, product and vendor data activities must be considered a vital part of any national data system. However, due to the vast scope, diversity, property right considerations, and other complicated legal, economic, and social factors associated with these data activities, centralized direction of system development for these data activities is unfeasible and of questionable value. Instead, initial attention should be directed toward upgrading of data activities in individual firms, followed by cooperative efforts within trade associations, manufacturing groups, etc. Increased effort should be directed to development of improved methods (e. g., computer controlled photocomposition of equipment catalogs, automated design programs, etc.) which can be applied to improvement of vendor data activities in a large number of industries.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

The problem posed, that of equipment, product and vendor service data, is indeed a significant one. The recommendation provided in the questionnaire does recognize the difficulties, but I am not confident that such proprietary operations should be explicitly included in a national system as such. Rather I would suggest that the national system confine itself to providing channels by which trade associations, manufacturing groups, etc., can develop effective and expeditious communication channels with their users. One aspect would be a computer-based index to equipment catalogs and to product and vendor data. I do not think that a national attempt to provide computer-controlled photocomposition of equipment catalogs is appropriate.

Science Communication
Washington, D C 200 07
Contract F44620-67-C-0023

Budget Bureau No. 21-S6700
Approval Expires March, 1969

PROBLEM STATEMENT: Certain mission-oriented industry activities, such as in the food industry, are presently evolving from craft arts to high-technology activities. National data systems must, therefore, be designed to accommodate data from such activities and to serve the communities concerned. What allowances should be made for such services, both at the present and in the future?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☐	Easily Solvable	☐	Highly Relevant	☐
Secondary Importance	☒	Difficult	☒	Relevant	☒
Not Important + at present	☐	Impossible	☐	Not Related	☐

POSSIBLE RECOMMENDATION: The less highly developed industries and technologies provide especially attractive opportunities for development of effective data systems. Scientists and technologists in these fields have not yet established highly structured and institutionalized data acquisition and application methodologies. Consequently, these scientists and technologists would be more receptive to trial of new methods and systems. In addition, if data systems evolve simultaneously with these technologies, the systems could contribute substantially to structuring of the knowledge bases required for rapid technological progress. Rapidly developing industries and technologies should be continuously monitored for identification of situations offering high promise for the introduction of a coordinated data system. Trade associations and industrial cooperatives, as well as mission-oriented Government agencies, should be encouraged to formulate and implement such systems.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

The problem is a good one, well-stated. I disagree with the recommendation only in certain specific aspects. I do not feel for example that these less highly-developed industries and technologies do provide especially attractive opportunities for effective data systems. The reason I disagree is that I think it would be extremely difficult to develop effective data systems for these industries because the industries themselves don't know what kind of data or what kind of data systems they need. Rather, I would suggest that the appropriate way to provide assistance for such industries and technologies is to make sure that the basic scientific and technological data which are known to be useful in industrial applications and research programs should be made extremely well accessible to these people, and that the scientists in them should be provided with assistance in reaching in to areas of data where they are themselves not familiar with the scientific approach or technology or phraseology involved. I think the best basis for national data and information systems is in the well-defined sciences and technologies not in the ones which are changing so rapidly in their orientation. By way of a specific example, let me point out that until a few years ago, there was no thought at all that food processing technology would have any concern with high energy radiation. Any attempt to develop an information or data system for that industry would have failed miserably and in fact would have probably inhibited the utilization of radiation as a food preservative technique because there would be the clear implication that it was not relevant. I am sure that any system we set up now for these highly evolving areas would be inadequate to their needs of two years hence.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67005
Approval Expires March, 1968

PROBLEM STATEMENT: Traditional means of informal communication at technical meetings and via direct correspondence or discussion among scientists and engineers perform some useful functions which cannot be easily assumed by formal data systems. What should be the function of informal communications in national data systems and how can these functions be best coordinated with highly structured data systems?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☐	Easily Solvable	☒	Highly Relevant	☒
Secondary Importance	☐	Difficult	☐	Relevant	☐
Not Important	☒	Impossible	☐	Not Related	☐

POSSIBLE RECOMMENDATION: Informal systems are used, essentially, to communicate conceptual information of a current awareness nature. Informal communications provide the lubricant at the biting edge of scientific and technological advances. In contrast, structured data constitute the machine tool for scientific and technological advance. Increasingly large quantities of scientific and technical data are being jointly used by groups of workers. Such joint use of data constitutes a communication system, in a very restricted sense. As more data files are structured and made directly accessible to workers, the volume of communication conducted via the data files and related access tools will increase rapidly. Informal communications will continue to perform social and motivational functions.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

The problem statement is appropriate because it recognizes the existence of these informal communication channels. I agree with the recommendation that the informal channels are highly valuable and that they will continue to perform social and motivational functions. However, I would go further and say that these informal communication channels will continue to be essential in many ways and the formally based highly structured data systems will benefit from whatever communication they provide. However, the informal channels cannot be built in to the formal systems without doing harm to the informal channels. I would suggest that they should be let alone and admired.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

PROBLEM STATEMENT: During recent years, much attention has been devoted to document handling systems (e.g., the Defense Documentation Center and the National Library of Medicine MEDLARS System). What should be the relationship between such document handling systems and data systems (i.e., systems that handle the factual information content of the document rather than the document itself)? Should operation of data systems be totally separate from that of document systems; should the two perform only complementary functions; or should they be totally integrated?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☐	Easily Solvable	☐	Highly Relevant	☐
Secondary Importance	☒	Difficult	☒	Relevant	☒
Not Important	☐	Impossible	☐	Not Related	☐

POSSIBLE RECOMMENDATION: The operations of existing data systems and document systems should be conducted so as to complement and supplement one another. For example, existing document handling systems should augment current indexing of conceptual content of documents to include adequate indexing of the data content of documents. Such indexing would facilitate identification of data for extraction and incorporation in data systems. Increasingly large quantities of useful data are not being published; consequently, data systems must also acquire input data from other sources. In fact, in the future it will often be desirable to by-pass publication of data and to transmit data from the point of measurement directly to the data system. The data system will perform many of the functions now served by publication (i.e., exposure for review and verification by colleagues, dissemination for use, and recording for archival or reference purposes). Therefore, data systems will in the future tend to supplant document systems, especially for archival purposes, for bench or console-level services to the technologist, and, to a lesser extent, for the scientist.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

The problem provides appropriate recognition of the difference between document handling and data systems. I agree in general with the recommendation, specifically with the first part of it. I disagree that data systems will in the future tend to supplant document systems especially for archival purposes. Rather I feel that there will always be a need for archival handling of the documents themselves. The most profound analysis of previous work is performed by scientists who go to the original papers themselves and perhaps back to the scientist who wrote the paper. Archival storage of documents is a necessity and data systems will not obviate the need for reference to the documents. However, for the bench and console level services to the technologist and to the scientist outside of his field, I agree that the data systems will be more useful than the document systems. I also agree that the two kinds of systems do provide complementary functions, they should not be totally integrated, and that increased efficiency in each will benefit the other.

One very difficult problem remains before this general problem could be resolved. As the scientific literature grows the document system as an archival storage will become increasingly unwieldy and access to the documents will be a very complicated process. Adequate indexing of all the content of a document is extremely difficult and it is not easy to see what archival system will be appropriate. I am not talking about the mechanical storage of the material itself because there are technological advances which will certainly provide a solution to reduction in bulk. However, I am talking about the difficulties inherent in making the inquirer aware of which documents contain material which may be relevant to his problem. Author indexing by itself is not adequate and external indexing is becoming increasingly expensive. I do not believe that computer searching of whole text of each document can ever be feasible and I don't think it should be attempted. There's a real problem here, but I am not the right person to recommend solutions.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

PROBLEM STATEMENT: It is unquestionably much easier to develop systems which provide the user the location of data rather than deliver data to the user. To what extent should data services be rendered by a referral center or network rather than a data retrieval and dissemination network or system?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☐	Easily Solvable	☒	Highly Relevant	☐
Secondary Importance	☒	Difficult	☐	Relevant	☒
Not Important	☐	Impossible	☐	Not Related	☐

POSSIBLE RECOMMENDATION: Referral centers and networks offer a logical stepping-stone from our current uncoordinated data efforts to future, more highly integrated data systems. In fact, even after highly integrated data systems are developed, a mechanism similar to a referral network will be required to direct inquiries to the location where the response data are available. The existing National Referral Center at the Library of Congress should be supplemented with specialized referral centers in specific areas of science and technology (e.g., engineering materials). Each specialized referral center should maintain indexes of scientific and technical data in the field served by the center.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

In my own thought the individual center covering a specialized area with a specialized body of users is the basic and key element of national systems. Therefore I agree with the recommendation that referral centers and networks offer a logical stepping stone and that they will continue to play an important role. I don't think a very highly integrated data system is ever going to emerge because I don't think we are ever going to become that clear in our scope and that unified in our way of attacking the problems. Therefore I see the referral centers and networks as being an essential part of all the future.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

PROBLEM STATEMENT: Users of scientific and technical data often wish to make retrospective searches. What criteria should be used for selection of the retrospective data to be incorporated into national systems?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☒	Easily Solvable	☐	Highly Relevant	☒
Secondary Importance	☐	Difficult	☒	Relevant	☐
Not Important	☐	Impossible	☐	Not Related	☐

POSSIBLE RECOMMENDATION: The rate of obsolescence of data varies considerably; consequently, where feasible the value of each type of data should be analyzed prior to input of back-logged data into a new system. In general, inputting of back-logged data does not appear highly desirable; rather, systems should concentrate on capture and use of current data. Exceptions to this general rule are useful data which cannot be regenerated (e.g., weather records, epidemiological data, etc.), data whose regeneration cost would considerably exceed the cost of maintaining the data in a system, or data whose utility can be substantially upgraded by incorporation into a more efficient system. In addition, existing data files, although perhaps they are not of substantial continuing value, can provide in some instances an economical test-bed for use in the structuring and testing of new systems.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

I agree with the problem and I disagree with the recommendation. It is true as the recommendation states that the rate of obsolescence of data varies considerably. However, I disagree as to the statement that systems should concentrate on capture and use of current data. There are many cases where it has been clearly indicated that older data are just as good and sometimes better than current data. It is true that systems which attempt to go back in time face a larger volume of material to cover and a more difficult job for that reason. However, the problem can be solved on a piece-meal basis by going back in time for selected key areas in which all of the past data can be surveyed, analyzed, and integrated with the present data. Any system which tends to overlook or brush aside the existing backlog of scientific data will be encouraging the unnecessary repetition of measurements already made and will therefore be inefficient in the long run.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67005
Approval Expires March, 1968

PROBLEM STATEMENT: There is increasing evidence that equipment developments are moving so rapidly in the information systems that they are controlling the structure of the automated data systems now being established. What can be done to assure coordination of the efforts of equipment and software suppliers with system requirements?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☐	Easily Solvable	☐	Highly Relevant	☐
Secondary Importance	☒	Difficult	☒	Relevant	☒
Not Important	☐	Impossible	☐	Not Related	☐

POSSIBLE RECOMMENDATION: Scientific and technical data system designers and users must define their requirements more explicitly. These requirements cannot be effectively satisfied by equipment and program languages designed for business data processing or for mathematical computations. Equipment manufacturers cannot develop optimal equipment to meet ill-defined, non-standardized system specifications. However, by analyses and prototype testing data system designers and users can systematically establish the functional characteristics of required equipment. In addition, data system designers must define, document, and publicize the current and future equipment market potential which exists in scientific and technical data systems. Equipment manufacturers and software firms have the basic capabilities required and can be expected to move quickly to meet economically valid equipment and programming requirements of scientific and technical data systems.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

The statement of the problem is valid. There is increasing evidence that equipment developments are moving so rapidly that the information system designers and information centers can't keep up. However, I think that the recommendation is pointed in the wrong direction. The simple fact is that we have another case here of unnecessary obsolescence. The equipment manufacturers are doing themselves and everybody else a disservice by making past equipment incompatible with present equipment. The information centers would do best to ignore the mad scramble toward more and more complex computers and concentrate on making their own systems readily machine-readable and flexible as to utilization. Whenever a small element of the system, that is a data center or information activity of narrow scope, gets itself well-established it should seize upon some flexible means of storage and retrieval of information and let the computer companies go ahead in their mad race. Eventually translation devices must and will be provided which will permit interconnection.
I think that it would be unfair to impose on the data centers and the system designers any requirement that they keep testing prototypes. They've got their own business to do and the equipment manufacturers should recognize that they've got to provide flexibility.

Science Communication
Washington, D. C. 20007
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

PROBLEM STATEMENT: Currently scientific and technical activities are conducted largely independently of the operation of formal data centers. More direct coupling of day-to-day work activity and archives of data is now technically feasible; exploitation of this possibility offers unprecedented potential for rapid and effective use of existing data. How can the potential of such systems be demonstrated?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☑	Easily Solvable	☐	Highly Relevant	☑
Secondary Importance	☐	Difficult	☑	Relevant	☐
Not Important	☐	Impossible	☐	Not Related	☐

POSSIBLE RECOMMENDATION: A small number of research and development projects or programs should be selected to test the applicability and effectiveness of automated data system concepts. In these test projects, all operations involving data would be automated and incorporated into a system serving the project. For example, a typical scientist or technologist working on the project would have direct access to several data files which would be used not only to facilitate his own work, but also to communicate with other members of the project team. Files directly accessible to the user should include the archival or reference files commonly found in data centers, as well as the frequently used work files often maintained either at the worker's desk or at the computing center. The operations of such projects should be carefully monitored and analyzed to identify methods applicable to other similar or larger scientific and technical program efforts.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

The problem is a good one and the recommendation has some valid points. However, I think that I would modify the recommendation and suggest that it is less important to put the scientist next to a fully automated data system, instead he should be put next to a flexible partially automated system which will include the services of some scientists and information scientists who can work with him to find out how he can best be served. This will be an education in both directions and will be much more beneficial than forcing a specific highly-automated system down the individual user's throat.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

PROBLEM STATEMENT: There is almost a complete absence of criteria for the evaluation of the economic performance of current data systems. Recently the concept of cost-effectiveness has gained wide acceptance as a basis for deciding whether or not a system would be developed or operated. Should cost-effectiveness be the principal criterion to determine whether future scientific and technical data systems will or will not be developed?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☑	Easily Solvable	☐	Highly Relevant	☐
Secondary Importance	☐	Difficult	☐	Relevant	☑
Not Important	☐	Impossible	☑	Not Related	☐

POSSIBLE RECOMMENDATION: The specific costs associated with data handling or with the development and operation of data systems should and can be assembled although such cost data are not currently known. The benefits accrued from data handling systems are difficult to quantify. Consequently, in the immediate future the primary emphasis of system development efforts should be on achievement of effectiveness. Cost optimization can then be achieved by selecting the more efficient system from the systems previously proven effective.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

It is difficult at best and usually impossible to do any honest and realistic cost-effectiveness studies of research. Our own experience in the Office of Standard Reference Data indicates that cost-effectiveness studies of data systems which are a support of research is equally impossible. It may be possible to apply cost-effectiveness and cost-optimization techniques to specific system elements, mechanical portions of information activities. However, I do not feel that any attempt to base system plans on cost-effectiveness can be productive.

Instead, I would suggest that user satisfaction and utilization of data system elements is a much better criterion and one which is far more easy to obtain. At the present time COSATI Panel #6 on Information Analysis Centers is concerning itself with effectiveness studies of information analysis centers, with reasons for closing information analysis centers, and related problems. During the forthcoming year we will probably have some conclusions to offer which are relevant to this problem.

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 1968

PROBLEM STATEMENT: Inability to define a single system structure applicable to the national requirements of data communication and use has been cited as a justification for not developing new or improved national data systems. What criteria should be employed in the planning of national data systems to determine whether centralized data systems, decentralized data networks, coordinated data exchange programs, data source referral centers, or no system should be implemented?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☑	Easily Solvable	☑	Highly Relevant	☑
Secondary Importance	☐	Difficult	☐	Relevant	☐
Not Important	☐	Impossible	☐	Not Related	☐

in concept. That doesn't mean that the solution comes automatically!

POSSIBLE RECOMMENDATION: The structures of data systems cannot be dictated by fiat from a top-level policy position. Rather, such structures must evolve from working-level responses to real needs. In fact, national systems are already developing in this fashion. The current need is for coordination and financial support of these evolving systems. Experimental data systems should be tested which tie into a network several of the systems and services which the scientist or engineer now must use separately. The experimental system components should include automated recorders, computing equipment, automated archives of relevant data, archives of computer routines, reactive display consoles, and automated report generators. Such experimental systems should be carefully monitored and evaluated.

CRITICISM OF THE RECOMMENDATION AND ALTERNATE RECOMMENDATIONS:

AMEN!

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-SC7003
Approval Expires March, 1968

FORECASTING QUESTION

Currently, much scientific and technical data are handled by informal means (telephones, conferences, etc.) as well as by formal systems such as data collection networks, data exchange cooperatives, government and commercial publishing, data centers, and data-document depositories (data libraries). Examples of relatively recent implementations of systems of national importance are the National Space Science Data Center and the National Standard Reference Data System.

Would you please give your opinion as to the formal systems of national significance which you expect to be implemented over the next 10 years. The list should include only those systems implementations which can be expected to significantly affect current practices for handling scientific and technical data. Please indicate the date when each system can be expected to be in operation.

Brief Description of Projected Formal System	Date When System Will Be Operational
Expansion and vitalization of National Referral Center with better national interconnections	1975
Emergence of the National Chemical Information System and a similar system for Physics	1975
Joint operation of EDUCOM, NSRDS, NCIS, NPIS	1980

Science Communication
Washington, D. C. 200 07
Contract F44620-67-C-0022

Budget Bureau No. 21-S67003
Approval Expires March, 196[illegible]

ISSUE EVALUATION QUESTIONNAIRE

YOUR PROBLEM STATEMENT: Within many Federal agencies, specialized data centers and information systems have been set up or are being set up in response to legitimate needs. Within private industries, similar steps have been taken, and more will be taken in the future. However, these centers and systems are not adequately recognized, even within the sponsoring organization, as a valuable resource. Management does not know about them. Staff does not use them. Utilization and awareness across organizational lines are even more scarce. How can the data system reach its users?

Problem Importance		Amenity to Resolution		Relevance of Your Experience	
Vital to National Systems	☑	Easily Solvable	☐	Highly Relevant	☑
Secondary Importance	☐	Difficult	☑	Relevant	☐
Not Important	☐	Impossible	☐	Not Related	☐

YOUR RECOMMENDATIONS:

Two kinds of action are necessary. First, COSATI should focus top-level R&D management attention on the resources which are available. Across-agency communication should be officially encouraged. Second, scientific and professional societies should set up, at national and regional meetings, display areas where selected information and data centers can offer their services, plus a guide-book on the existence of other centers.

Science Communication
Washington, D. C. 200 07
COSATI Data Systems Study
Final Report - F44620-67-C-0022

30 April 1968

SOURCES OF QUOTATIONS INCLUDED IN THE FOREWORD

Sources of Quotations Included in the Foreward

1. McLuhan, Marshall, "Technology: Its Influence on the Character of World Trade and Investment," in Proceedings of a Symposium: Technology and World Trade, National Bureau of Standards, U.S. Department of Commerce, November 16-17, 1966, p. 11.

2. Swanson, Rowena, Information System Networks....Let's Profit From What We Know, Office of Aerospace Research, U.S. Air Force, Arlington, Virginia, June 1966, p. 1.

3. Shipman, Joseph, "The Mounting Crisis in Primary Literature," Engineering Societies and Their Literature Programs: Proceedings of a Critical Appraisal, Engineers Joint Council, New York, New York, January 17-18, 1967, p. 28.

4. Kochen, Manfred, The Growth of Knowledge: Readings on Organization and Retrieval of Information, John Wiley & Sons, Inc., New York, New York, 1967, p. xi.

5. North American Aviation, Inc., Final Report: DoD User-Needs Study Phase II: Flow of Scientific and Technical Information Within the Defense Industry, Anaheim, California, November 30, 1966, pp. 1-9.

6. TRW Space Technology Laboratories, Technical Data Management and Implementation Program - Final Report, Thompson Ramo Wooldridge, Inc., May 28, 1964, p. 2-2.

7. Committee on Government Operations, Summary of Activities Toward Interagency Coordination, Report to the United States Senate, 89th Congress, 1st Session, June 24, 1965, p. 16.

8. Cahn, Julius N., A System of Information Systems, a paper presented at Fourth Institute on Information Storage and Retrieval, School of Government and Public Administration, American University, Washington, D.C., February 12, 1962, p. 2.

9. The President's Science Advisory Committee, Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information (The Weinberg Report), The White House, Washington, D.C., January 10, 1963 p. 7.

Quotation Sources, Continued...

10. Campbell, Colonel Charles T., "The Army Concept and Organization for Technical Data Management," in Minutes of Eighth Annual Meeting of the Engineering Documentation Section, American Ordnance Association, Washington, D.C., April 1966, pp. D2-14.

11. Committee on Government Operations, Summary of Activities Toward Interagency Coordination, Report to the United States Senate, 89th Congress, 1st Session, June 24, 1965, p. 19.

12. Cohen, Dr. Karl, "Interaction of Industry with Information," in Proceedings of DoD/NSIA Technical Information Symposium for Management, Statler Hilton Hotel, Los Angeles, California, May 26-27, 1965, p. 32.

13. Reily, Philip K., "Technical Information Transfer and Economic Growth," Keynote Address presented at the Institute on the Transfer of Technical Information, The American University, October 23, 1967, p. 3.

14. Carlson, Walter M., "Management of Scientific Information," in Proceedings of DoD/NSIA Technical Information Symposium for Management, Statler Hilton Hotel, Los Angeles, California, May 26-27, 1965, p. 44.

15. Ignatius, Paul R., "Defense Data Management Program Objectives," in Proceedings of DoD/NSIA Technical Information Symposium for Management, Statler Hilton Hotel, Los Angeles, California, May 26-27, 1965, p. 57.

16. Morris, Thomas D., "Keynote Address," in Proceedings of DoD/NSIA Technical Information Symposium for Management, Statler Hilton Hotel, Los Angeles, California, May 26-27, 1965, p. 3.

17. Select Committee on Government Research, Documentation and Dissemination of Research and Development Results, Study Number IV, House of Representatives, Washington, D.C., November 20, 1964, p. 91.

Quotation Sources, Continued...

18. Adkinson, Burton W., as quoted in Interagency Coordination of Information, Hearings before the Subcommittee on Reorganization and International Organizations of the Committee on Government Operations, U.S. Senate, September 21, 1962, p. 42.

19. Johnson, President Lyndon Baines, Presidential Message upon signing of the State Technical Services Act, P.L. 89-182, September 14, 1965.

20. Select Committee on Government Research, National Goals and Policies: Study Number X, Report to the House of Representatives, December 29, 1964, p. 64.

21. Battelle Memorial Institute, Proceedings of 1st Ad Hoc Forum of Scientific and Technical Information Analysis Center Managers, Directors, and Professional Analysts, November 9-11, 1965, p. 4.

22. Fubini, Eugene G., "Scientific Information," in Proceedings of DoD/NSIA Technical Information Symposium for Management, Statler Hilton Hotel, Los Angeles, California, May 26-27, 1965, p. 29.

23. Hornig, Donald, A Bill to Provide a Standard Reference Data System, Hearings before the Subcommittee on Science, Research, and Development of the Committee on Science and Astronautics, U.S. House of Representatives, June 28-30, 1966, p. 10.

24. Committee on Government Operations, Summary of Activities Toward Interagency Coordination, Report to the U.S. Senate, 89th Congress, 1st Session, June 24, 1965, p. 19.

25. Select Committee on Government Research, National Goals and Policies: Study Number X, Report to the House of Representatives, December 29, 1964, p. 51.

26. Hoshovsky, A.G., and Album, H.H., "Toward a National Technical Information System," in American Documentation, Volume 16, No. 4, October 1965.

27. Committee on Government Operations, Summary of Activities Toward Interagency Coordination, Report to the U.S. Senate, 89th Congress, 1st Session, June 24, 1965, p. 16.

Quotation Sources, Continued...

28. Hoshovsky, A.G., and Album, H.H., "Toward a National Technical Information System," in American Documentation, Volume 16, No. 4, October 1965, p. 313.

29. Klein, Stanley, "A Summary and a Perspective of This Conference," in Engineering Societies and Their Literature Programs: Proceedings of a Critical Appraisal, Engineers Joint Council, New York, New York, January 17-18, 1967, p. 7.

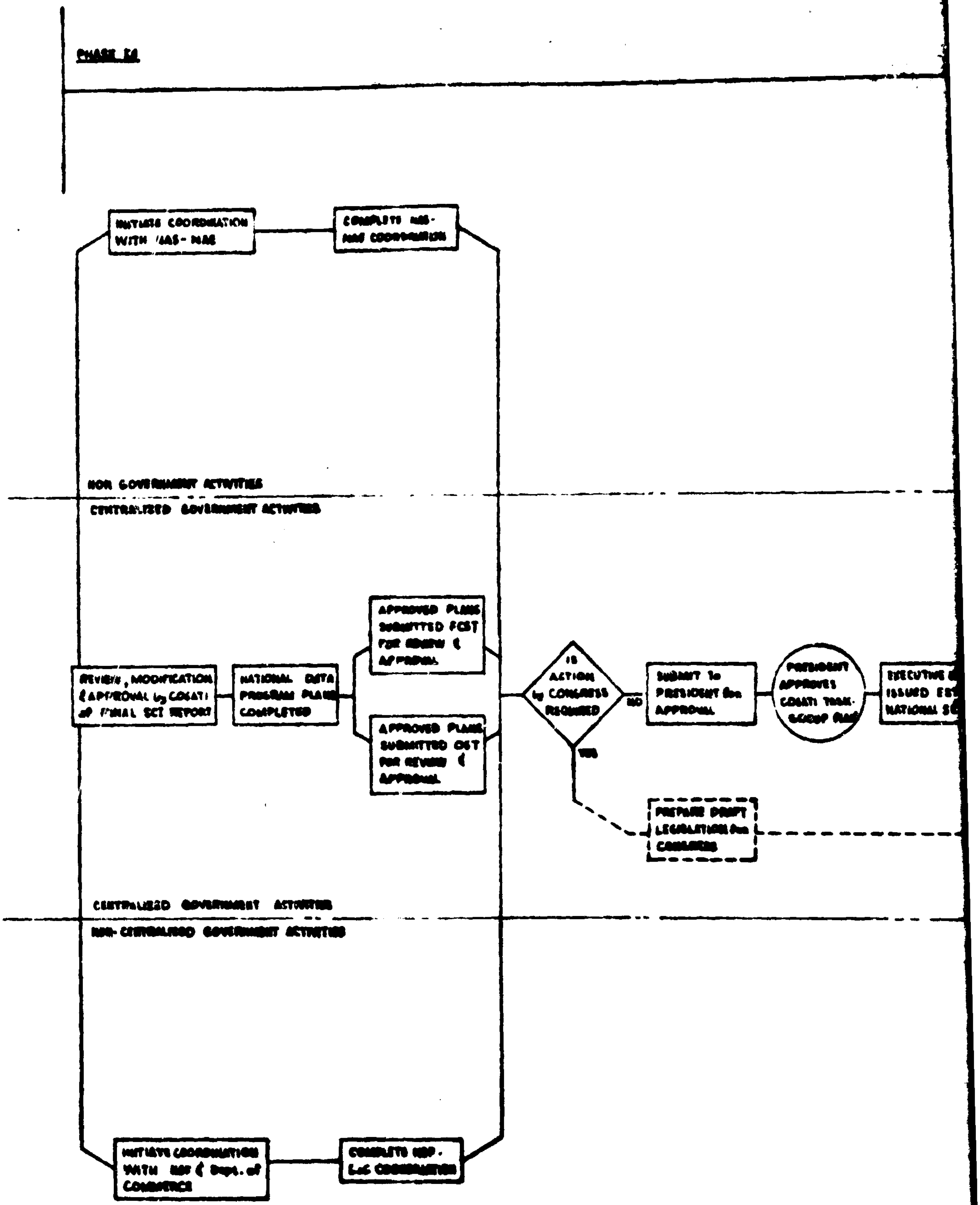
30. Fubini, Eugene G., "Scientific Information," in Proceedings of DoD/NSIA Technical Information Symposium for Management, Statler Hilton Hotel, Los Angeles, California, May 26-27, 1965, p. 31.

31. President's Science Advisory Committee, Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information (The Weinberg Report), The White House, Washington, D.C., January, 1963, p. 21.

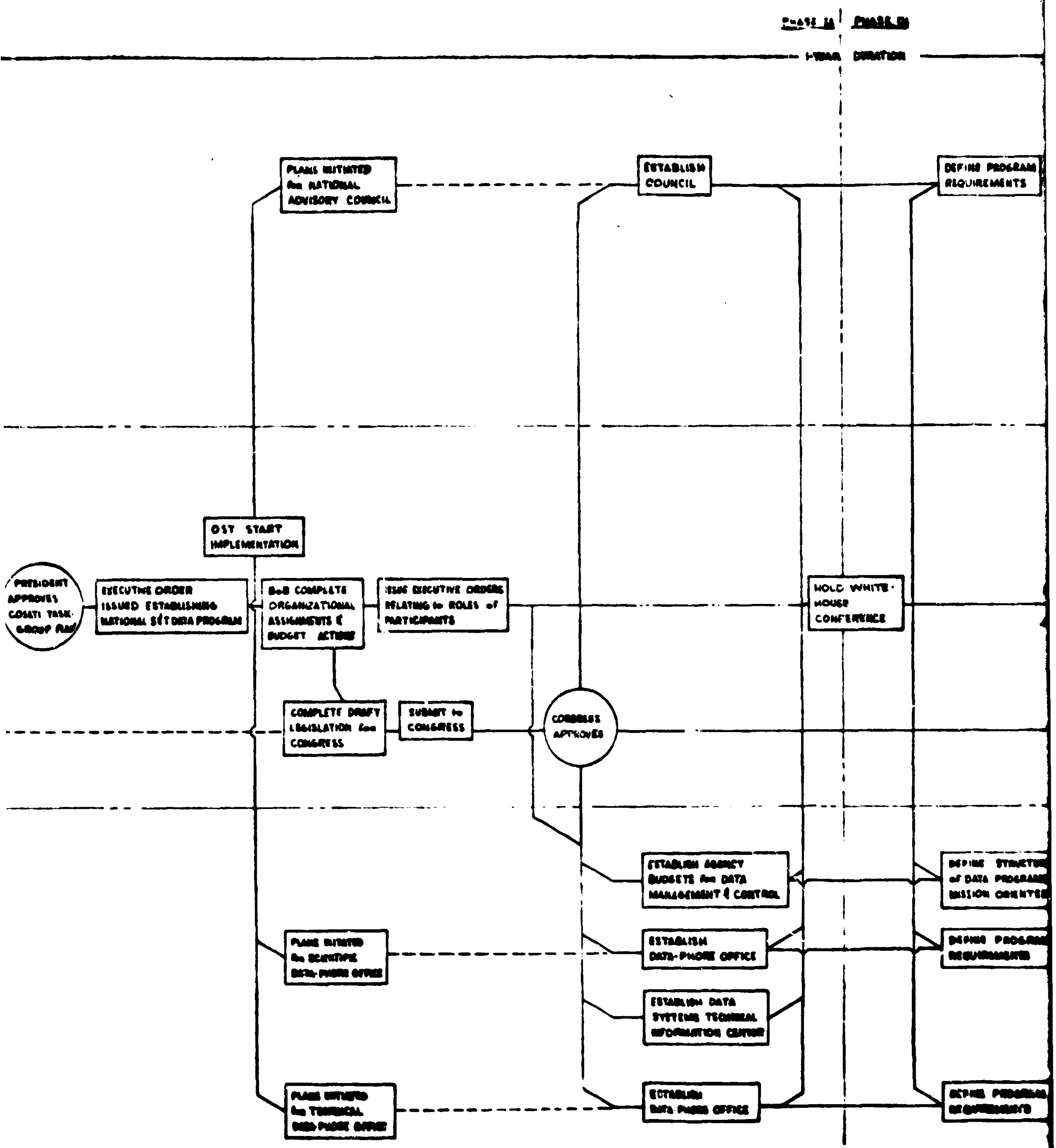
32. Wiesner, Dr. Jerome, Interagency Coordination of Information, Hearings of the Committee on Government Operations, U.S. Senate, September 21, 1962, p. 31.

33. Terry, Luther L., Surgeon General, U.S. Public Health Service, as quoted in Interagency Coordination of Information, Hearings before the Subcommittee on Reorganization and International Organizations of the Committee on Government Operations, U.S. Senate, September 21, 1962, p. 104.

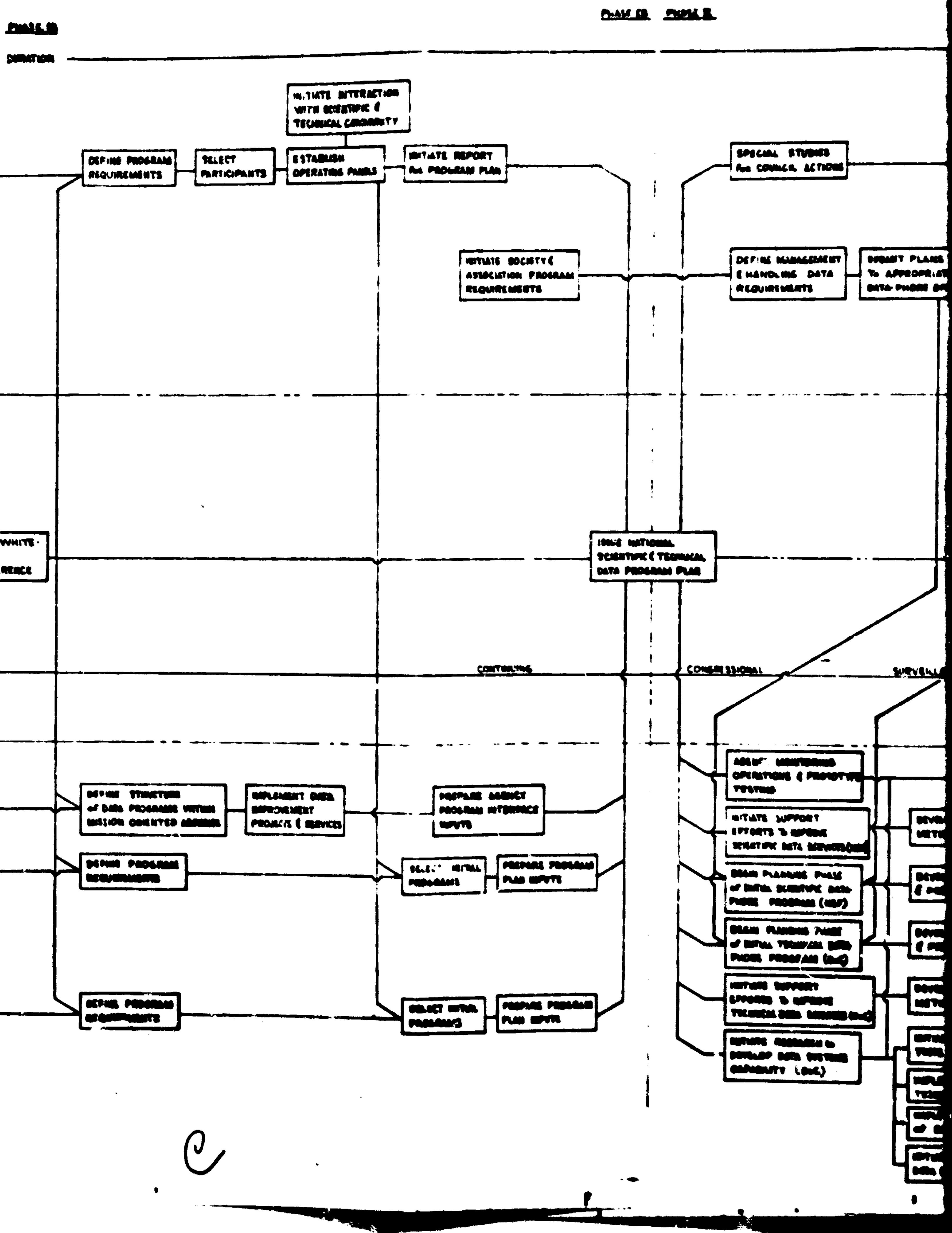
34. de Solla Price, Derek J., "Communication in Science: The Ends--Philosophy and Forecast," reprinted from Ciba Foundation Symposium on Communication in Science: Documentation and Automation, Anthony de Reuck and Julie Knight, Editors; J. & A. Churchill, Ltd., Publishers; London, 1967, p. 209.

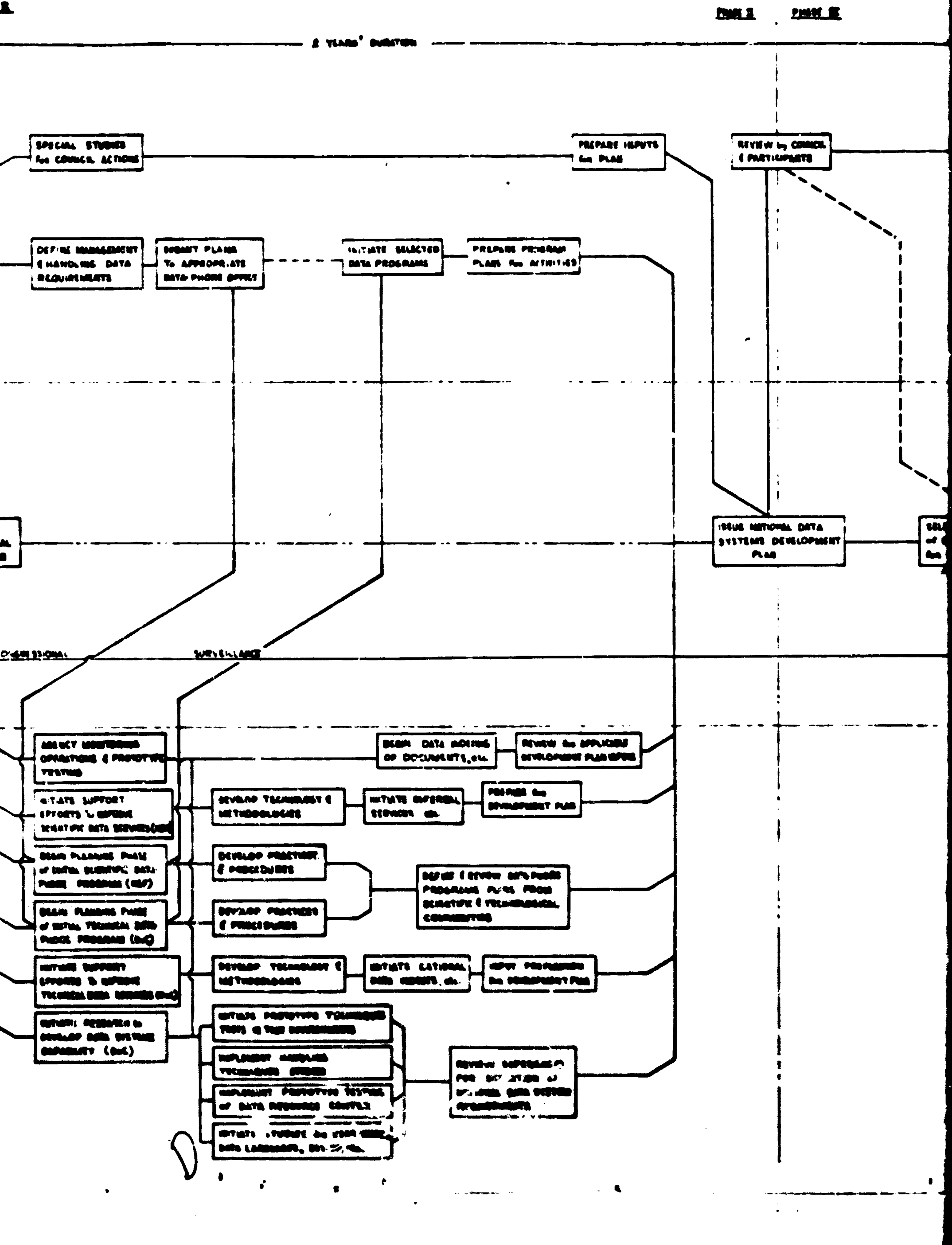


INITIATE COORDINATION WITH NAS-NAE
NON GOVERNMENT ACTIVITIES
CENTRALIZED GOVERNMENT ACTIVITIES
REVIEW, MODIFICATION & APPROVAL by COSATI of FINAL SCI REPORT
NATIONAL DATA PROGRAM PLANS COMPLETED
APPROVED PLANS SUBMITTED FCST FOR REVIEW & APPROVAL
APPROVED PLANS SUBMITTED OST FOR REVIEW & APPROVAL
IS ACTION by CONGRESS REQUIRED
NO
YES
SUBMIT to PRESIDENT for APPROVAL
PRESIDENT APPROVES COSATI TASK GROUP PLAN
EXECUTIVE
ISSUED EST
NATIONAL S
PREPARE DRAFT LEGISLATION for CONGRESS
CENTRALIZED GOVERNMENT ACTIVITIES
NON-CENTRALIZED GOVERNMENT ACTIVITIES

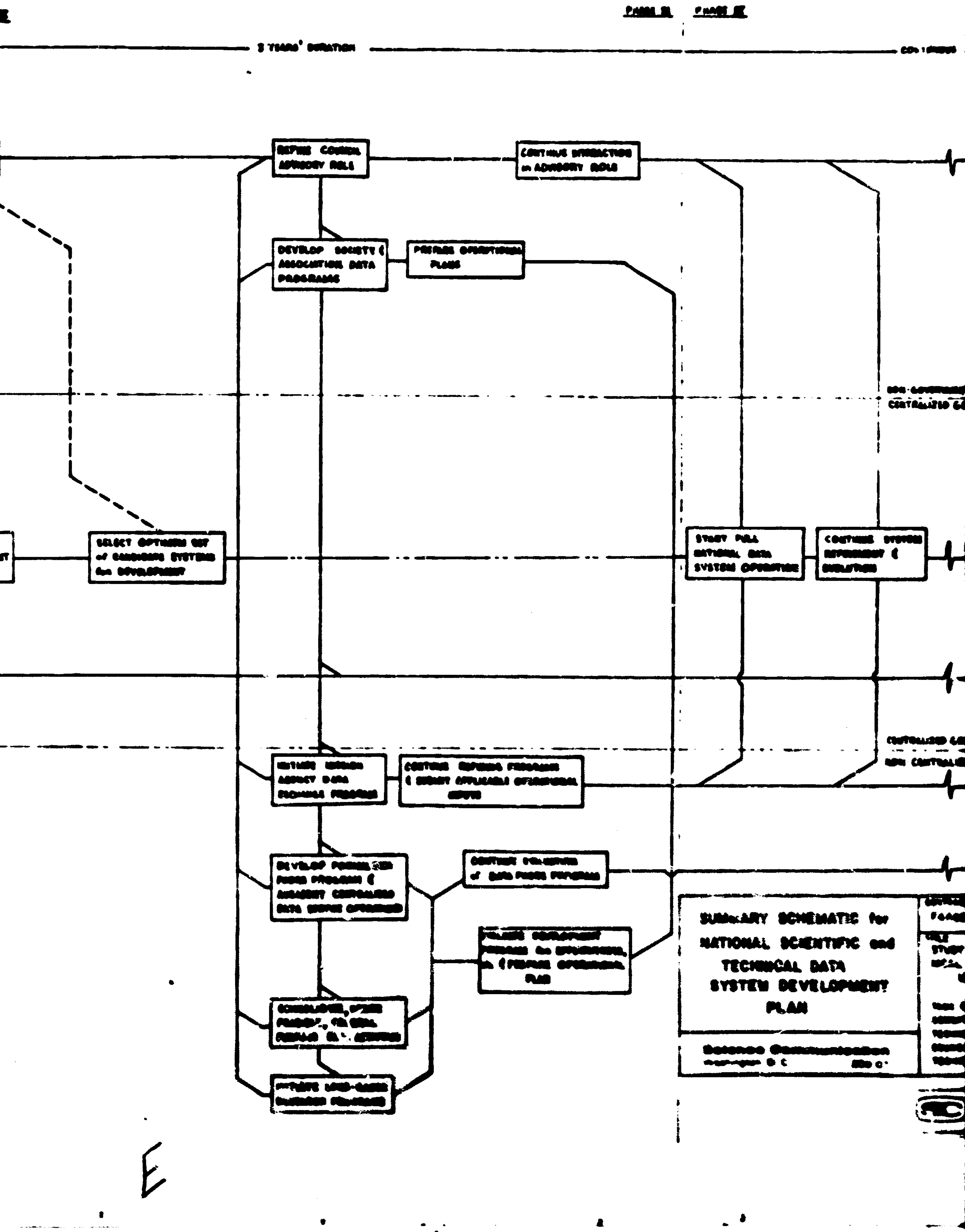

1-YEAR DURATION
PLANS INITIATED for NATIONAL ADVISORY COUNCIL
ESTABLISH COUNCIL
DEFINE PROGRAM REQUIREMENTS
OST START IMPLEMENTATION
PRESIDENT APPROVES COSATI TASK GROUP PLAN
EXECUTIVE ORDER ISSUED ESTABLISHING NATIONAL S&T DATA PROGRAM
BoB COMPLETE ORGANIZATIONAL ASSIGNMENTS & BUDGET ACTIONS
ISSUE EXECUTIVE ORDERS RELATING to ROLES of PARTICIPANTS
HOLD WHITE-HOUSE CONFERENCE
COMPLETE DRAFT LEGISLATION for CONGRESS
SUBMIT to CONGRESS
CONGRESS APPROVES
ESTABLISH AGENCY BUDGETS for DATA MANAGEMENT & CONTROL
DEFINE STRUCTURE of DATA PROGRAM
ESTABLISH DATA SYSTEMS TECHNICAL INFORMATION CENTER
DEFINE PROGRAM REQUIREMENTS
DEFINE PROGRAM REQUIREMENTS

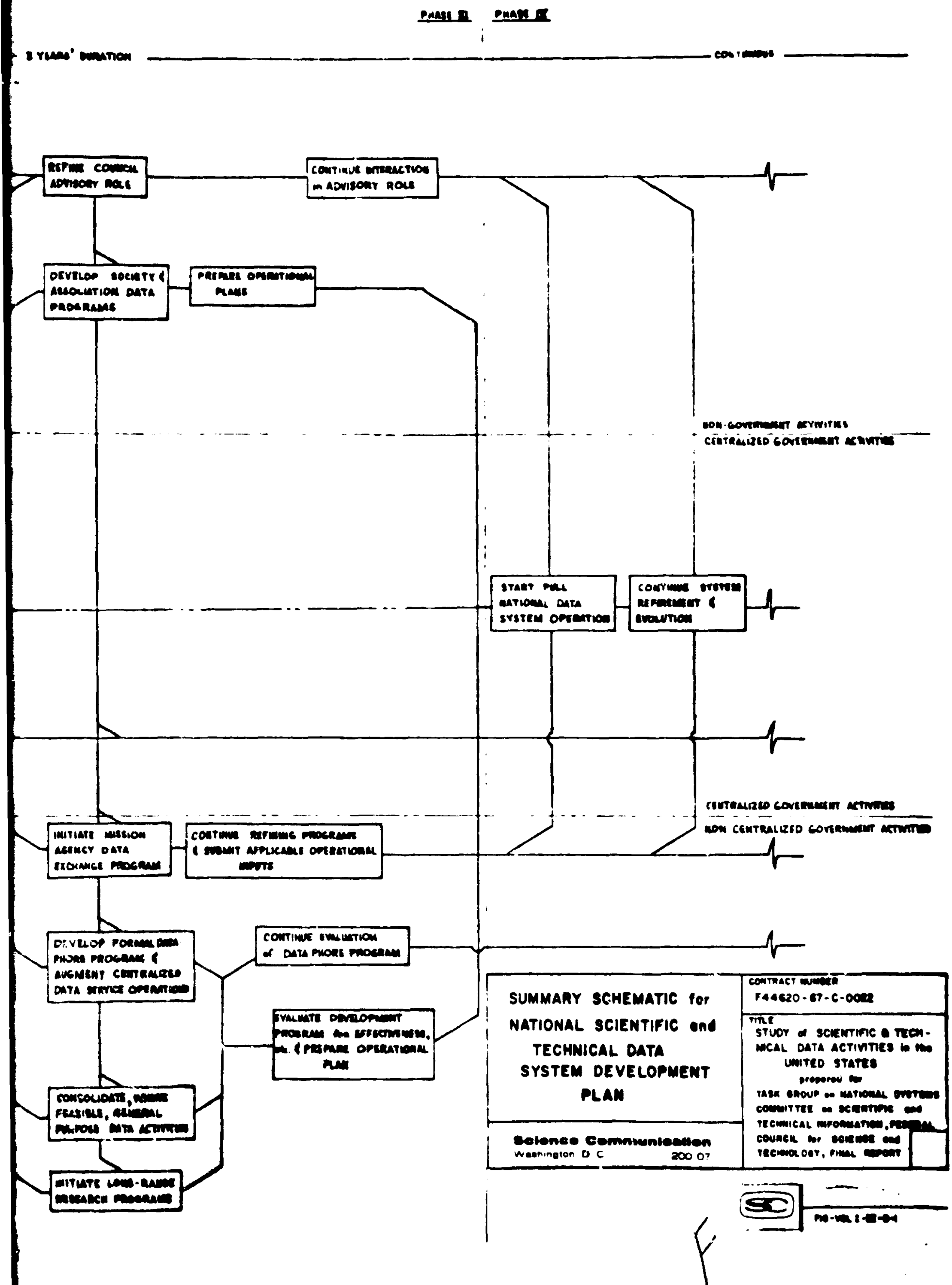
B


PHASE III
PHASE III PHASE IV
DURATION
INITIATE INTERACTION WITH SCIENTIFIC & TECHNICAL COMMUNITY
DEFINE PROGRAM REQUIREMENTS
SELECT PARTICIPANTS
ESTABLISH OPERATING PANELS
INITIATE REPORT FOR PROGRAM PLAN
SPECIAL STUDIES FOR COUNCIL ACTIONS
INITIATE SOCIETY & ASSOCIATION PROGRAM REQUIREMENTS
DEFINE MANAGEMENT & HANDLING DATA REQUIREMENTS
SUBMIT PLANS TO APPROPRIATE DATA PHASE [illegible]
WHITE-
RENCE
ISSUE NATIONAL SCIENTIFIC & TECHNICAL DATA PROGRAM PLAN
CONTINUING
CONGRESSIONAL
SURVEILLA
DEFINE STRUCTURE of DATA PROGRAMS WITHIN MISSION ORIENTED AGENCIES
IMPLEMENT DATA IMPROVEMENT PROJECTS & SERVICES
PREPARE AGENCY PROGRAM INTERFACE INPUTS
DEFINE PROGRAM REQUIREMENTS
SELECT INITIAL PROGRAMS
PREPARE PROGRAM PLAN INPUTS
DEFINE PROGRAM REQUIREMENTS
SELECT INITIAL PROGRAMS
PREPARE PROGRAM PLAN INPUTS
AGENCY MONITORING OPERATIONS & PROTOTYPE TESTING
INITIATE SUPPORT EFFORTS TO IMPROVE SCIENTIFIC DATA SERVICES (NSF)
BEGIN PLANNING PHASE of INITIAL SCIENTIFIC DATA PHASE PROGRAM (NSF)
BEGIN PLANNING PHASE of INITIAL TECHNICAL DATA PHASE PROGRAM (DoC)
INITIATE SUPPORT EFFORTS TO IMPROVE TECHNICAL DATA SERVICES (DoC)
INITIATE RESEARCH to DEVELOP DATA SYSTEMS CAPABILITY (DoC)

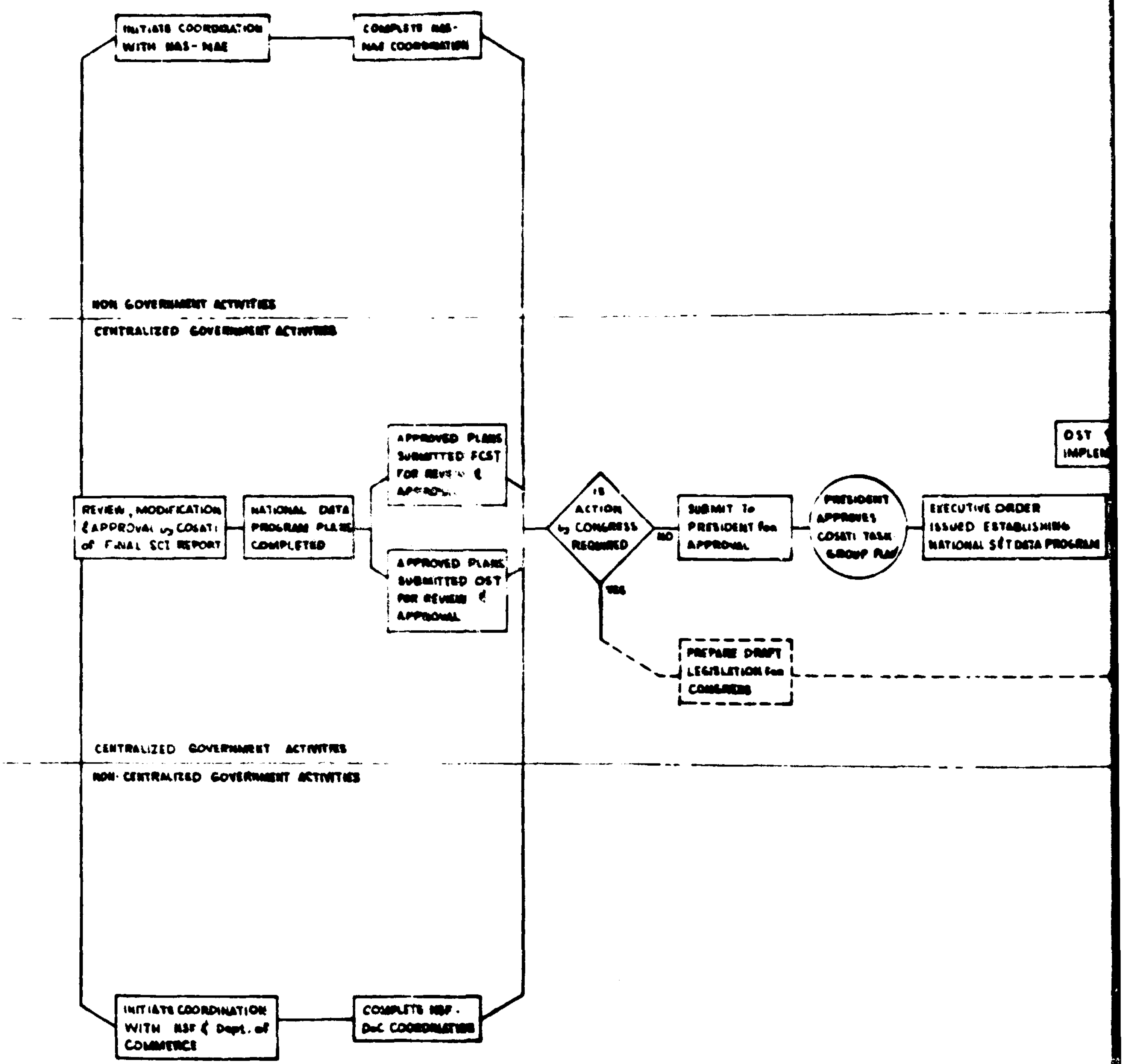


2 YEARS' DURATION
SPECIAL STUDIES for COUNCIL ACTIONS
PREPARE INPUTS for PLAN
REVIEW by COUNCIL & PARTICIPANTS
DEFINE MANAGEMENT & HANDLING DATA REQUIREMENTS
INITIATE SELECTED DATA PROGRAMS
PREPARE PROGRAM PLANS for ACTIVITIES
ISSUE NATIONAL DATA SYSTEMS DEVELOPMENT PLAN
SURVEILLANCE
BEGIN DATA INDEXING OF DOCUMENTS, etc.
DEVELOP TECHNOLOGY & METHODOLOGIES
DEVELOP PRACTICES & PROCEDURES

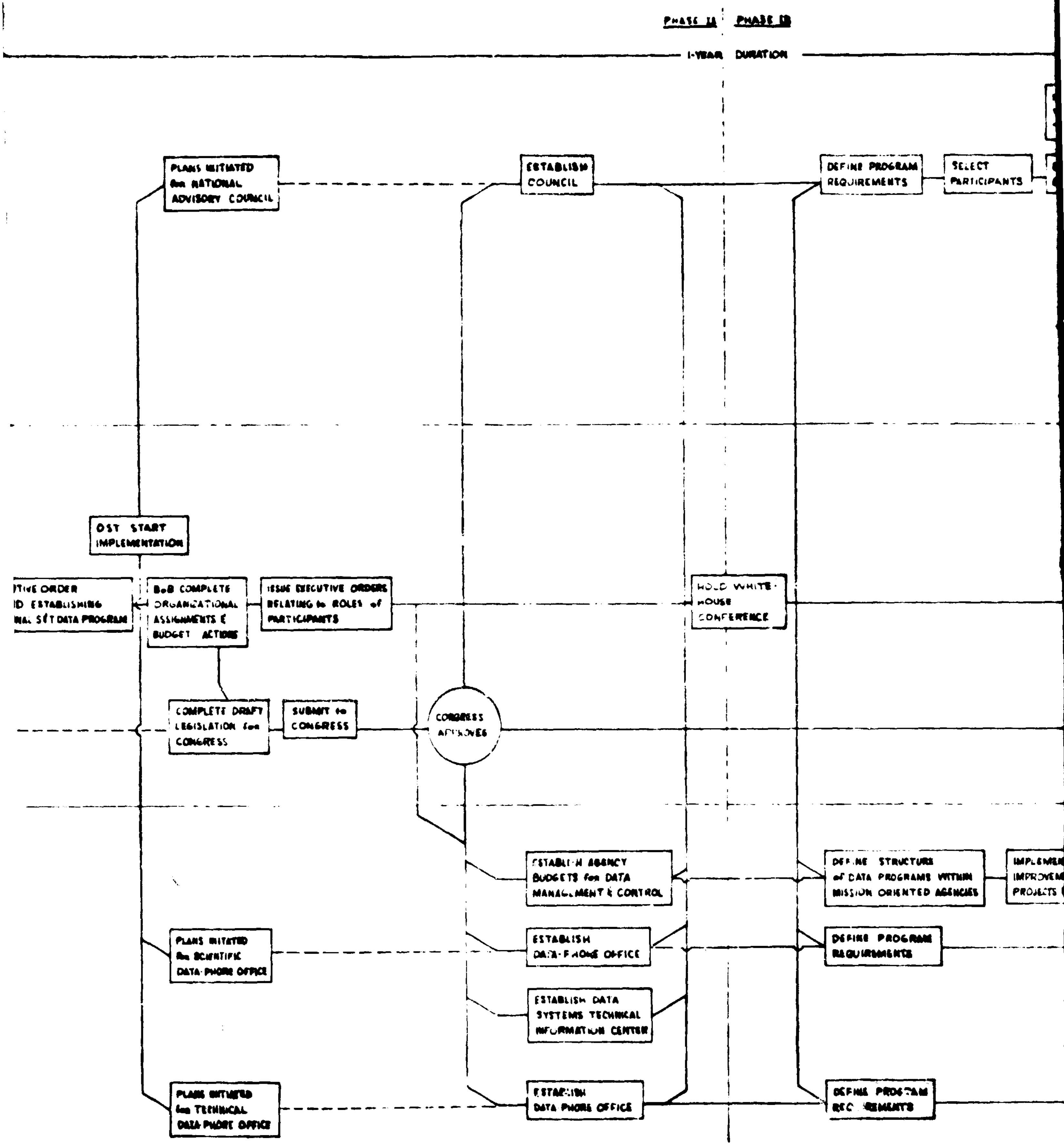

3 YEARS' DURATION
REFINE COUNCIL ADVISORY ROLE
CONTINUE INTERACTION in ADVISORY ROLE
DEVELOP SOCIETY & ASSOCIATION DATA PROGRAMS
PREPARE OPERATIONAL PLANS
SELECT OPTIMUM SET of CANDIDATE SYSTEMS for DEVELOPMENT
START FULL NATIONAL DATA SYSTEM OPERATION
CONTINUE SYSTEM REFINEMENT & EVOLUTION
SUMMARY SCHEMATIC for NATIONAL SCIENTIFIC and TECHNICAL DATA SYSTEM DEVELOPMENT PLAN
Science Communication
Washington D C


PHASE III
PHASE IV
3 YEARS' DURATION
CONTINUOUS
REFINE COUNCIL ADVISORY ROLE
CONTINUE INTERACTION in ADVISORY ROLE
DEVELOP SOCIETY & ASSOCIATION DATA PROGRAMS
PREPARE OPERATIONAL PLANS
NON-GOVERNMENT ACTIVITIES
CENTRALIZED GOVERNMENT ACTIVITIES
START FULL NATIONAL DATA SYSTEM OPERATION
CONTINUE SYSTEM REFINEMENT & EVOLUTION
CENTRALIZED GOVERNMENT ACTIVITIES
NON CENTRALIZED GOVERNMENT ACTIVITIES
INITIATE MISSION AGENCY DATA EXCHANGE PROGRAM
CONTINUE REFINING PROGRAMS & SUBMIT APPLICABLE OPERATIONAL INPUTS
DEVELOP FORMAL DATA PHORE PROGRAM & AUGMENT CENTRALIZED DATA SERVICE OPERATIONS
CONTINUE EVALUATION of DATA PHORE PROGRAM
EVALUATE DEVELOPMENT PROGRAM for EFFECTIVENESS, etc. & PREPARE OPERATIONAL PLAN
CONSOLIDATE, WHERE FEASIBLE, GENERAL PURPOSE DATA ACTIVITIES
INITIATE LONG-RANGE RESEARCH PROGRAMS
SUMMARY SCHEMATIC for NATIONAL SCIENTIFIC and TECHNICAL DATA SYSTEM DEVELOPMENT PLAN
Science Communication
Washington D C 200 07
CONTRACT NUMBER
F44620-67-C-0022
TITLE
STUDY of SCIENTIFIC & TECHNICAL DATA ACTIVITIES in the UNITED STATES
prepared for
TASK GROUP on NATIONAL SYSTEMS COMMITTEE on SCIENTIFIC and TECHNICAL INFORMATION, FEDERAL COUNCIL for SCIENCE and TECHNOLOGY, FINAL REPORT
SC

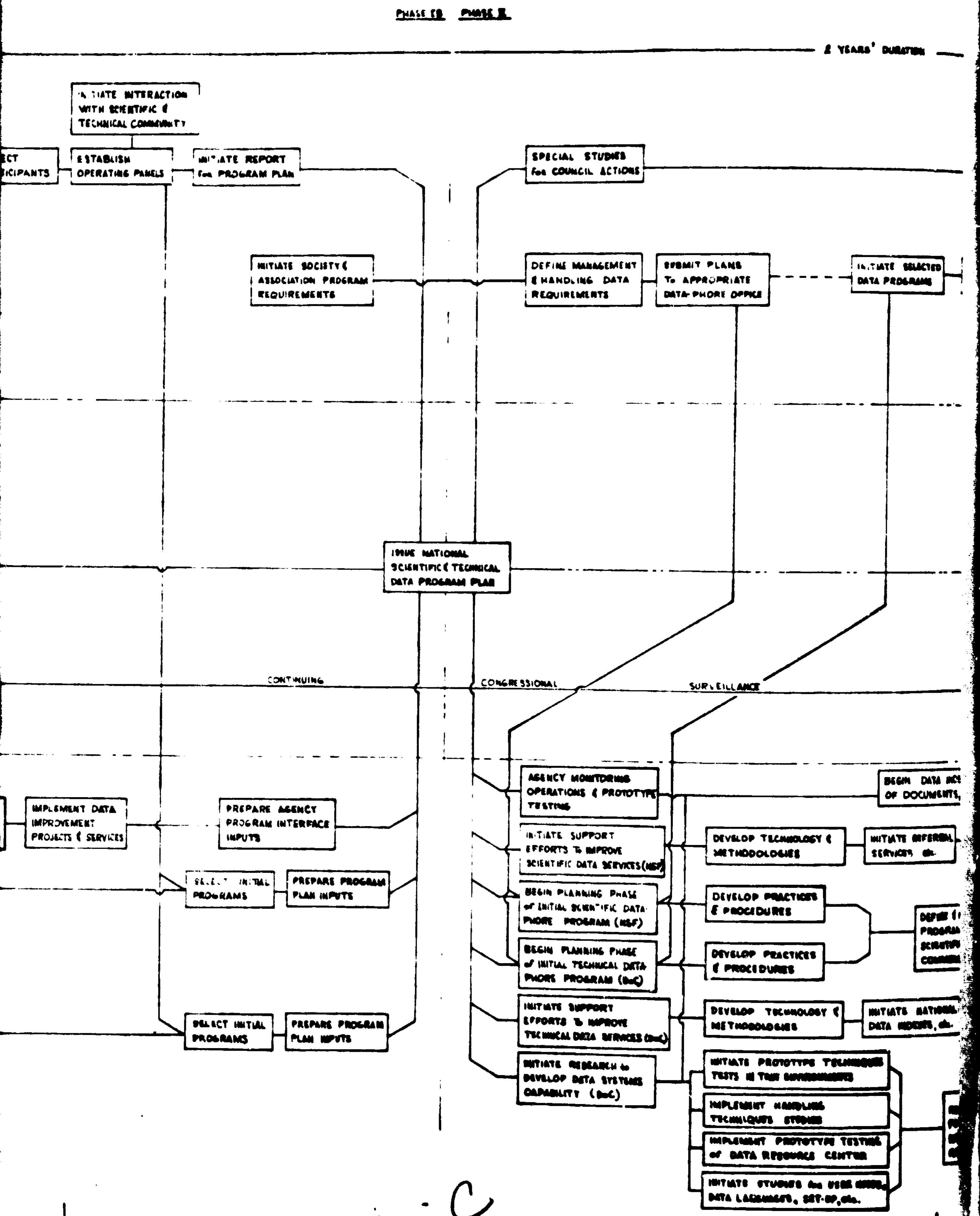
PHASE IA



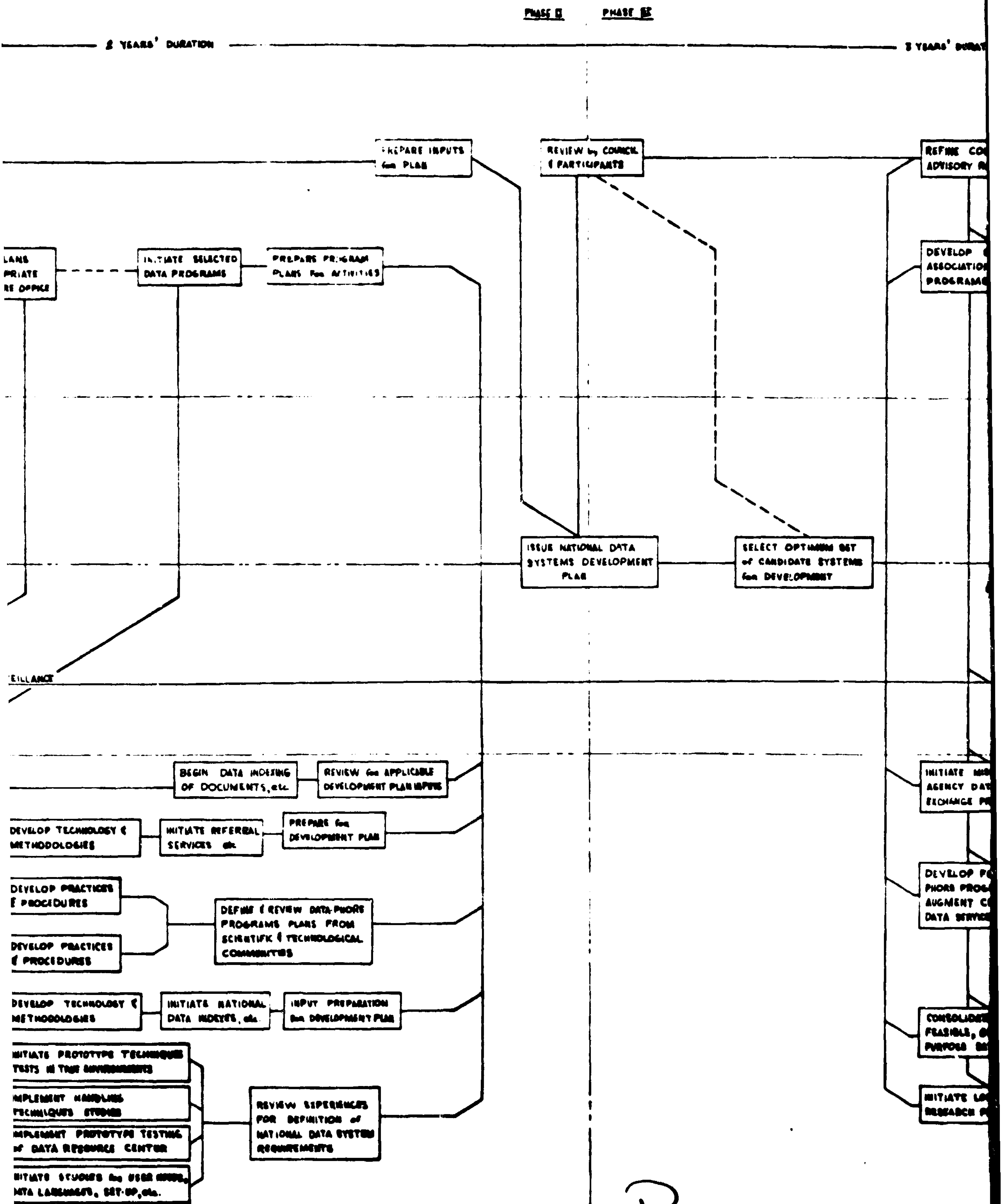
INITIATE COORDINATION WITH NAS-NAE
COMPLETE NAS-NAE COORDINATION
NON GOVERNMENT ACTIVITIES
CENTRALIZED GOVERNMENT ACTIVITIES
REVIEW, MODIFICATION & APPROVAL by COSATI of FINAL SCI REPORT
NATIONAL DATA PROGRAM PLANS COMPLETED
APPROVED PLANS SUBMITTED FCST FOR REVIEW & APPROVAL
APPROVED PLANS SUBMITTED OST FOR REVIEW & APPROVAL
IS ACTION by CONGRESS REQUIRED
NO
YES
SUBMIT To PRESIDENT for APPROVAL
PRESIDENT APPROVES COSATI TASK GROUP PLAN
EXECUTIVE ORDER ISSUED ESTABLISHING NATIONAL S&T DATA PROGRAM
OST
IMPLEM
PREPARE DRAFT LEGISLATION for CONGRESS
CENTRALIZED GOVERNMENT ACTIVITIES
NON-CENTRALIZED GOVERNMENT ACTIVITIES
INITIATE COORDINATION WITH NSF & Dept. of COMMERCE
COMPLETE NSF-DoC COORDINATION


1-YEAR DURATION
PLANS INITIATED for NATIONAL ADVISORY COUNCIL
ESTABLISH COUNCIL
DEFINE PROGRAM REQUIREMENTS
SELECT PARTICIPANTS
OST START IMPLEMENTATION
BoB COMPLETE ORGANIZATIONAL ASSIGNMENTS & BUDGET ACTIONS
ISSUE EXECUTIVE ORDERS RELATING to ROLES of PARTICIPANTS
HOLD WHITE-HOUSE CONFERENCE
COMPLETE DRAFT LEGISLATION for CONGRESS
SUBMIT to CONGRESS
CONGRESS APPROVES
ESTABLISH AGENCY BUDGETS for DATA MANAGEMENT & CONTROL
DEFINE STRUCTURE of DATA PROGRAMS WITHIN MISSION ORIENTED AGENCIES
PLANS INITIATED for SCIENTIFIC DATA-PHONE OFFICE
ESTABLISH DATA-PHONE OFFICE
DEFINE PROGRAM REQUIREMENTS
ESTABLISH DATA SYSTEMS TECHNICAL INFORMATION CENTER
PLANS INITIATED for TECHNICAL DATA-PHONE OFFICE
ESTABLISH DATA PHONE OFFICE
DEFINE PROGRAM

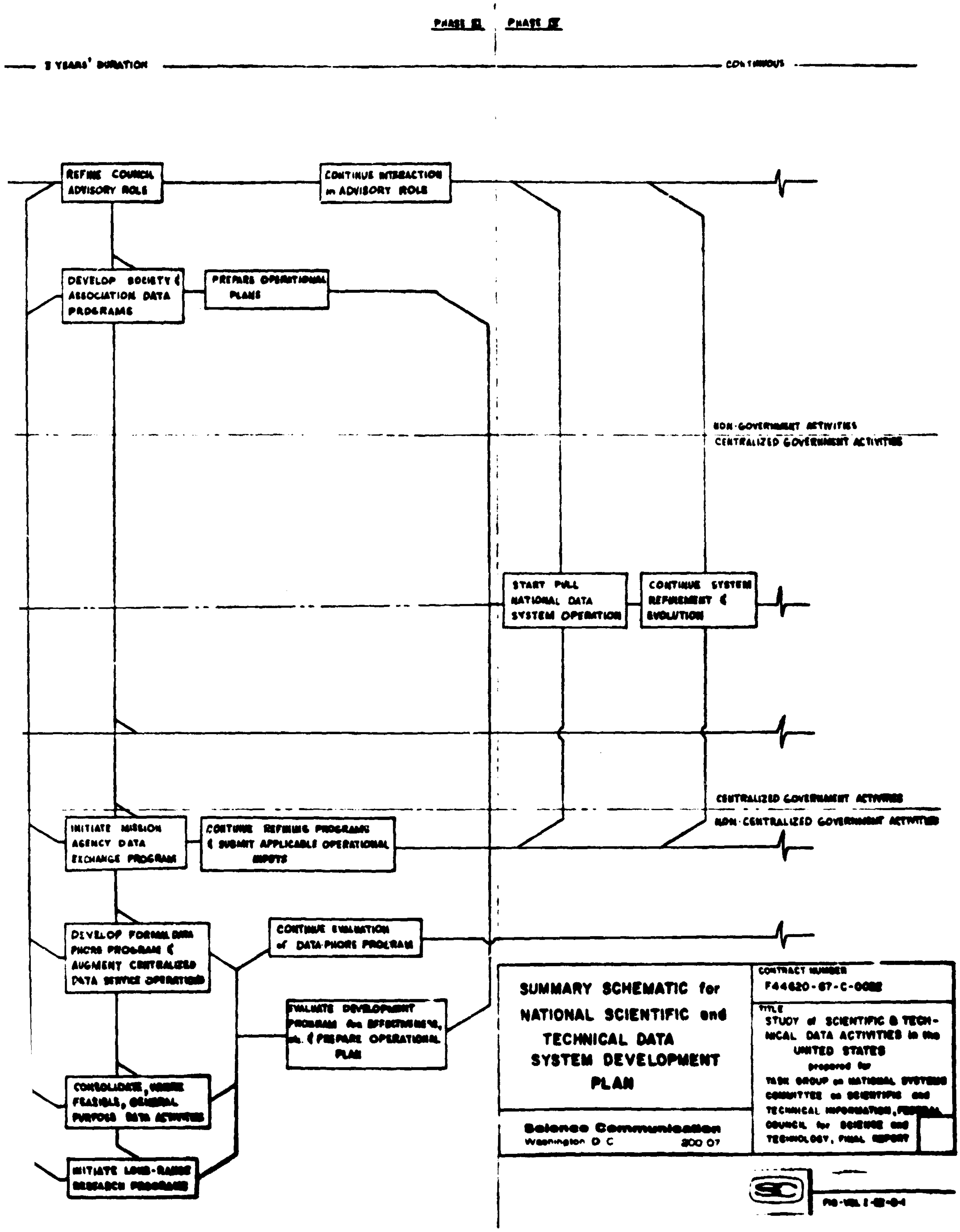
B


PHASE IB PHASE II
2 YEARS' DURATION
INITIATE INTERACTION WITH SCIENTIFIC & TECHNICAL COMMUNITY
ESTABLISH OPERATING PANELS
INITIATE REPORT FOR PROGRAM PLAN
SPECIAL STUDIES FOR COUNCIL ACTIONS
INITIATE SOCIETY & ASSOCIATION PROGRAM REQUIREMENTS
DEFINE MANAGEMENT & HANDLING DATA REQUIREMENTS
SUBMIT PLANS TO APPROPRIATE DATA-PHORE OFFICE
INITIATE SELECTED DATA PROGRAMS
ISSUE NATIONAL SCIENTIFIC & TECHNICAL DATA PROGRAM PLAN
CONTINUING
CONGRESSIONAL
SURVEILLANCE
AGENCY MONITORING OPERATIONS & PROTOTYPE TESTING
IMPLEMENT DATA IMPROVEMENT PROJECTS & SERVICES
PREPARE AGENCY PROGRAM INTERFACE INPUTS
INITIATE SUPPORT EFFORTS TO IMPROVE SCIENTIFIC DATA SERVICES (NSF)
DEVELOP TECHNOLOGY & METHODOLOGIES
SELECT INITIAL PROGRAMS
PREPARE PROGRAM PLAN INPUTS
BEGIN PLANNING PHASE of INITIAL SCIENTIFIC DATA-PHORE PROGRAM (NSF)
DEVELOP PRACTICES & PROCEDURES
BEGIN PLANNING PHASE of INITIAL TECHNICAL DATA-PHORE PROGRAM (DoC)
DEVELOP PRACTICES & PROCEDURES
INITIATE SUPPORT EFFORTS TO IMPROVE TECHNICAL DATA SERVICES (DoC)
DEVELOP TECHNOLOGY & METHODOLOGIES
SELECT INITIAL PROGRAMS
PREPARE PROGRAM PLAN INPUTS
INITIATE RESEARCH to DEVELOP DATA SYSTEMS CAPABILITY (DoC)
INITIATE PROTOTYPE TECHNIQUES TESTS IN TRUE ENVIRONMENTS
IMPLEMENT HANDLING TECHNIQUES STUDIES
IMPLEMENT PROTOTYPE TESTING of DATA RESOURCE CENTER
INITIATE STUDIES for USER NEEDS, DATA LANGUAGES, SET-UP, etc.

- C



PHASE II
PHASE III
2 YEARS' DURATION
3 YEARS' DURAT
PREPARE INPUTS for PLAN
REVIEW by COUNCIL & PARTICIPANTS
REFINE CO
ADVISORY R
INITIATE SELECTED DATA PROGRAMS
PREPARE PROGRAM PLANS for ACTIVITIES
DEVELOP
ASSOCIATION
PROGRAMS
ISSUE NATIONAL DATA SYSTEMS DEVELOPMENT PLAN
SELECT OPTIMUM SET of CANDIDATE SYSTEMS for DEVELOPMENT
BEGIN DATA INDEXING OF DOCUMENTS, etc.
REVIEW for APPLICABLE DEVELOPMENT PLAN INPUTS
INITIATE
AGENCY DAT
EXCHANGE P
DEVELOP TECHNOLOGY & METHODOLOGIES
INITIATE REFERRAL SERVICES etc.
PREPARE for DEVELOPMENT PLAN
DEVELOP PRACTICES & PROCEDURES
DEFINE & REVIEW DATA-PHORE PROGRAMS PLANS FROM SCIENTIFIC & TECHNOLOGICAL COMMUNITIES
DEVELOP PRACTICES & PROCEDURES
DEVELOP F
AUGMENT C
DATA SERVICE
DEVELOP TECHNOLOGY & METHODOLOGIES
INITIATE NATIONAL DATA INDEXES, etc.
INPUT PREPARATION for DEVELOPMENT PLAN
CONSOLIDATE
FEASIBLE,
INITIATE PROTOTYPE TECHNIQUES TESTS IN TRUE ENVIRONMENTS
IMPLEMENT HANDLING TECHNIQUES STUDIES
IMPLEMENT PROTOTYPE TESTING of DATA RESOURCE CENTER
INITIATE STUDIES for USER NEEDS, DATA LANGUAGES, SET-UP, etc.
REVIEW EXPERIENCES FOR DEFINITION of NATIONAL DATA SYSTEM REQUIREMENTS
INITIATE LO
RESEARCH
D


PHASE III
PHASE IV
3 YEARS' DURATION
CONTINUOUS
REFINE COUNCIL ADVISORY ROLE
CONTINUE INTERACTION in ADVISORY ROLE
DEVELOP SOCIETY & ASSOCIATION DATA PROGRAMS
PREPARE OPERATIONAL PLANS
NON-GOVERNMENT ACTIVITIES
CENTRALIZED GOVERNMENT ACTIVITIES
START FULL NATIONAL DATA SYSTEM OPERATION
CONTINUE SYSTEM REFINEMENT & EVOLUTION
CENTRALIZED GOVERNMENT ACTIVITIES
NON-CENTRALIZED GOVERNMENT ACTIVITIES
INITIATE MISSION AGENCY DATA EXCHANGE PROGRAM
CONTINUE REFINING PROGRAMS & SUBMIT APPLICABLE OPERATIONAL INPUTS
DEVELOP FORMAL DATA PHORE PROGRAM & AUGMENT CENTRALIZED DATA SERVICE OPERATIONS
CONTINUE EVALUATION of DATA-PHORE PROGRAM
EVALUATE DEVELOPMENT PROGRAM for EFFECTIVENESS, etc. & PREPARE OPERATIONAL PLAN
CONSOLIDATE, WHERE FEASIBLE, GENERAL PURPOSE DATA ACTIVITIES
INITIATE LONG-RANGE RESEARCH PROGRAMS
SUMMARY SCHEMATIC for NATIONAL SCIENTIFIC and TECHNICAL DATA SYSTEM DEVELOPMENT PLAN
Science Communication
Washington D C 200 07
CONTRACT NUMBER
F44620-67-C-00
TITLE
STUDY of SCIENTIFIC & TECHNICAL DATA ACTIVITIES in the UNITED STATES
prepared for
TASK GROUP on NATIONAL SYSTEMS
COMMITTEE on SCIENTIFIC and TECHNICAL INFORMATION, FEDERAL COUNCIL for SCIENCE and TECHNOLOGY, FINAL REPORT
SC
FIG-VOL I

E

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Science Communication, Inc. 1050 31st Street, N. W. Washington, D. C. 20007	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP NA

3. REPORT TITLE

Study of Scientific and Technical Data Activities in the United States
Volume I - Plan for Study and Implementation of National Data System Concepts

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report

5. AUTHOR(S) (First name, middle initial, last name)

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
30 April 1968	287	Numerous

8a. CONTRACT OR GRANT NO. Contract F44620-67-C-0022 b. PROJECT NO. ARPA Order 892 as amended c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY ~~Office of Aerospace Research~~ AFOSR ~~1400 Wilson Boulevard~~ Arlington, Virginia 22209

13. ABSTRACT

This volume presents a plan for study and implementation of national scientific and technical data system(s) concepts. The plan reported was developed as a part of a broader planning effort by the Task Group on National System(s) of the Committee on Scientific and Technical Information (COSATI). COSATI is a committee of the Federal Council for Science and Technology.

Major objectives of the plan are: (1) management of scientific and technical data resources in a manner optimal for maintenance of a strong science and technology, (2) improvement of existing data management programs and data handling services by better use of available technologies and methodologies, (3) development of the personnel, institutional, and methodological capabilities required to support future data-management and data-handling systems, and (4) identification of procedures and designation of responsibilities for actions to facilitate the development of new systems of data management and data handling.

The plan envisions the achievement of those objectives within a National Program for Scientific and Technical Data. Significant elements of the National Program include organization of a National Advisory Council for Scientific and Technical Data and establishment of two Program Offices - one for scientific data activities and one for technical data activities. (continued on back)

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Scientific data						
Technical data						
Data management						
National data systems						
Data handling systems						
Scientific communications						
Data storage and retrieval						
Data collection networks						
Data service centers						
Data libraries						
Data processing						
Computers						
Microfilming						

13. Abstract (continued)

The plan presented ~~in this volume is based in~~ part on an extensive survey-study of data activities as currently conducted in government, industry, and the professions. The results of this background study are reported in Volume II of this report.

UNCLASSIFIED
Security Classification