DOCUMENT RESUME

ED 162 904 SE 025 443

AUTHOR Wood, Barbara L.

TITLE Review of Scientific & Technical Numeric Data Base Activities.

INSTITUTION King Research, Inc., Rockville, Md.
SPONS AGENCY National Science Foundation, Washington, D.C. Div. of Science Information.
PUB DATE Mar 77
CONTRACT NSF-C-878
NOTE 176p.; Contains occasional marginal legibility
AVAILABLE FROM National Technical Information Service (NTIS), Operations Division, Springfield, Virginia 22161 ($8.00)

EDRS PRICE MF-$0.83 HC-$10.03 Plus Postage.
DESCRIPTORS Data; *Data Bases; Data Collection; *Information Sources; *Resources; Sciences; *Scientific Research; *Statistics; Tables (Data)

ABSTRACT

This examination of numeric data base activities grew out of a continuing effort to develop a system of statistical indicators of scientific and technical communication. A broad mix of directories, listing by subject field, reports, mail inquiries and telephone and face-to-face interviews was used to collect information. Scientific and technical numeric data bases are identified and described in the following fields: physical, environmental, life and social sciences. Principal federal agencies supporting these data base activities include the National Oceanographic and Atmospheric Administration, Environmental Protection Agency, National Cancer Institute, Energy Research and Development Administration, National Space Science Data Center, United States Geological Survey, National Standard Reference Data System, National Archives and Records Service, U.S. Naval Observatory, Federal Energy Administration, and activities associated with Information Analysis Centers and federal statistical programs. This report provides information individually about some numeric data bases and collectively about some groups of data bases. There is some information about non-numeric aspects of data bases which also contain numeric data. The information collected about numeric data bases differs markedly from data base to data base. For some data bases, numbers of items in the file can be quoted and perhaps even growth rates. For some data bases, information is available concerning dollar costs, number of users, or number of queries. (Author/BB)

PB-278 254

Review of Scientific and Technical Numeric Data Base Activities

[illegible], Inc, Rockville, Md

Prepared for

National Science Foundation, Washington, D C Div of Science Information

Mar 77

REVIEW OF SCIENTIFIC & TECHNICAL NUMERIC DATA BASE ACTIVITIES

for

National Science Foundation
Division of Science Information
Under Contract No. NSF-C878;
"Development of Statistical
Indicators of Scientific
and Technical Communication"

Prepared by

Barbara L. Wood

King Research, Inc.
6000 Executive Blvd.
Rockville, Md.

March 1977

BIBLIOGRAPHIC DATA SHEET	1. Report No.	2.	3. Recipient's Accession No. PB278254

4. Title and Subtitle
Review of Scientific & Technical Numeric Data Base Activities.

5. Report Date
March 1977

6.

7. Author(s)
Barbara L. Wood

8. Performing Organization Rept. No. 6

9. Performing Organization Name and Address
King Research, Inc.
6000 Executive Boulevard
Rockville, Maryland 20852

10. Project/Task/Work Unit No. 9232

11. Contract/Grant No.
NSF-C878

12. Sponsoring Organization Name and Address
National Science Foundation
Division of Science Information
1800 G Street, N. W.
Washington, D. C. 20550

13. Type of Report & Period Covered
Research Report

14.

15. Supplementary Notes

16. Abstracts

This report is an examination of numeric data base activities collected as part of a continuing effort to develop a system of statistical indicators of scientific and technical communication. Data were collected by means of a broad mix of directories, listings by subject field, reports, mail inquiries and telephone and face-to-face interviews was used to collect information. Numeric data bases reported on reside largely within the Federal government. Many agencies are unable to attach funding levels specifically to support of numeric data bases. However, the Federal government alone provides funds of more than one billion dollars per year to support the group of numeric data activities where expenditures or obligations could be identified.

Scientific and technical numeric data bases are identified and described in the following fields: Physical, environmental, life and social sciences. Principal Federal agencies supporting these data base activities include; the National Oceanographic and Atmospheric Administration, Environmental Protection Agency, National Cancer Institute, Energy Research and Development Administration, National Space Science Data Center, United States Geological Survey, National Standard Reference Data System, National Archives and Records Service, U.S. Naval Observatory, Federal Energy Administration, and activities associated with Information Analysis Centers and Federal statistical programs.

17. Key Words and Document Analysis. 17a. Descriptors

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement
Available through NTIS.

19. Security Class (This Report) UNCLASSIFIED

20. Security Class (This Page) UNCLASSIFIED

22. Price PC A08 / MF A01

FORM NTIS-35 (REV. 10-73) ENDORSED BY ANSI AND UNESCO. THIS FORM MAY BE REPRODUCED USCOMM-DC 8265-P74

PORTIONS OF THIS REPORT ARE NOT LEGIBLE.

HOWEVER, IT IS THE BEST REPRODUCTION

AVAILABLE FROM THE COPY SENT TO NTIS

This material is based upon research supported by the National Science Foundation under Contract No. NSF C-878.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

ABSTRACT

This examination of numeric data base activities grew out of a continuing effort to develop a system of statistical indicators of scientific and technical communication. The larger multi-year effort, supported by the National Science Foundation, primarily addressed narrative, ink-print communication media. An exploratory study of the magnitude and growth of machine-readable numeric data, as an STI resource, was conducted to augment information about the other media.

No comprehensive list of scientific and technical numeric data bases, nor of organizations producing or maintaining them, currently exists. A broad mix of directories, listings by subject field, reports, mail inquiries and telephone and face-to-face interviews was used to collect information. Numeric data bases reported on reside largely within the Federal government. Many agencies are unable to attach funding levels specifically to support of numeric data bases. However, the Federal government alone provides funds of more than one billion dollars per year to support the group of numeric data activities where expenditures or obligations could be identified.

Scientific and technical numeric data bases are identified and described in the following fields: Physical, environmental, life and social sciences. Principal Federal agencies supporting these data base activities include; the National Oceanographic and Atmospheric Administration, Environmental Protection Agency, National Cancer Institute, Energy Research and Development Administration, National Space Science Data Center, United States Geological Survey, National Standard Reference Data System, National Archives and Records Service, U. S. Naval Observatory, Federal Energy Administration, and activities associated with Information Analysis Centers and Federal statistical programs.

Although rapid growth in the size and number of numeric data bases was frequently mentioned in the literature and in interviews, quantification of data base size was rarely the case. In addition, a great deal of numeric data was identified which, although not always machine-readable, was published or otherwise available in such forms as maps, charts, papers, sheets or reels of film, photographic prints and microforms.

There is a growing awareness on the part of many Federal agencies and the Congress, of the need for improving identification of numeric data sources and enhancing their use.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF TABLES

LIST OF TABLES (Cont'd)

LIST OF TABLES (Cont'd)

SECTION 1

INTRODUCTION

In 1975, the Center for Quantitative Science* began a study to develop statistical indicators of communication in science and technology. The scope of the study includes a broad interpretation of 'science and technology' as well as of 'communication.' The indicators which have been developed are based almost exclusively on printed communications media (i.e., the technical literature as it is normally understood). Therefore, the measures of the printed works (numbers of publications, costs, etc.) cited in subsequent sections of this report are subsumed within the data presented in the other "Statistical Indicators" reports issued in 1976 and 1977 (51, 52, 53, 102).

A very limited investigation into numeric data, conducted during 1975, indicated that (ink print) published numeric data might constitute only a small portion of the numeric data available in a communicable form to scientists and engineers. Technological developments within the past 15 years or so have created new media which are particularly suited to the recording, storage, and analysis of numeric data. This is especially true for those types of data which are voluminous or which require extensive mathematical or statistical manipulation. During 1976 an expanded investigation was conducted with the purpose of identifying and quantifying measures of the extent and utilization of numeric data bases in the United States.

This effort concentrated largely on machine-readable files (and products or services associated with them) in order to complement the literature data. King Research, Inc., sought to be able to make some more definitive statements than the often heard, "They're big and getting bigger," based on statistics for number and size of numeric files, costs, and usage.

The study made it very apparent that it is a far from simple task even to identify all (nearly all, all significant) numeric data bases in science and

*Then affiliated with Market Facts, Inc., currently King Research, Inc.

technology. Excluded from primary consideration are all data bases which support purely management information systems. Thus, a file about on-going research projects which contains narrative descriptions of the projects plus funding levels would be considered a science and technology data base with respect to the project descrptions but would not be considered a science and technology numeric data base merely by virtue of the numeric funding values.

At another point along the boundary of what should or should not be included are heavily numeric data bases which are collected and exist largely outside the realm of science and technology, but which are heavily used for work within this domain. Thus, fish catch statistics or chemical production figures may be largely economic statistics, yet they may be crucial to certain ecologic research. Demographic statistics may be intended primarily for governmental purposes, such as election districting and siting of schools, but they also underpin much epidemiological research.

There does not exist any comprehensive compilation of numeric data bases. There have been, from time to time, directories prepared which identify and, to some extent, characterize numeric data bases and associated activities, such as the International Compendium of Numeric Data Projects (14) and Critical Data in Britain. There are various compilations of information resources for selected subject matter, e.g., A Directory of Information Resources in the United States. Physical Sciences, Engineering (80), or for types of activities, e.g., Directory of Federally Supported Information Analysis Centers (79). The descriptions provided in these compilations, however, only rarely specify whether any of the activities include compilation of or provision for access to numeric data bases.

It has been far less difficult, not surprisingly, to obtain at least some information about government-operated and government-sponsored numeric data bases than about proprietary ones, even when the latter were identified. Technically oriented business firms are naturally reluctant to disclose many characteristics of their in-house information systems and data bases. However, even the managers of public systems and data bases were notably reluctant and/or poorly equipped to supply information about file sizes, costs, number of users, number of uses, and other measures for numeric data bases. No proprietary scientific numeric data bases were identified with public availability comparable

to the bibliographic and textual data bases available, for example, through SDC or Lockheed.

This report provides information individually about some numeric data bases and collectively about some groups of data bases. There is some information about non-numeric aspects of data bases which also contain numeric data. The information collected about numeric data bases differs markedly from data base to data base.

For some data bases, numbers of items in the file can be quoted and perhaps even growth rates. For some data bases, information is available concerning dollar costs, number of users, or number of queries. For no single numeric data bank did we obtain all of these measures for a single year, much less for time series; no single measure was available for all numeric data bases. Even those data bases for which, for example, the numbers of items had been identified, were incommensurable along this dimension. Indeed, different pieces of the same data collections could not be aggregated nor compared because the units of measure included bits, bytes, records, reels of magnetic tape, reels of videotape, sheets of paper, punched cards, microfiche, and feet of photographic film and "station-months," among others.

The conclusion is inescapable that data are presently lacking which would permit aggregating or reliable determining trends in the size or activity of numeric data base undertakings. In the main body of this report, individual numeric data bases and groups of such data bases have been characterized in those terms and along those dimensions which were elucidated during the study.

Almost exclusively, Federal data base holdings and activities are described. Some foreign data base activities are mentioned in the context of international efforts in which the United States participates.

Many files are produced and/or made available commercially for a fee or for internal use by industry. These files are important to science and technology. However, little information on proprietary files was available, beyond identifying their existence. The publicly available proprietary files that were identified appeared to be so closely tied to bibliographic services or to financial data that they seemed beyond the scope of this investigation.

The centers, files and systems identified in this report do not comprise a complete list. The descriptions contained are limited to a few factors which help to characterize the magnitude of operations. Descriptions of organization, mission and purpose are intentionally brief. Readers interested in additional detail on these characteristics may refer to the documents in the Bibliography and also to the bibliography in the chapter on Numeric Data Bases in the upcoming ARIST volume (60). Itemized listings of files and centers have been provided in the Appendices as an additional aid primarily because such listings are considered useful to future investigators.

Section 2 addresses the general problems encountered and gives a brief description of the survey pre-test. A fuller summary of the pre-test results and a copy of the survey instrument are in Appendix A. Section 3 provides limited information on the level of Federal effort. Section 4 provides an historical background of the development of numeric data compilations. Subsequent sections present the findings either for specific agencies, centers, or systems or for fields of science which rely heavily on numeric data.

SECTION 2

COLLECTION OF DATA ABOUT DATA

The "Indicators" contract originally specified that statistical data be collected from secondary (i.e., published) sources. This limitation, however, prevented the development of adequate indicators of scientific and technical communication, even of that portion which was channeled through the literature. As a result, several surveys were developed and conducted to fill gaps in the information available from published sources.

This study of numeric data also began with a literature search and requests for annual reports, directories, and other listings. Much of the written material that was available addressed the needs, problems, and development of numeric data and dealt extensively with conceptual and technological details of building large (and/or critically needed) numeric data files. A major finding was the existence of a dearth of "data on data." From the beginning, the information from printed publications was supplemented by hundreds of phone calls, letters, and personal interviews. In addition, a survey was also planned but, for reasons stated below, it was not conducted.

2.1 Survey Pre-Test

At the end of 1975, it was evident that sifting through the literature on numeric data was not providing adequate detail to meet the data needs of the study. Not only were inadequate statistical data found on current quantitative levels of activity and effort, but historical statistics which are required to estimate growth rates, were virtually non-existant. A survey of a sample of data centers was clearly indicated.

As a source from which to select a sample of centers, the Second International Edition of the Encyclopedia of Information Systems and Services (56) was chosen. This is one of the more comprehensive directories that was found, especially in its inclusion of centers and systems from all fields of science. This 1,280-page book contains descriptions of about "1,750 organizations concerned with new forms, new media, and new methods for providing information services." It contains 13 indexes. One of these, "Data Collection and Analysis

Centers," was reviewed and generated many of the contacts used in subsequent months. The index contains some, but not all, of the centers included in the National Standard Reference Data System (NSRDS), the Federally-supported Information Analysis Centers (IACs), the Census Service Centers, and the Federally Funded Research and Development Centers (FFRDCs). A number of bibliographic, non-numeric data, and/or non-science data collections are included as well as "service-only" organizations for which no information is provided on the existence of any data holdings. Only about five percent of the entries in the index appear to be non-U.S.

The information contained in the Encyclopedia was obtained directly from each center by questionnaire, letter, or telephone contact, so it may be expected to be reasonably accurate as of the date collected. The location and/or personnel of many centers have subsequently changed. A bibliography, which lists over 50 of the sources used to identify potential entries in the Encyclopedia, is itself a useful reference to an assortment of guides, directories, surveys, registers, etc.

A questionnaire was designed to collect data on file sizes, data sources, fields of science, user populations, sales of products and services, and data publications. Both current (1975) information and historical information (for a decade) were requested. Care was taken to avoid ambiguous wordings; questions were eliminated to which, we judged, many centers would be unable to respond. Funding level was one of the questions eliminated. (See Appendix A for the questionnaire used in the pre-testing.) A small number of centers (22) was selected from the Encyclopedia index to pre-test the survey instrument. An analysis of the responses was expected to provide additional information with which to improve the questionnaire.

Two counterintuitive conclusions were drawn from the responses:

- A summary of the data collected would be meaningless.
- A mail survey with an extensively revised questionnaire would still be unlikely to generate the needed data.

Questionnaires received from the 15 respondents were filled in only spottily. Sufficient data were received from four Federal centers to allow their incorporation into the parts of this report concerning those centers. However, even this was difficult, as identification (by name) of individual files was lacking in the survey. Commercial firms specified that data provided were confidential. In general, respondents did not feel that the information they were providing in the survey was in any way indicative of level of activity, mission goals, or growth. Appendix A contains further description of the problems encountered. For reasons stated here and in Appendix A, the survey was aborted.

2.2 Conceptual and Definitional Problems in Quantification

Pervasive problems exist in fundamental concepts, including definition of some of the dimensions along which to measure data bases. These problems must be solved in order to obtain statistical and other information which might be useful for comparative purposes.

2.2.1 File Size

Although data files are discussed in terms of their size, this characteristic does not indicate their relative importance to the scientific community. For example, data files containing recommended values of fundamental constants* may be relatively small files (and may not be available in machine-readable form). Nevertheless, their critical importance to scientists and engineers is so widely accepted, that substantial effort is expended to publicize and distribute any changes.

Growth rates are also suspect as determinants of importance or utility since some types of numeric files have maximum achievable sizes. Certain historical and geographic files, once they are complete, are of this type. Files of this type would be created for single experiments. Files may remain at the same size despite changes in the contents, e.g., a file of the maximum temperatures ever recorded on each day of the year will never contain more than 366 entries for each weather station. For this type of file, growth rates cannot be projected

*For example, those data files maintained by the centers of the National Standard Reference Data System, National Bureau of Standards.

beyond a certain size. Other types of files can grow indefinitely as new data are entered.

Numbers of files and file sizes are not clearcut terms. One can comparatively easily define the number of books, journals, or articles, or the number of pages, number of words per page, and so forth, when referring to textual material. However, the concept of "a file" is somewhat more nebulous. An individual file may be created from data extracted from a number of other files or may be a summary created from a much larger file. File size may be described in terms of numbers of items of physical recording media, e.g., number of cards, magnetic tapes, or disc packs. It may be measured in terms of numbers of records contained. Or, it may be measured in terms of numbers of bits or bytes.

None of these measures are interchangeable. Not even ready interconversion of measures is possible: in contrast to the rule-of-thumb average, for example, of about six characters per word of running English text, the number of bytes per machine record can in entirely plausible instances vary from about a dozen to many thousands.

In many cases, the computer programs which allow access, manipulation and output are integral parts of the file. Even when they are not, certain quantities of identifiers, labels, and control and edit characters are. Therefore, even a bit count is not necessarily a good indicator of the quantity of <u>data</u> contained in the file.

Newer technology further compounds the file-size problem. A tape library containing 100 magnetic tapes may "double" its data content and still show the same number of tapes (or fewer tapes) due to compression of data.

2.2.2 Number of Users

The relative significance of a file is also not necessarily determined by the number of its users. Maximum potential use of data files (number of users and number of uses) and distribution of data file products is largely governed by both the scope of the data (e.g., useful to scientists in a number of fields)

and the number of scientists in a specific discipline (in cases where the utility of the file is restricted to a narrow sub-discipline). A file may be directly important to only a few scientists but indirectly important to many by the extensive influence of the work of the few.

Growth rates for the numbers of users cannot be interpreted in vacuo. Slow or zero growth rates in the number of users of a file may indicate that the file is "unadvertised" and underutilized, or it may mean that all potential users are already users and the number of scientists in that subject specialty is static. Rapid growth in the number of users may derive from market penetration, growth in the number of specialists, or new importance of the data to an additional specialty. Projections in any case require a market study.

Federally created files frequently pose problems in identifying the true number of users, because they are public information and their use is not controlled by copyright or patent limitations. Any user who obtains a copy is free to reproduce and distribute copies at will. Therefore, counts of direct use may not represent total use.

2.2.3 Costs and Fees

There were very few cases in which even rough approximations of costs were obtainable. To a large extent this was due to the fact that the development and maintenance of the numeric file was an integral part of the operating budget of an organization and not separately budgeted. In many cases files were originally developed as staff tools, and costs were pretty well buried within day-to-day operations. When such a file was tailored to be used by outsiders, the user charges set may or may not have borne a relationship to actual costs of creating and maintaining the file. When a major operation exists for the collection of data, it may be futile to attempt allocating costs to the data base. For example, in the case of a system which monitors and records environmental data; how much of the cost of the data collection equipment can reasonably be included in the cost of creating and maintaining the file?

Procedures for setting user charges vary from flat stated dollar prices for copies of card decks or magnetic tapes (e.g., those set by NTIS*); to hourly charges for use of files (e.g., most commercially available files); to payment such as exchanges of data files, submission of blank tapes onto which files are copied by the provider, or submission of excess blank tapes or cards in exchange for return of "filled" tapes or cards (U.S. Naval Observatory); to files which are provided completely free to the user.

The material presented in this report is not complete not only because some numeric data bases are not covered; the characterization of some that are included represents only a portion of actual activity in their creation, use, and costs. In most situations caveats concerning the problems of measuring numeric data files are irrelevant as no measures were readily available, certainly not without a far more extensive research effort.

*National Technical Information Service, Department of Commerce.

SECTION 3

LEVELS OF FEDERAL INVOLVEMENT

There is a reasonably convincing argument that in terms of investment of effort, the bulk of activity is on a national and international scale. Certainly, the kinds of numeric files which contain critical values in the physical and engineering sciences are available on a worldwide scale. Regional, national, global, and space data (i.e., demographic, astronomical, environmental, and other phenomenological) require government support in order for both the collection effort and the distribution to be broad enough in scope to be meaningful.

Although no information is currently available on the total Federal level of funding of numeric data activity, an examination of Federal legislation in recent years indicates that the proportion of bills calling for the establishment of information programs and numeric data bases had increased from one percent in 1966 to five percent in 1974 (113).

For a small subset of those centers, programs, and files described in this report, funding levels were determined, as shown in Table 3.1. Exactly what is included as funding varied considerably from contractor costs (which may include activities other than data handling and does not include the data handling operations of the sponsoring agency) to total appropriation. The study findings do not allow any statement as to what proportion of total Federal involvement is represented by the amounts shown. It is as likely to be ten percent as it is to be 50 percent. It represents at best a minimum commitment level by the government.

If scientific and technical information (STI) is the end product of scientific and technical investigation--of all research and development--then scientific and technical data most certainly make up a set of byproducts as complex, as large, and growing as rapidly as STI.

It appears, following investigation of the area, that only the "tip of the iceberg" is being widely discussed in the literature of numeric data. Certainly within the Federal government, a majority of data-handling (collection, storage, analysis, dissemination) is internal, and the products are not readily

Table 3.1 ANNUAL FUNDING LEVELS FOR SELECTED FEDERAL SCIENTIFIC DATA PROGRAMS

Program	Millions of Dollars	Fiscal Year	Type
Principal Federal Statistical Programs	458	1975	Obligations
National Climatic Center	16	1976	Obligations
National Meteorological Center			
EPA Data Programs	11	1976	Obligations
USDA Research ADP	10	1975	Projected Obligations
NCI-SEER	6	1976	Contractor Costs
Federal Mapping & Cartography	305	1975	Obligations
ERDA Technical Information Services Electronic Systems	1	1977	Obligations
Federal IACs	85	1975	Projected Expenditures
NODC/NOAA	3	1975	Appropriation plus Reimburseables and Transfers
NSSDC/NASA	2	1975	
EDS/NOAA	7	1975	
NSRDS	3	1975	
National Air Data Branch (EPA)	1	1975	
Total	908		

SOURCE: King Research, Inc.

available to the general scientific public. As systems and procedures are refined, these same data bases may become publicly available. However, the funding levels of such operations are currently buried within the operating budgets of individual agencies. Other measures of the magnitude of these operations are also not separately stated.

Between 1962 and 1970, the National Science Foundation collected and reported data on the level of Federal obligations for "General-Purpose Scientific Data." For this data, NSF used the following definition:

> "General-purpose scientific data are those which, either separately or in combination with other information, can be applied to useful, general scientific purposes. Included as general-purpose scientific data are statistics, observations, specimens, readings, or other facts gathered from surveys, field investigations, or compilations of operating records, and similar types of information.
>
> Excluded are data used solely for internal administrative or operating purposes. General-purpose scientific data may pertain to the physical, life, mathematical, engineering, and environmental sciences as well as to the psychological and social sciences. These data are used by many organizations and individuals, including public agencies--Federal, State, and local--private foundations or associations, research investigators, and the general public."

Volume XX of Federal Funds for Research and Development and Other Scientific Activities (83) contains the last tabulation of these data, and provides estimates of them through 1972. When questioned about the cessation of scientific data in the series, NSF personnel mentioned three factors: lack of interest in the results, inadequacies in the definition, and (partially as a result) unreliability of the data.

Not all of these problems have been fully resolved for the STI data which are still being collected and reported! The comparison of the two data sets in Table 3.2 indicates that scientific data consistently exceeded STI during the decade. Erratic fluctuations in the rates of increase for scientific data could result from real differences or from definition and reliability deficiencies. However, the large increase in 1970 is caused by the activity surrounding the 1970 Census. Federal STI funding levels are currently (1977) estimated at $446 million in Volume XXIV of the above mentioned source (83). Projections of Federal STI funding are found in Volume V of "Statistical Indicators of Scientific and Technical Communication" (53), where the 1980 level is shown as $624 million.

Table 3.2 FEDERAL OBLIGATIONS FOR SCIENTIFIC INFORMATION AND DATA: 1962-1972

(Millions of dollars)

Year	Scientific and Technical Information		General-Purpose Scientific Data	
	Funds	Annual % Change	Funds	Annual % Change
1962	128	n.a.	220	n.a.
1963	164	28	268	22
1964	203	24	309	15
1965	225	11	343	11
1966	278	24	325	-5
1967	324	17	368	13
1968	359	11	380	3
1969	362	1	385	1
1970	387	7	550	43
1971[e]	400	4	526	-4
1972	413	3	537	2

SOURCE: National Science Foundation, Federal Funds for Research and Development and Other Scientific Activities, 1971.

SECTION 4

HISTORICAL HIGHLIGHTS IN NUMERIC DATA COMPILATION

Compilation of numeric data goes well back to ancient times. Thus astronomic observations allowed the Mayans, Egyptians, and Babylonians to establish 365/366-day calendars very early in history. Rabbi Hillel was able to shift the complex lunar-solar Jewish calendar from an ad hoc observational basis to a predictive basis some 1900 years ago. Extensive compilations of astronomical data were available to Copernicus, Drake, and Gallileo, not to mention the early navigators. Tables of astronomic values to this day constitute an important aid to navigation.

In the rest of this section are mentioned some published forms of numeric data in science and technology and some of the organizations which conduct data activities. They promote scientific and technical data compilation and establish standards to assure that data in the several fields of science and technology are intercomparable.

4.1 Published Data

The collection and recording of numerical data has traditionally been an integral part of scientific and technical research and development. Historical reviews of the collation and formal dissemination of numerical values in tabular format* usually begin with the publication in 1883 of the first edition of the Landolt-Börnstein Tabellen in Germany. This edition contained fewer than 300 pages. By the sixth edition in the 1950's it had grown to about 2,000 pages. Two other early landmark documents were the French Tables Annuelles de Constantes et Données Numériques which appeared in 10 volumes between 1910 and 1930. Eight volumes of the International Critical Tables of Numerical Data of Physics, Chemistry, and Technology appeared in the United States between 1926 and 1933. Copies of this work are still sold and in use today. In fact, a 1965 survey of American Chemical Society members indicated that half of the respondents cited this document as the "most frequently consulted data compilation

*All material in Section 4.1, unless otherwise noted, is from CODATA Bulletin #16 (19).

for property data. In the mid-1950's it was apparent that updating of such a comprehensive work was impossible, in spite of the demand. Subsequent efforts to produce data compilations leaned heavily toward more selected data sets, resulting in the publication of a great number of handbooks designed for use by scientists and engineers in specific disciplines or subdisciplines.

Other noteworthy handbooks published in the U.S. and abroad include:

Kay and Laby Tables of Physical and Chemical Constants (1 volume, 13 editions, 1911-1967)

Tables de Constantes Selectionées

Handbook of Biological Data

Tabulae Biologicae

Biological Handbook Series

Biochemists' Handbook (1961, 1192 pages)

Handbook of Chemistry and Physics (1968, roughly 50 editions)

Handbook of Biochemistry (1968-70, 2 editions, 2nd edition 1600+ pages)

Handbook of Microbiology (1973, 4 volumes, 3000 pages)

Document Guigy Scientific Tables (-1968, 7 editions, 800 pages)

Atlas of Protein Sequence and Structure (1969 and 1972 editions were each 100% larger than previous editions)

Standard Handbook for Electrical Engineers (ref. 22)

Basic Data of Plasma Physics (ref. 22)

Crystal Data Determinative Tables (1972-73, 3rd edition) (ref. 15)

Engineering Properties of Ceramics (ref. 128)

Handbook of Electronic Materials (ref. 128)

This list is by no means complete, it is merely a sampling of important and widely used handbooks.

In addition to the large number of handbooks now in use, a more recent development in published numeric data is the ever-expanding list of journals that concentrate largely on the publication of data tabulations, rather than solely narrative articles. The Journal of Physical and Chemical Reference Data first appeared in 1972. This quarterly publication is a joint effort of the American Chemical Society, American Institute of Physics, and the National Bureau of Standard's National Standard Reference Data System. Other examples of data journals of international scope are:

Atomic Data and Nuclear Data Tables
Journal of Chemical and Engineering Data
Atomic Data
Organic Magnetic Resonance
International Journal of Chemical Kinetics
Journal of Chemical Thermodynamics
Mass Spectrometry Bulletin

There are many others.

In some cases it is difficult to make a distinction between handbooks and journals since updates to handbooks (or to selected portions) may appear in journals associated with them. Subsequently, these update tables may be combined in a new edition of the handbook.

Several other aspects of published data need also to be included. The first is the seemingly endless quantity of statistical data which are periodically generated and updated on a local, national, and international scale. Although many of these data may be considered only marginally within the scope of science and technology, since they cover not only a great deal of industrial information, but many aspects of social science (such as population and economics) as well; it seems unwise to exclude all such statistical activity. It was, furthermore, the U.S. Bureau of the Census that introduced machine-readable data to deal with the problems of handling large numeric data bases. Additionally, most of the mission-oriented programs of recent years combine such statistical data with data from the "hard" sciences. No attempt is made in this section to list even a sample of such data programs. Several are described in subsequent sections of this report.

Another area of tabular data (as opposed to narrative information) includes the compilations of standards and/or taxonomies which are generally non-numeric but which are necessary for interpretation of much numeric data. Examples include:

Le Système International d'Unités (SI) (NBS translation is Special Publication 330)

Symbols, Units and Nomenclature in Physics (1965)

Manual of Symbols and Terminology for Physico-Chemical Quantities and Units (1969)

International Code of Nomenclature of Bacteria (1966)

International Code of Botanical Nomenclature (1966)

International Code of Zoological Nomenclature (1961)

CBE Style Manual (1972)

International Glossary for Hydrological Investigation of the Soil

Rules for IUPAC Notation for Organic Compounds (1961)

The Wiswesser Line-Formula Chemical Notation (1968)

To all of these types of publications, must be added those which attempt to index sources of data. These are not abstracting and indexing services in the sense that they provide access to the literature - nor even to the published data; rather they tend to provide references to sources of data compilations; to the organizations and their "unpublished" compilations.

The classic example is the International Compendum of Numerical Data Projects (14) which, while admittedly incomplete, listed 150 projects in 26 countries. The Third Consolidated Guide to World Data Centers (96), "Geological Data Files, Survey of International Activity" (13), and "A Catalog of Compilation and Data Evaluation Activities in Chemical Kinetics, Photochemistry and Radiation Chemistry" (Codata Bulletin #3) have also been published by ICSU/CODATA. Critical Data in Britain and the second edition, Data Activities in Britain, identified 100 projects. The Directory of Federally Supported Information Analysis Centers (79) identifies about 100 U.S. projects, but not all of these produce data compilations. The Encyclopedia of Information Systems and Services (56) lists 1,750 organizations of which almost 25 percent are indexed as data collection and analysis centers. A 1974 Directory of Data Bases in the Social and

Behavioral Sciences (110) describes almost 700 centers (80 percent in the U.S.) holding files in the social sciences alone. Within the Federal government there are a large number of directories of data bases, data tapes, information sources and resources, and so on. None of the above directories purport to be complete and all are somewhat out of date by the time they are published.

4.2 The Role of Discipline-Oriented Societies

For over a century, professional societies have been active in addressing the issues posed by the need for reliable numeric data. Societies already in existence paid progressively more attention to such data, and new societies were founded with major orientation to data collection, organization, and interpretation. A few of the well-known U.S. and international associations include (19):

Europaeische Gradmessung (geodesy, founded 1864)
International Association of Geodesy (founded 1886)
World Meteorological Organization
American Chemical Society
American Institute of Physics
Federation of Astronomical and Geophysical Services
International Astronomical Union
International Telecommunications Union
American Petroleum Institute
Manufacturing Chemists Association
Institute of Electrical and Electronics Engineers
World Federation of Engineering Organizations
World Federation of Culture Collections
International Association on the Properties of Steam
International Union of Pure and Applied Physics
American Institute of Biological Sciences
International Atomic Energy Agency

Quite obviously from the titles, some of these are individual member organizations, some are associations of organizations, and others approach the status of private research institutes or government or intergovernmental agencies. The formation of national and international organizations, while a century-old phenomenon, appears to be accelerating as "sub-disciplines" and "cross-disciplines" evolve.

4.3 Other Organizations

Among the mission-oriented associations and agencies are included a variety of types of activities. The classical "mission-orientation" combines various aspects of a multitude of disciplines. Among these are organizations built around such broad topics as "environment", "health", "defense", "transportation". In addition to these classical missions, we are inclined to add some that are specific to scientific information and especially scientific data. So perhaps we need also to include, for example:

General Conference on Weights & Measures
American Documentation Institute
World Health Organization
United Nations' UNISIST
ICSU's CODATA and World Data Centers
American National Standards Institute
NAS-NRC's Office of Critical Tables and, later, Numerical Data Advisory Board
American Society for Testing Materials

and an assortment of organizations in the areas of computer technology, programming and standards, telecommunications, social statistics, and satellite and other remote data collection mechanisms. All have a vested interest in data collection, analysis, documentation, and standardization.

Not mentioned thus far is the growing number of data centers, be they Information Analysis Centers, data compilation projects, data dissemination organizations, or referral centers, nor data activities which produce numerical data for the internal use of the organization producing it. Many

government agencies collect, compile, store, and disseminate numeric data. During its existence, the Interagency Committee on Scientific and Technical Information (COSATI) struggled with procedures and methods to coordinate both information and data activities and even to standardize some aspects of these activities. The one government agency which must be singled out because of its mission is the National Bureau of Standards, which was, in the 1960's, briefly but not officially referred to as the National Institutes of Science and Technology. It was founded in 1901 specifically to develop and to promulgate reliable methods to measure quantifiable phenomena; this remains an important part of its mission.

The above discussion barely scratches the surface of an immense and growing activity in numeric data. Advances in technology which made possible the exploration of space, the examination of ever-smaller particles of matter and the measurement of previously elusive biological phenomena are added to electronic sensing, recording, and data storage technologies to create a situation in which the quantity of data which "could" be collected is far larger than could foreseeably be used. The growth in the quantity of data is accompanied by a growth in the number of organizational entities concerned with them. Efforts are launched to improve analytic and compression capabilities. Solutions to environmental and social problems require ever increasing amounts of data. Legislative mandate creates new agencies with new requirements for data and also creates special agencies specifically responsible for data handling.

SECTION 5

NUMERIC DATA AND MAJOR DISTRIBUTION AGENCIES

There are three United States Government agencies which do not themselves compile scientific or technical data, but which play major roles in disseminating numeric data compiled by or for other parts of the Government. The Government Printing Office (GPO), the Defense Documentation Center (DDC), and the National Technical Information Service (NTIS) are all major central distribution agencies for information products. A fourth agency, the National Archives and Records Service (NARS), serves primarily as a repository of records but also disseminates copies. In all of these cases dealing with numeric data is only a small fraction of the agency's overall activity. It is, however, that part of their activities which is here pertinent.

So far, GPO deals exclusively with publishing and distributing printed documents. Although these documents include maps, graphs, and reproduction of computer printouts and other tabulations, GPO itself is not involved in the compilation or computer generation of these products nor in the sale of machine-readable products.

The Defense Documentation Center acquires and distributes copies of reports generated by the Department of Defense and its contractors. Non-classified materials also become part of the NTIS inventory. DOD technical information includes the products of many Information Analysis Centers and other data-generating activities. Recently DDC has begun to explore its potential for handling numeric data materials other than published reports.

Sales of magnetic tape and card decks have been a small but growing part of NTIS for several years. NTIS processes primarily government reports and some journals. Its prominent position as a central source of documentation has been instrumental in the recent naming of NTIS as the Federal Software Depository.

5.1 Data Files Available through NTIS

NTIS has an extensive announcement system. This system has been and continues to be predominantly concerned with the distribution of technical reports. However, there has been a growing coverage of machine-readable data files over the past few years. The data files which can be purchased from NTIS are announced, along with report literature, in the Weekly Government Abstracts. In addition, in recent years NTIS has produced three directories of data files. Each succeeding edition has included a broader coverage. The contents of these directories are of particular interest as they describe a set of data files which are widely accessible.

An initial objective was to compare the three directories in terms of the number of entries for each field of science and in terms of the number of generating agencies. The intent was to get a general trend of the growth rate. Comparisons based on designations of the field of science proved difficult due to changing designations used by NTIS for the fields; the three tables of contents are not consistent. In addition, we noted (accidentally) that a particular set of tapes (the USDA Survey of Purchases of Fish Products) were included in the sections for "Marketing" in 1973, for "Consumer Affairs" in 1974, and for "Agriculture & Food" in 1976. All three designations have a certain rationale; neither would it have been surprising to find them included instead under "Commerce", "Economics", "Ocean Science", or even "Income, Expenditures & Health" or "Health Statistics" (assuming nutrition as part of health). Part of the problem is the innate difficulty of assigning a data file to any single category, especially if there is any conceivable cross-disciplinary or mission-oriented character to the file. An additional dimension is added when we realize that the nature of many of the tapes is Federal statistical information and related products.

The three directories are:

NTIS Software & Data Files, 1973 (not widely distributed) (89)

Directory of Computerized Data Files and Related Software, 1974 (87)

A Directory of Computerized Data Files, Software and Related Technical Reports, 1976 (88)

Summaries of the characteristics of the entries in the three directories are presented in Tables 5.1 through 5.3.

The 1973 document is called a catalogue and includes only items obtainable from NTIS. It contains 248 entries in 19 categories. Of these, 144 entries were available in machine-readable (tape or card) form. Of these, 96 were data files (i.e., not exclusively software nor bibliographic) and 32 were software. The 96 data files contained 120 reels of tape and 8,400 cards (8 decks).

The 1974 document is subtitled "available from Federal agencies" and includes a substantial portion of entries which are not directly available from NTIS. It contains 524 entries in 42 categories (an additional 16 categories are provided in the list of subject fields, but contain no entries). Virtually all of the entries were in machine-readable form (only one was questionable) although many were also available as hard copy or microforms. We identified five entries which were for "data center" operations and for which no additional information was available (from this source) on file sizes. Four hundred and eleven entries had data file products, while 89 entries referred to software only. The data files (not counting the products of the data centers) contained 3,782 reels of tape and 36,300 cards.

The above information is for all fields. NTIS personnel estimate that the 1976 directory contains approximately 1200 entries. It contains 46 subject field categories. From these categories entries in only 28 were examined (one in economics, four in social science, and all 23 of the science and technology fields). Out of the 745 entries in these fields, 558 were machine-readable. Of these, 257 are data files (as identified in the directory itself), an additional 18 are "data base reference services", and 272 are software products.

According to NTIS, the number of Federal agencies included in the 1974 directory was about 50, while 75 were included in 1976. NTIS' move to become the Federal Software Center may certainly be expected to increase its holdings of software programs. How much impact this move will have on data files remains

Table 5.1 SUMMARY OF ENTRIES IN NTIS SOFTWARE & DATA FILES, 1973

Subject Field	Total Entries	Machine-Readable Entries				Data Files	
		Total[1]	Software Only	Data Files[2] No.	Data Files[2] % of Total Entries	Tape Reels	Cards (000)
1. Bibliographic	2	2	0	0	-	-	-
2. Chemistry	10	7	2	5	50	7	-
3. City Games	5	3	0	0	-	-	-
4. Communications	8	7	2	5	62	24	-
5. Computer Software	16	9	0	0	-	-	-
6. Corporate Finance	2	1	0	0	-	-	-
7. Demography	30	21	0	21	70	23	-
8. Earth Sciences: Cartography	9	5	3	2	22	2	-
9. Earth Sciences: Gas Sample Analysis	1	1	0	1	100	1	-
10. Earth Sciences: Geochemistry	7	5	0	5	71	5	-
11. Earth Sciences: Tunneling & Rock Structure Analysis	6	2	2	0	-	-	-
12. Environmental Pollution & Control	28	13	11	2	7	2	-
13. Federal Supply System	8	8	0	8	100	8	-
14. Industrial Growth & International Commerce	20	11	4	8	40	11	-
15. Library & Information Science[3]	8	6	1	5	62	1	8
16. Marketing[4]	12	9	0	9	75	11	-
17. Medical Science	12	8	2	5	42	5	-
18. Transportation	43	9	5	3	7	3	-
19. U.S. Budget	21	17	0	17	81	17	-

See footnotes at end of table.

(Continued)

Table 5.1 (cont.) SUMMARY OF ENTRIES IN NTIS SOFTWARE & DATA FILES, 1973

Subject Field	Total Entries	Machine-Readable Entries		Data Files		Data Files	
		Total[1]	Software Only	No.	% of Total Entries	Tape Reels	Cards (000)
Sub-total "S&T"[5]	149	93	23	60	40	81	8
% of Total	60	65	72	62	--	68	100
Total	248	144	32	96	39	120	8

[1]Difference between 'Total' & sum of 'Software' + "Data" is bibliographic files; Difference between "Total Entries" and total "Machine-Readable Entries" is reports in hard copy (or microform) only.

[2]Includes combination of software & data.

[3]Mostly social science & geographic code information.

[4]Mostly agricultural & income tax information.

[5]Fields 2, 4, 5, 7-12, 15-17.

SOURCE: King Research, Inc. (Based on: NTIS Software and Data Files, 1973, National Technical Information Service, 1974.)

Table 5.2 SUMMARY OF CONTENTS OF DIRECTORY OF COMPUTERIZED DATA FILES AND RELATED SOFTWARE, 1974

Subject Field[1]	Total Entries	Machine-Readable Entries		Data Files[3]		Data Files	
		Total[2]	Software Only	No.	% of Total Entries	Tape Reels	Cards (000)
Demography, Social Science & Government	126	126	12	113	90	2,768[4]	1
2. Consumer Affairs	9	9	0	9	100	10	0
3. Education	8	8	0	8	100	25	0
5. Environment & Geography	4	4	4	0	-	-	-
6. Government Publications	1	1	0	0	-	-	-
8. Law Enforcement & Justice	4	4	0	4	100	3	1
9. National Defense	1	1	0	1	100	44	0
10. Population	50	50	8	42	84	2,378	0
11. Public Lands, Parks, Recreation & Travel	9	9	0	9	100	23	0
12. Social Insurance & Welfare Services	23	23	0	23	100	32[4]	0
14. Vital Statistics	11	11	0	11	100	26	0
15. Quick Query Service	6	6	0	6	100	227	0
Economics	209	209	8	201	96	900[4]	23
21. Banking, Finance & Insurance	5	5	1	4	80	13	0
22. Business Enterprise	32	32	0	32	100	41[4]	18
23. Comparative International Statistics	2	2	0	2	100	13	0
24. Federal Government Finances & Employment	21	21	0	21	100	24	0
25. Foreign Aid & Commerce	25	25	0	25	100	223	0
26. Income Expenditures & Wealth	21	21	0	21	100	24	0
27. Housing & Construction	3	3	0	3	100	29	0
28. Labor Force, Employment & Earnings	44	44	0	44	100	321	0
29. Price Statistics & Price Indexes	1	1	0	1	100	1	0
30. State & Local Government & Finances	8	8	0	8	100	48	0
31. Transportation	47	47	7	40	85	163[4]	4
Science & Technology	189	188	69	98	52	115[4]	12
41. Aeronautics & Aerodynamics	8	8	1	7	88	6	4
42. Agriculture & Food	13	13	1	5	38	4	0
43. Astronomy & Astrophysics	1	1	0	1	100	2	0
44. Atmospheric Sciences	11	11	0	11	100	11[4]	0

See footnotes at end of table.

(Continued)

Table 5.2 (cont.) SUMMARY OF CONTENTS OF DIRECTORY OF COMPUTERIZED DATA FILES AND RELATED SOFTWARE, 1974

Subject Field[1]	Total Entries	Machine-Readable Entries: Total[2]	Software Only	Data Files[3]: No.	Data Files[3]: % of Total Entries	Data Files: Tape Reels	Data Files: Cards (000)
45. Behavioral & Social Sciences	1	1	0	1	100	1	0
46. Biological & Medical Sciences	31	31	3	25	81	57[5]	0
47. Biomedical Technology & Engineering	4	4	0	4	100	4	0
49. Chemistry	7	7	2	5	71	7	0
50. Civil, Structural & Marine Engineering	11	11	11	0	-	-	-
51. Communications System	10	10	3	7	70	25	4
52. Computers, Control & Information Theory	25	24	23	1	4	0	0
53. Earth Sciences	25	25	7	18	72	37[4]	0
55. Geocoding & Dictionary Files	6	6	0	6	100	3	4
58. Libraries & Information Science	10	10	0	1	10	1[5]	0
59. Materials Sciences	1	1	0	0	-	-	-
60. Mathematical Sciences	11	11	11	0	-	-	-
64. Ocean Science & Technology	1	1	0	1	100	1[4]	-
65. Physics	1	1	0	0	-	-	-
69. Simulation & Models	2	2	1	1	50	2	0
72. Urban Technology	10	10	6	4	40	4	0
Subtotal "S&T"[6]	307	306	81	204	66	3,081[4,5]	12
% ot Total	59	59	91	50	-	81	33
Total	524	523	89	412	79	3,783[4,5]	36

[1]The following fields listed in the Directory are not included in the table as they contained no entries:

Demography, Social Science & Government
- 1. Congress, Legislation and Committees
- 4. Elections
- 7. Immigration
- 13. Veteran Affairs

Science and Technology
- 48. Building Technology
- 54. Electrotechnology
- 56. Industrial Engineering
- 57. International Relations
- 61. Mechanical Engineering
- 62. Nonpropulsive Energy Conversion
- 63. Nuclear Science & Engineering
- 66. Propulsion, Engines & Fuels
- 67. Reprography & Recording Devices
- 68. Safety Engineering & Protection
- 70. Space Technology
- 71. Test Methods, Instrumentation & Equipment

(Continued)

Table 5.2 (cont.) SUMMARY OF CONTENTS OF DIRECTORY OF COMPUTERIZED DATA FILES AND RELATED SOFTWARE, 1974

[2]Difference between 'Total' & sum of 'Software' + "Data" is bibliographic files; Difference between "Total Entries" and total "Machine-Readable Entries" is reports in hard copy (or microform) only.

[3]Includes combination of software & data.

[4]Plus an unspecified number. Usually indicates that at least one entry is a data center, or an agency which provides tailor-made tapes of available data.

[5]Plus unspecified number of files available on-line.

[6]Fields 5, 10, 14, 15, 23, 28, 29, 41-72.

SOURCE: King Research, Inc. (Based on: National Technical Information Service, Directory of Computerized Data Files and Related Software, 1974.)

Table 5.3 SUMMARY OF ENTRIES IN SELECTED FIELDS FROM A DIRECTORY OF COMPUTERIZED DATA FILES, SOFTWARE AND RELATED TECHNICAL REPORTS, 1976

Subject Field[1]	a. Total Entries	Machine-Readable Entries				Data Files	
		b. Total[2]	c. Software Only	d. Data Files[3] No.	d. Data Files[3] % of Total Entries	Tape Reels	Cards (000)
Economics	42	40	1	38	90	152	0
5. Employment, Earnings & Labor	42	40	1	38	90	152	0
Social Sciences	173	119	39	77	45	2,458	1
11. Demography & Population	48	48	4	43	90	2,358[4]	0
17. Library & Information Sciences	36	[illegible]	7	21	58	72[5]	1
18. Municipal Information Systems	67	27	27	0	-	-	-
23. Vital Statistics	22	15	1	13	59	28	0
Science & Technology	530	399	233	159	1	3,621	7
24. Aerodynamics & Fluid Dynamics	6	4	4	0	-	-	-
25. Agriculture & Food	29	25	3	16	55	24[4]	0
26. Atmospheric Sciences & Astronomy	5	5	1	4	80	2,653[4]	0
27. Behavioral Sciences	3	2	0	2	67	2	0
28. Biological Sciences	11	11	7	4	36	14	0
29. Cartography	27	20	18	2	7	16	0
30. Chemistry	14	10	1	9	64	12[4]	0
31. Civil & Structural Engineering	77	60	59	1	1	1	0
32. Communications	43	37	3	34	79	136	0
33. Computer Sciences	73	40	37	2	3	2	0
34. Electrotechnology	4	4	4	0	-	-	-
35. Energy Sources, Generation & Transmission	27	25	12	13	48	17[4]	0
36. Environmental Pollution & Control	40	27	19	8	20	14[4]	0
37. Industrial & Mechanical Engineering	3	3	3	0	-	-	-

See footnotes at end of table. (Continued)

Table 5.3 (cont.) SUMMARY OF ENTRIES IN SELECTED FIELD FROM A DIRECTORY OF COMPUTERIZED DATA FILES, SOFTWARE AND RELATED TECHNICAL REPORTS, 1976

Subject Field[1]	Total Entries	Machine-Readable Entries				Data Files	
		Total[2]	Software Only	Data Files[3] No.	Data Files[3] % of Total Entries	Tape Reels	Cards (000)
38. Materials Sciences	18	15	14	1	6	1[4]	0
39. Mathematics & Statistics	26	19	17	2	8	2	0
40. Medical Sciences	31	26	7	18	58	16[4]	3
41. Natural Resources, Geology & Hydrology	49	37	4	33	67	261	0
42. Navigation & Guidance	6	5	1	4	67	3[4]	4
43. Ocean Science & Technology	12	7	3	4	33	445	0
44. Physics	5	5	4	1	20	1	0
45. Soil & Rock Mechanics	19	10	9	1	5	1	0
46. Test Methods & Metrology	2	2	2	0	-	-	-
Total "S&T"	745 (62% of total entries)	558	273	274	37	6,231[4,5]	8

[1] The following fields are listed in the directory but entries were not examined:

Economics

1. Banking, Finance & Economic Trends
2. Business Enterprise
3. Commerce - United States
4. Commerce - Foreign
6. Finances - Federal Government
7. Finances - Local & State Government
8. Housing & Construction
9. Income, Expenditures & Wealth

Social Sciences

10. Consumer Affairs
12. Education
13. Government Administration - Federal, State & Local
14. Health Care
15. Health Statistics
16. Law Enforcement & Criminal Justice
19. Public Land, Parks, Recreation & Travel
20. Social Services
21. Transportation - Air
22. Transportation - Surface

[2] Difference between 'Total' & sum of 'Software' + "Data" is bibliographic files; Difference between "Total Entries" and total "Machine-Readable Entries" is reports in hard copy (or microform) only.

[3] Includes combination of software & data.

(Continued)

Table 5.3 (cont.) SUMMARY OF ENTRIES IN SELECTED FIELD FROM A DIRECTORY OF COMPUTERIZED DATA FILES, SOFTWARE AND RELATED TECHNICAL REPORTS, 1976

[4]Plus an unspecified number. Usually indicates that at least one entry is a data center, or an agency which provides tailor-made tapes of available data.

[5]Plus unspecified number of files available on-line.

SOURCE: King Research, Inc. (Based on: A Directory of Computerized Data Files, Software and Related Technical Reports, 1976, National Technical Information Service, 1976.)

to be seen. However, the fact that many large publically-supported data centers; the WDC-As* particularly, recognize that one of their major problems is the (lack of) marketing of products and services could serve as an impetus to marketing through NTIS.

5.2 DDC Looks at Fact Services

One of the objectives recommended in a DDC 10-Year Requirements Study (49) is the provision of "fact services". A few comments from this study are presented here since they are relevant to numeric data development. Fact services, as defined in the DDC study, include numeric data plus other "answer providing" services and management information services. The results of a survey of DDC users provided rationale for specifying this objective as, "users are convinced that much information is unreviewed, unverified and unreliable," users "will need more information by 1980", and, "fact-type information is not easily available to users currently" outside of handbook type materials which are "not as up-to-date as needed".

The tasks defined to meet this objective include the creation of a machine-readable data base of critical and most heavily used data. To be included in this data base are the evaluated data from the NSRDS.

Over 50 percent of DDC users have used at least one Information Analysis Center (58 percent used the Infrared Information and Analysis Center, smaller percentages used other centers). Nevertheless, the results of asking these users about preference for format and media are not encouraging for computer-generated numeric data. This fact may underscore that marketing is one of the major issues in providing numeric data. Less than 10 percent of the survey respondents preferred handbook/manual or numeric data forms. Fewer than 1 percent preferred numeric data alone. Those choosing computer-readable or computer output media (printouts, cards, tapes, and cathode ray tube) comprised just over 3 percent. Most preferred textual material such as technical reports and journal articles.

*See Section 7.2 (World Data Centers)

Nevertheless, the report summarizes the "fact services" issue for DDC as follows:

> The most conspicuous gap in [DDC] information services, according to the Expert Panels, is the provision of fact information services. "Fact information" includes numeric values and also pieces of discrete data capable of satisfying inquiries without further reference. The technology is predicted to be capable of supporting such a service, and it is rated most desirable and extremely important, yet the probable timing places this as a mid- to long-range event. [Note: mid-range = 1985-1995, long-range = after 1995]
>
> This event must be regarded as a high-payoff area for DDC to pursue. There appears to be no explanation for the lack of progress in fact services other than neglect on the part of the information planners and designers.

5.3 National Archives and Records Service

The National Archives and Records Service (NARS) is responsible for the preservation and availability of basic records of the Federal government. The records for which NARS assumes this responsibility are those relating to the

- necessary processes of government
- protection of public and private rights
- interests of scholars, students, and the general public.

A Data Archive Staff was established in 1969. In 1975 they produced their first widely circulated (2,000 copies) Catalogue of Machine-Readable Records in the National Archives of the U.S. (8) One thousand one hundred and fifty (1600 bpi) tapes had been acquired. Since the NTIS (1974 and 1976) data file directories were prepared in cooperation with NARS, it is assumed that all of the files in the NARS catalogue also appear in the NTIS directory. Tape holdings at NARS* have increased from 300 (800 bpi) in 1972, to 600 (800 bpi) in 1973 and 1,200 (1600 bpi) in 1974.

*As reported to King Research, Inc. on the data collection form.

The files in the NARS catalogue all fall in the social sciences field. They consist of both statistical and administrative data. NARS provides copies (on tape or card) with few restrictions. They handle the public distribution of all data files of the Civil Aeronautics Board (CAB) and Securities & Exchange Commission (SEC). It is not clear whether any of the same tapes are available from both NARS and NTIS.

In 1969-70 NARS conducted a census (104) of tape reels in the Federal government. Five and one-half million reels were identified, over half of them in the Department of Defense. The Bureau of the Census and the Social Security Administration held approximately 200,000 reels each. The census of tape reels concentrated on files in the Washington, D.C. area and is admittedly incomplete, especially for agencies located in other parts of the country. Approximately 300 file series (in which the 1960 Census, for example, is one "file series") were located by NARS which were deemed worthy of accessioning.

Sales from archival tapes since 1970, when the first magnetic tape holdings were acquired, were: FY1971 - 50 reels, FY1972 - 130 reels, FY1973 - 150 reels (104). The addition of CAB files increased sales to the point where the first four months of FY1974 showed sales of 200 reels. Although sales appear small compared to the large number of files stored by NARS, most Federal agencies do not depend upon the Archives for active or current data distribution and only provide tapes to NARS when the usage of the files has dropped to the level of considering them archival only. The January 1974 article from which much of the above detail was obtained refers to burgeoning data activity in environmental monitoring and NARS' desire to remain cognizant of such efforts. The subsequent years can be expected to be ones in which Archives has broadened its holdings into fields other than social sciences. For instance, NASA's National Space Science Data Center currently (1976) archives many tapes with NARS rather than in its own facility.

Additional information collected in the data collection pre-test by KRI is not completely compatible with that reported above by Rosencrantz and follows:

Table 5.4 NATIONAL ARCHIVES AND RECORDS SERVICE REQUESTS AND SALES: 1972-75

Item	1972	1973	1974	1975
Requests	3	30	107	188
Sales ($000)	0	4	17	30

SOURCE: King Research, Inc.

SECTION 6

INFORMATION ANALYSIS CENTERS AND RESEARCH AND DEVELOPMENT CENTERS AS POTENTIAL SOURCES OF DATA

Federally Funded Research and Development Centers (FFRDCs) are wholly or largely Federally supported and include some but not all of the "National Laboratories". They are administered extramurally by universities, industrial firms, or non-profit institutions. Similar characteristics apply to Federally supported Information Analysis Centers (IACs). In a number of situations, one or more IACs are housed within or attached to an FFRDC. Even when this is the case, no consistent pattern of relationship was identified. A list of FFRDCs is provided in Figure 6.1, indicating both sponsoring agency and organizational affiliation. In 1968 Kertesz (50) identified 21 IACs within the Oak Ridge National Laboratory (ORNL) FFRDC alone (See Figure 6.2).

It can safely be assumed that both IACs and FFRDCs handle numeric data. IACs may also handle documents rather than data. Although FFRDCs certainly generate numeric data for internal use as part of the research process, they may not necessarily be accumulating them in machine-readable or externally useable formats. The Lawrence Berkeley Laboratory, however, is investing heavily in a large, general-use social-science-environment data base (SEEDIS).

The definition of Information Analysis Center implies an information coordinating and marketing emphasis which would put IAC operations more securely into the scientific and technical information area than are those of the more basically research oriented FFRDCs.

The following definition for IACs was adopted by the former Committee on Scientific and Technical Information (COSATI) and appears in the 1974 Directory of Federally Supported Information Analysis Centers (79).

> An information analysis center is a formally structured organizational unit specifically (but not necessarily exclusively) established for the purpose of acquiring, selecting, storing, retrieving, evaluating, analyzing, and synthesizing a body of information and/or data in a clearly defined specialized field or pertaining to a specific mission with the intent of compiling, digesting, repackaging, or otherwise organizing and presenting pertinent information and/or data in a form most authoritative, timely, and useful to a society of peers and management.

Two previous directories appeared in 1968 (20) and 1970 (21). Although Weinberg (129) identified approximately 400 IACs in 1963, the total numbers of Federally supported centers included in each directory are only 113 (1968), 119 (1970), 108 (1974). These directories are not all-inclusive; some Federally-supported IACs are specifically excluded. Among the exclusions which are of particular interest to this exploration of numeric data are those "centers devoted exclusively to the following types of information services:" (79)

- Management information services.
- Holders of raw data files.
- Mapping and charting activities.
- Regional or state information services (e.g., technological or agricultural utilization services).

Nevertheless, the lists do include some holders of data files. A reading of the 1974 list (see Figure 6.3) also indicates that there is partial overlap with National Standard Reference Data System (NSRDS) centers.

Over this period of time, 68 IACs appear in all three directories. Of those remaining, 9 were identified (by Kruzas (56)) as having terminated operation, while 23 appear to be still in existence. An additional 17 were not located in the Kruzas Encyclopedia. Some name changes had occurred, or centers had been combined, or divided. From the descriptions in the Kruzas Encyclopedia, it appears that there are a substantial number of additional centers which might have been included as Federally supported IACs.

Using the directory descriptions to classify the 1974 centers by field of science, and allowing a center to be classed in more than one field, the largest proportion is in the physical sciences (32 percent). Engineering was second with 27 percent, followed by social sciences with 23 percent. In addition, environmental sciences and life sciences also each had over 10 percent. When a center was classed as multidisciplinary if it covered more than one field of science, the multidisciplinary proportion was almost one quarter (23 percent). (See Table 6.1)

Because the definition of IACs used in the directories specifically excludes organizations which are primarily "holders of raw data files" and do not perform analysis, this population of IACs presumably includes only some unspecified portion of the organizations involved with numeric data files. However, their products are often announced and distributed by NTIS. Therefore, these centers may be the most widely known, core group of numeric data analysis organizations. Nevertheless, the 1976 "Special Technology Group Catalog"(90) from NTIS gives only one IAC that sells data in machine-readable form! The remainder of the organizations which are included in this catalog (a total of 17, of which the Office of Standard Reference Data (OSRD) is counted as one) produce paper or microfilm publications. The descriptions in the catalog do not specifically identify any of these publications or tabulations. However, many are handbooks.

Weisman (130) analyzed and summarized the characteristics of the IACs appearing in the 1970 Directory. A similar analysis was performed by KRI staff for all three directories to determine whether any trend, especially concerning the handling of numeric data, could be identified. The analysis frequently involved subjective judgements in extracting detail from the textual descriptions in the directories. Some differences appeared between our results and Weisman's. The results for the three periods are shown in Tables 6.2 and 6.3. Our results for 1970 were used in the comparisons by characteristic in order to eliminate the confounding effect of these differences in interpretation. For services and products we retained Weisman's counts for 1970, so the comparison over time may be somewhat less valid.

An increasing proportion of IACs listed in the directories appear to be providing the following services: consultations and answering inquiries,

Table 6.1 INFORMATION ANALYSIS CENTERS BY FIELD OF SCIENCE: 1974

Field of Science	Number	Percent
Total	108	100
Physical Sciences	35	32
Mathematics	2	2
Computer Sciences	3	3
Environmental Sciences	19	18
Engineering	29	27
Life Sciences	14	13
Psychology	5	5
Social Sciences	25	23
Other Sciences, NEC	2	2
Multidisciplinary	25	23

SOURCE: King Research, Inc. - (Based on Federally Supported Information Analysis Centers 1974, National Referral Center, Library of Congress.)

Table 6.2 CHARACTERISTICS OF FEDERALLY SUPPORTED INFORMATION ANALYSIS CENTERS: 1968, 1970, 1974

Characteristic		Directory Year					
		1968		1970		1974	
		#	%	#	%	#	%
Total		113	100	119	100	108	100
Type:	Government	43	38	43	36	45	42
	Academic or Educational Assoc.	37	33	38	32	36	33
	Research Laboratory	20	18	21	18	18	17
	Private Company	13	12	17	14	9	8
Age*:	<1 year	1	1	0	-	0	-
	1-5	63	56	44	37	17	16
	6-10	22	19	39	33	35	32
	>10	23	20	36	30	55	51
	Unknown	4	4	0	-	1	1
User Qualifications:							
	Clearance	31	27	16	13	17	16
	Professional	56	50	64	54	43	40
	Anyone	25	22	39	33	48	44
	Unknown	1	1	0	-	0	-
Staff:	1-2 members	20	18	17	14	16	15
	3-5	18	16	17	14	18	17
	6-10	22	19	23	19	30	28
	11-20	18	16	26	22	16	15
	>20	25	22	28	24	23	21
	Unknown	10	9	8	7	5	5
Number of Sponsors:							
	1	83	73	85	71	80	74
	2	20	18	22	18	13	12
	3	5	4	4	3	5	5
	4	1	1	3	3	7	6
	>4	2	2	3	3	3	3
	Unknown	2	2	2	2	0	-

*Age in year of directory

SOURCE: King Research, Inc.

Table 6.3 NUMBER AND PERCENT OF FEDERALLY SUPPORTED INFORMATION ANALYSIS CENTERS OFFERING SELECTED SERVICES AND PRODUCTS: 1968, 1970, 1974.

Service, Products	Directory Year 1968 No.	1968 %	1970 No.	1970 %	1974 No.	1974 %
Store, Retrieve Data	64	57	104	87	78	72
Analyze, Evaluate Data	58	51	89	75	59	55
Critical Reviews, State of the Art Reports	50	44	50	42	36	33
Critical Compilations, Handbooks	33	29	41	34	29	27
Answer Inquiries, Consultations	61	54	72	61	93	86
Reprint, Library Services & Facilities	13	12	19	16	46	43
Bibliographies, Abstracts, Referrals	46	41	62	52	70	65
Translations	2	2	1	1	0	0
Film, Maps, Graphics	7	6	10	8	11	10
Data on Tapes and Cards	8	7	15	13	27	25
Training, Workshops, Standards	6	5	1	1	14	13
R&D Conducted, Coordinated	11	10	11	9	5	5
SDI	55	49	82	69	100	93
Stimulation or Leadership in Field	4	4	5	4	3	3
Total	113	100	119	100	108	100

SOURCE:

1968, 1974 King Research, Inc.

1970, Weisman, Herman, Information Systems, Sources and Centers, 1972.

reprint library services and facilities, film and maps and graphics, data on tapes and cards, and SDI. A decreasing proportion of IACs offer critical reviews and state of the art reports, translations, and conduct or coordination of R&D. No real pattern should be presumed as the distribution may reflect the selection of IACs included in the directory rather than real trends in activity.

No noteworthy trends were found in age, type, size of staff, or number of sponsors. However, there did appear to be a reduction in the restrictions for users of IAC services. A larger proportion of IACs (44 percent) indicated in 1974 that services were available to anyone than had so indicated in 1968 (22 percent), and a smaller proportion (16 percent compared to 27 percent) required specific clearance.

Although two current NSF-funded studies are addressing the questions of IAC management with particular emphasis on financial administration, there is still no information on total levels of Federal support for IACs. Nor does there appear to be a way of determining from the literature what proportion of IAC activity deals with numeric data.

SECTION 7

INTERNATIONAL DATA COORDINATION

In this section are described the data activities of two international organizations in which the United States participates. CODATA concerns itself largely with identifying data compilation undertakings and with assuring transferability of data. The group of so-called World Data Centers concerns itself with making actual data compilations available.

The International Council of Scientific Unions (ICSU) formed two groups in the late 60's which have as their primary area of concern, the handling of numeric data. These groups are the Committee on Data for Science and Technology (CODATA), formed in 1966 and the Panel on World Data Centers (WDCs), in 1968. Activities of both of these groups are highlighted in this section with particular emphasis on their relationship to data handling in the United States. Although this section deals only with ICSU, the reader is reminded of the reference to other international organizations in Section 2.0, and should not assume that only ICSU and its members are active in the international data scene.

7.1 CODATA

One of the most important of the organizational and coordinating efforts in recent years has been the activities of CODATA. This group held its Fifth International Conference in Denver, Colorado in July 1976. Since its establishment in 1966, CODATA has been studying the problems associated with the myriad aspects of the generation, storage, and accessibility of numeric data. Initially, it was concerned almost exclusively with physical sciences data, more recently its concerns have extended to the biological and geological sciences; there currently is pressure to broaden the scope to encompass the social sciences.

Although CODATAs overall purpose is the coordination of international data programs, there is a strong underlying philosophy that coordinated national systems form the base for international exchange. Therefore in making recommendations for standards, procedures, and collaborative efforts national, regional, and local data collection efforts are frequently emphasized.

The report of a 1975 Study on the "Problems of Accessibility and Dissemination of Data for Science and Technology" (published by both UNESCO (121) and CODATA (19)) provides among other things, a review of some historical developments in the area of numeric data. Selected events from as long ago as 1883 are considered relevant.

The CODATA task group which conducted the 1975 study (Task Group on Accessibility and Dissemination of Data), produced a comprehensive framework for categorizing the many types of data projects. Summary tables of the categories are reproduced as Tables 7.1 and 7.2. Although the summary specifies only three broad scientific discipline areas, the categories can be extended to the social sciences with some imagination. Thus, for example, demographic data would fall into categories a_2, b_2, c_1, d_1 or d_2, e_1, f_1 or f_2, and g_3 of Tables 7.1 and 7.2 and production statistics into categories a_2, b_2, c_2, d_2, e_1, f_1 or f_2, and g_2. A separate example of category b_1 might be developmental stages of pre-school children; while category b_2 could be examplified by characteris[illegible] of visitors to national parks.

The categorization scheme helps to highlight the fact that the many disciplines and subdisciplines may tend to have markedly different types of data and data needs. In addition, there are the requirements, again different among themselves, of mission-oriented and cross-disciplinary data programs which are required to cope with a multitude of different categories of data in some coordinated manner. Although this categorization scheme provides a way to attach type identifiers to specific data (e.g., given a data item or element, one can list the characteristics), it is less useful for "typing" data files or data banks which may consist of diverse data elements with all of the characteristics listed. The second table provides characteristics of user-specificity which are more easily applied to data banks. However, this level of categorization alone is a rather inadequate typography for grouping data files pro forma. Another approach is suggested by the "data levels" in which bit quantity is expressed in Table 7.7 (See Section 7.2). In this breakdown, Level I applies to raw, unanalyzed data, Level II is processed data and Level III is analyzed and/or summarized data. With this approach a particular file would be expected to contain data in only one level, while a "data bank" consisting of numerous files or a "data system" could contain files in all levels.

Table 7.1 VARIETIES OF CATEGORIES OF DATA

	CATEGORIES OF DATA	CHEMISTRY/PHYSICS	GEO-/ASTRO-SCIENCES	BIOSCIENCES
a_1	Data which can be measured repeatedly	Most data	Geol. structures, rocks Accel. due to gravity Fixed stars	Most data
a_2	Data which can be measured only once		Volcanic eruptions Solar flares, novae	Rare specimens Fossils
b_1	Location-independent	Most data	Minerals Global Tectonics	Most data, excluding extraterrestrial
b_2	Location-dependent		Rocks, fossils Astronomical data Meteorological data	Rare specimens Fossils
c_1	Primary observational or experimental data	Optical spectra Crystallographic F-values	Solsmographic records Weather charts	Physiological data (e.g., respiration rates, blood volumes, etc.)
				Biochemical data (e.g., composition of tissues and organs)
c_2	Combinations of primary data with the aid of a theoretical model	Fundamental constants Crystal structures	Fossil zoning Temp. distribution in Sun	Genetic code Body surface area Model of vascular bed Dimensions of tracheo-bronchial tree
c_3	Data derived by theoretical calculation	Molecular properties, calculated by quantum mechanics	Solar eclipses predicted by celestial mechanics	Prediction of phenotypic expression from genotypes
d_1	Determinable data	Most macroscopic data	Elements of planetary orbits	Gene loci Chromosome numbers
d_2	Stochastic data	Polymer data Structure-sensitive properties	Soil and rock composition Solar flares Frequency of visible meteors per unit interval	Most data
e_1	Quantitative data	Most data	Seismic data Meteorological data	Physiological data Biochemical data
e_2	Semiquantitative data	Mohs hardness scale	Wind force scale	
e_3	Qualitative data	Chemical struc. formulae Properties of nuclides	Rock classification Classification of stellar spectra Fossil shapes	Amino acid sequences Taxonomic classification of organisms
f_1	Data presented as numerical values		Meteorological data	Physiological data Biochemical data
f_2	Data presented as graphs or models	Phase diagrams Stereoscopic molecular diagrams Molecular models	Geological maps Weather maps Sky mapping at a particular radio frequency (e.g., 21cm)	Metabolic pathways Electrocardiograms Electroencephalograms
f_3	Symbolic data		Lithology in bore hole data	

Note: A given group of data can be categorized simultaneously by several 'facets' a, b, c etc.: for instance, the nature of meteorological data characterized as a_2, b_2, c_2, d_2, e_1 and f_1 (or f_2).

SOURCE: Committee on Data for Science and Technology, ICSU, "Study on the Problems of Accessibility and Dissemination of Data for Science and Technology", CODATA Bulletin No. 16, 1975.

Table 7.2 VARIETIES OF CATEGORIES OF DATA BASED ON USER CHARACTERISTICS

CATEGORIES OF DATA	CHEMISTRY/PHYSICS	GEO-/ASTRO-SCIENCES	BIOSCIENCES
E_1 Data generated in a specific discipline and used almost exclusively by specialists in the same discipline	Para- or diamagnetic susceptibility of compounds Electric quadrupole moments of atomic nuclei Constants representing the anharmonicity of normal modes of vibration of molecules Density of frequency distribution of normal vibration of crystals High resolution infrared spectra Crystal structure factors of diffracted rays in crystal structure analysis: F(h, k, l)	Seismographic records of earthquakes Magnetospheric data obtained by artificial satellites	Electrocardiograms Electroencephalograms
E_2 Data that are also used by research workers in a limited number of related disciplines	Characteristics of ferromagnetic materials Binding energy of atomic nuclei from protons and neutrons High resolution NMR spectra Infrared and Raman spectra Optical transition probability Rate constants of chemical reactions Steam tables		Number of chromosomes in cells for biological species or strains Genetic code
E_3 Data used more widely	Fundamental physical constants Physical properties of materials Physico-chemical properties of organic and inorganic compounds Electronic structure of atoms Atomic structures of common molecules	Geological structures Tide tables Catalogue of stars brighter than 3rd magnitude Basic data on moon, planets, sun, stars, galaxies Simple lists of Fraunhofer lines, sunspot numbers, nebulae, binaries, variable stars, quasars, pulsars	Toxicity of chemicals Human visual sensitivity to colours Physical size at various stages of growth and development

SOURCE: Committee on Data for Science and Technology, ICSU, "Study on the Problems of Accessibility and Dissemination of Data for Science and Technology", CODATA Bulletin No. 16, 1975.

In a recent CODATA Bulletin (19), there is a report of a French study (9) on the typology of data banks. The study report states that in order "to cover some 60 scientific and technological fields on a world scale", about 1,500 data banks would be required. The study team estimated that there are presently between 150 and 200 data banks: 80-100 in the U.S., 50 in France, 20-30 in the USSR, 15 in Germany, the remainder in the U.K., Italy, Spain, Japan and Canada. (The 1969 CODATA directory (14) of numeric data projects listed 150 projects in 26 countries.) Thus, there are fertile areas for growth and expansion. No indication is provided in the article as to the range of scientific fields included (especially whether the social sciences are considered), nor the definition of "data banks". The above list of countries, furthermore, is not exhaustive in terms of data banks mentioned elsewhere in the literature.

7.2 The World Data Centers; International Transfer of Numeric Data in the Geophysical Sciences

The impetus toward establishing World Data Centers occurred in the early 1950's during the planning for International Geophysical Year (IGY). The difficulties experienced in obtaining data from the previous such international program (the Second International Polar Year: 1932-33), made the establishment of a systematic approach for providing world-wide access to IGY data imperative. ICSUs Comité Special pour l'Année Géophysique Internationale (CSAGI) was responsible for international planning for IGY and in 1955 authorized the establishment of "at least three IGY World Data Centers" (96), each to consist of various sub-specialty parts. Offers to conduct WDCs came from individual countries and institutions. In accepting the designation of WDC, the administering agency agreed to abide by CSAGI principles for storing and disseminating data.

Subsequently WDC-A was established in the USA, WDC-B in the USSR and WDC-C in various other nations. WDC-C subsequently has been designated WDC-C1 (Europe) and WDC-C2 (Asia).

The organization of the World Data Centers took place in the mid-1950's as a mechanism for handling the data to be collected during the International Geophysical Year of 1957-58. World Data Centers have been of such importance that they were made into permanent organizations in 1964. They serve for the collection and transfer of numeric data in the geophysical sciences and studies of the sun. They "deal in general with observations with four basic parameters: object, property, time and place of observation" (96).

Since the IGY, the World Data Centers have received, organized, disseminated, and archived the international program data generated during the International Quiet Sun Year (IQSY), the Upper Mantle Project (UMP), the program on Recent Movements of the Earth's Crust, the Global Atmospheric Research Program (GARP), the International Hydrological Decade, the Integrated Global Ocean Station System (IGOSS), the International Magnetospheric Study (IMS), and others. Since 1968 they have operated under the auspices of the International Council of Scientific Unions (ICSU), coordinated by a Panel on World Data Centers (Geophysical and Solar).

The scope and range of World Data Center operations represent only a portion of total international exchange of numeric data. Quite aside from the usual routes of publication and international professional society meetings, scientist exchanges, and correspondence among the invisible colleges there is official and quasi-official data exchange typified by weather reports for international flights and time signals from, e.g., the U.S. National Bureau of Standards and the Canadian Dominion Observatory for navigation. In addition, although emphasis appears to be broadening to other disciplines/fields (the social and life sciences especially) the World Data Centers were originally designed to deal with "Geophysical Data". Thus any account of World Data Center activity will place a disproportionate (to total real international activity) emphasis on the associated disciplines. In the areas of health and demography for example, international exchange of data has been taking place extensively for a number of years and other international organizations carry these responsibilities. The World Data Centers do appear to account for the most organized of the exchange programs in the geophysical sciences.

To a large extent, if not exclusively, the World Data Centers are supplementary "front offices" for large national data collection and/or analysis operations. That is, they rely on the same data files which are accumulated by and supported by the large national centers. Thus the file sizes of the World Data Centers (where these figures are available) are not additional to those of national systems; World Data Center activity deals primarily with the transfer of data from these files to and from other countries.

Table 7.3 lists the WDCs, the programs (discipline areas) and location. Twenty-two WDC-As are listed in the 1973 Guide (96) and 24 WDC-Bs. WDC-As and WDC-Bs include at least one center each for each program. WDC-C1s total 22 but cover only 10 of the 22 programs. There are 12 WDC-C2s covering 7 programs. The fifteen so-called "Permanent Services" affiliated with these same programs are poorly defined. For the most part they antedate IGY. Their operating procedures are not designated by the Panel on WDCs. Generally though, data are available to WDCs from the Permanent Services and vice versa.

The operating philosophy of the World Data Centers is to make geophysical data (and often bibliographic information) available to the scientific community from central collections. When there are multiple WDCs (A, B, C1 and C2) for a single program the same data are generally provided to and available from all of the WDCs.

The Guide to World Data Centers (96) describes the types of data to be accumulated by each WDC, the types of access mechanisms (distribution, catalogs, etc.), and other responsibilities of WDCs. The ICSU panel relies upon appropriate international scientific bodies for the specification and approval of any changes to the guide. The Third Consolidated Guide significantly notes, "as the annual volume of observation data increases...there is a trend toward requiring only that the WDCs be informed where data are located...and how they may be obtained (instead of requiring that the data be transmitted to the WDCs)." Although the Guide provides relatively detailed lists of the kinds of data (including freuqency and format) at each center and specifications for interactions between centers, it provides no information on data quantities. For data use it specifies some announcement and publication procedures.

Table 7.3 LOCATION OF GEOPHYSICAL WORLD DATA CENTERS AND RELATED PERMANENT SERVICES

PROGRAM	WDC-A	WDC-B	WDC-C1	WDC-C2	PERMANENT SERVICE
SOLAR-TERRESTRIAL PHYSICS					
Solar and Interplanetary Phenomena	Boulder	Crimea Kiev Moscow	Arcatri Freiburg Meudon Munich Pic-du-Midi Rome Zürich	Tokyo (Univ) Toyokawa	Boulder (IUWDS)
Ionospheric Phenomena	Boulder	Moscow	Slough	Tokyo	Boulder (IUWDS)
Flare-Associated Events	Boulder	Moscow	Meudon Ondrejov Slough Umeå	Itabashi Toyokawa	Boulder (IUWDS)
Geomagnetic Variations	Boulder	Moscow	Charlottenlund Hailsham	Bombay Kyoto Tokyo (Univ)	Boulder (IUWDS) De Bilt Göttingen Roquetas Boulder (IUWDS)
Aurora	Boulder	Moscow	Edinburgh Kiruna Umeå		
Cosmic Rays	Boulder	Moscow		Itabashi Tokyo (Univ) Mitaka	Boulder (IUWDS) St. Maur des Fossés (SPARMO)
Airglow	Boulder	Moscow	Paris		
ROCKETS AND SATELLITES	Greenbelt[1]	Moscow	Slough		
METEOROLOGY	Asheville	Moscow		Tokyo[5]	
OCEANOGRAPHY	Rockville[4]	Moscow			Birkenhead
GLACIOLOGY	Tacoma[2]	Moscow	Cambridge		Zürich
SOLID-EARTH GEOPHYSICS					
Seismology	Boulder[4]	Moscow	Strasbourg		
Tsunamis	Honolulu	Moscow			
Gravity	Boulder	Moscow			Paris
Earth Tides	Boulder	Moscow	Uccle		
Recent Movements of the Earth's Crust	Boulder	Moscow			
Rotation of the Earth	Washington[3]	Moscow			
Marine Geology and Geophysics	Rockville[4]	Moscow			Mizusawa (IPMS), Paris
Magnetic Measurements	Boulder	Moscow			
Paleomagnetism and Archeomagnetism	Boulder	Moscow			
Volcanology	Boulder	Moscow			Tokyo (IAV), Strasbourg (BCIS)
Geothermics	Boulder	Moscow			

[1] NASA

[2] U.S. Geological Survey

[3] U.S. Naval Observatory

[4] Moved to Boulder

[5] WDC for nuclear radiation data only.

NOTES: The data for programs such as the IGY, IQSY and UMP that have been completed are archived in appropriate WDCs or in an appropriate scientific institution. IGY meteorological data collected by WDC-C1 are archived in WMO, Geneva. The data for the UMP are archived in WDC-A Boulder and WDC-B Moscow. All WDC-As are administered by NOAA except for those footnoted 1-3.

SOURCE: ICSU Panel on World Data Centres. Third Consolidated Guide to International Data Exchange through the World Data Centres, 1973.

7.2.1 World Data Centers in the United States

In the United States, the National Oceanographic and Atmospheric Administration (NOAA) has the primary responsibility for the operation of the WDC-As. Only three other organizations are involved, each having responsibility for one World Data Center:

Rockets and Satellites	- National Aeronautics and Space Administration
Glaciology	- U.S. Geological Survey
Rotation of the Earth	- U.S. Naval Observatory

Each World Data Center has a separate organizational name for the associated national center. For example, WDC-A/Meteorology is colocated with the National Climatic Center.

In a report by the National Academy of Sciences in 1975 (65), these 22 WDC-As are reduced to a list of seven. The first 5 of the major programs of Table 7.3 are referred to as individual centers eliminating the sub-programs of Solar Terrestrial Physics. Solid-Earth Geophysics is divided into three centers only - Tsunamis, Rotation of the Earth (sometimes called 'Longitude and Latitude'), and Solid Earth Geophysics (which includes the remaining Solid Earth Geophysics sub-programs shown in the Table except Volcanology and Geothermics, which are not mentioned). The last two "sub-centers" are apparently in preliminary stages of organization, as no description is given in the Guide nor in the NAS reports.

Two NAS reports (65), (81) provide some information on use and file sizes. The two reports use somewhat different approaches to quantification. Neither report distinguishes between World Data Center files and those of the five so-called National Centers; nor are the actual physical locations of files clearly stated. There specifically are some discrepancies between figures presented in these reports and those given directly by NOAA and NASA. (See Section 9.)

The tabular material presented in the remainder of this section serves several purposes and requires some cautionary statements to avoid undue confusion

on the part of the reader. As might be fairly anticipated, it provides indications of the magnitude of holdings and, in some cases, of transfer of data. The quantities of data shown are presented pretty much as defined in the sources from which they came, with some aggregating and provide examples of the various media and formats in which data may be recorded. The tables serve the specific purpose of graphically illustrating the complexity (impossibility) of the task of aggregation across data formats or media.

There are some terms which remain only nebulously defined. They are used by the sources as measures of data quantity, but the terms are undefined and the concept remains only vaguely similar to a record or group of records. These terms appear not only in this section, but in subsequent sections of this report, and the reader will find no clear definitions stated. These terms are observations, stations, station-months, serial stations.

Assessment of the "Impact of World Data Centers on Geophysics" (65) summarizes two studies (1960 and 1962) of WDC-As by NAS and a 1974 survey of WDC-A users. All WDC-As together responded to approximately 12,000 requests from 3,000 requestors from 1970 through 1974. The survey queried 500 of these 3,000 users, about half of whom responded. These respondents listed 1,600 papers by 800 authors which utilized the data supplied by the WDC-As. The number of authors exceeded the number of respondents because the authors included colleagues of the proximate users. Many respondents stated that the data were used by 5-10 colleagues, nine by 20 or more colleagues. ('Use' meant generation of a paper.)

Examples of U.S. versus foreign data flow (see Table 7.4 below) for only two kinds of data indicate that the larger proportion, 80 percent or more, of the archive of solar-terrestrial physics and marine observation data came from sources outside the United States. Data requests on the other hand were predominately U.S. (72 percent) for solar-terrestrial physics data, and non-U.S. (65 percent) for marine observation data. In terms of the data that was requested, however, 37 percent of the solar-terrestrial physics data requested was U.S. data, 63 percent foreign data.

Table 7.4 presents, extracted from narrative statements, some feel for the approximate quantities of data received and distributed and for the numbers of persons receiving data. Persons receiving data included both regular subscribers to center services and products and those who made individual requests.

Table 7.4 U.S. VERSUS FOREIGN DATA FLOW FOR SELECTED WDC-As: 1974

Data Subject	Data Receipts		Data Requests	
	U.S.	Other	U.S.	Other
Solar Terrestrial Physics	20%	80%	72%	28%
Marine Observations .....:	18%	82%	35%	65%

SOURCE: National Academy of Sciences, An Assessment of the Impact of World Data Centers on Geophysics, 1975.

Measures of the quantities of some types of data received and distributed are summarized from the same report in Table 7.5. The table is incomplete (a common problem in this data collection effort) and terms, as mentioned above are undefined.

"Geophysical Data Centers: Impact of Data-Intensive Programs" (81) addresses the volume of holdings in four National Geophysical Data Centers (and associated WDC-As) and particularly the problem associated with the "exponential rise in data quantity" with which these data centers are currently coping. (It is not clear whether "exponential" is intended literally or is merely substituted for "very rapid".)

An example of the "exponential rise" is given by the National Geophysical and Solar Terrestrial Data Center (NGSDC), which estimated its tape library at 600 reels in the report (updated to 1,000 in 1976 in Table 7.6), but also anticipated an annual input of 14,000 reels due to planned programs. One of the five synchronous meteorological satellites (SMS) of the First GARP Global Experiment (FGGE) will generate 5×10^{13} bits of data over an 18-month period. If these data were to be recorded on 2400-ft reels of tape at 1600 bpi, 200,000 to 400,000 reels would be filled for this one part of the FGGE program alone.

Table 7.5 EXAMPLES OF QUANTITIES OF DATA RECEIVED AND DISTRIBUTED BY WDC-As: 1973-1974

	Received Annually	Distributed Annually	Distribution-Inquiry Rate
Glaciology			20-30/month +600 subsc./quarter
Longitude and Latitude			Other WDCs only
Oceanography	40,000 stations		250/year
Rockets and Satellites			462/yr. (1973) 42 countries
Solar-Terrestrial Physics	8,470 station months	29 x annual accum., 44% of holdings	1,730/yr. (1973) + mailings to 1,400 addresses
WDC-A (grouped)	1,600 station months of magnetograms	29,000 sta. months	12,000 requests, 3,000 requesters over 5 years (1970-1974)
	2,800 station months of ionograms & hourly values	16,600 sta. months	
	58,800 marine observation serial stations	26,000 serial stations	

SOURCE: National Academy of Sciences, An Assessment of the Impact of World Data Centers on Geophysics, 1975.

The data volumes anticipated in the near future suggest the need to purge files, not only to reduce storage problems, but also to hold processing time to reasonable lengths. For example, if the data were compressed and if processing speed were 10×10^6 bps, it would still require 75 days to process the above set of FGGE SMS data. Few potential users are likely to be able to afford such an expense. Thinning or discarding of raw data requires a number of difficult decisions, not the least of which is that even if there were universal agreement within the international scientific community that certain data could be deleted, the volumes of new data so dwarf the old data that, for any appreciable impact on file size, new data rather than old must be discarded.

Substantial pressures obviously exist to utilize new technologies for storage and processing. One such technology is video tape storage. While one FGGE SMS data set would require 200,000 to 400,000 reels of conventional magnetic tape, all 5 SMS data sets could be stored on 5,000 video tapes. The National Climate Center is already storing some data on video tapes. However, processing rates are still a problem.

Table 7.6 SUMMARY OF HOLDINGS, SELECTED WORLD DATA CENTERS AND NATIONAL DATA CENTERS

Center	Type	Medium	Quantity
WDC-A Solar Terrestrial Physics[1]	Analog data	35 mm film	11 M ft.
	Tabulated data	35 mm film	1.4 M ft.
	Digital data	Magnetic tape	400 2,400 ft. reels
	Digital data	Cards	1.6 M
	Tabulated data	Sheets or pub's	5,100 cu. ft.
WDC-A total[1]	Marine observations	-	890,000
National Climatic Center[2]	Manuscripts	Sheets	76 M
	Manuscripts	Microfilm	86,000 100 ft. reels
	Punched Cards	Microfilm	377 M images
	-	Magnetic tape	77,000 reels
	Radar film	35 & 16 mm film	19,800 100 ft. reels
	Satellite film	10"x10" negatives	175;000
	Original records & pub's	Microfiche	58,000
	Unpub'd data	Tabulations	24,300
	Back issues of climatic data	Publications	194,000
National Geophysical & Solar Terrestrial Data Center[2]	-	35 mm film	12 M ft.
	-	Magnetic tape	1,000 reels[3]
	-	Punched cards	1.6 M
	Data, prints & publications	Sheets	5,100 cu. ft.
National Space Science Data Center[2]	-	Microfilm	27,519 100 ft. reels
	-	Micrifiche	18,198
	Digital data	Magnetic tape	40,931 reels
	Photographic film	Misc. widths	1.5 M ft.
	Photographic film	Assorted sizes indiv. sheets	26,678
	Hard copy	Sheets	188,612

See footnotes at end of table (Continued)

Table 7.6 (cont.) SUMMARY OF HOLDINGS, SELECTED WORLD DATA CENTERS AND NATIONAL DATA CENTERS

Center	Type	Medium	Quantity
National Oceanographic Data Center[2]	Station data geosort	-	550,000 stations
	Other digitized geosort	-	941,000 observations
	Analog prints	-	820,000 observations
	Digital biolog. data	-	13,000 stations
	Biological data	Papers	23,000
	Surface current data	-	4 M observations
	Digital data	Magnetic tape	8,000 reels*

M = Million

* anticipated annual input of 14,000 reels (1976)

SOURCES:

[1] National Academy of Sciences, An Assessment of the Impact of World Data Centers on Geophysics, 1975.

[2] The National Research Council, Geophysical Data Centers: Impact of Data Intensive Programs, 1976.

[3] Private communication, Newton Page, EDS, NOAA. 2/76.

Although much of the data holdings of these centers is in conventional formats other than magnetic tape, computer services are reported to account for the following percentages of center budgets:

NGSDC	20%
NCC	25%
NODC	20%
NSSDC	15%

Although the report contains no dollar cost figures, there is a strong plea that researchers include in their budgets 5 to 10 percent of the total costs of research for the costs of depositing data in an accessable format in the National Data Centers. It is feared that otherwise data center budgets will be inadequate and costs to users could escalate to the point where data are ignored.

Both documents described above provide an assortment of data on data quantities. Table 7.6 displays in somewhat condensed form the holdings of WDC-As and National Data Centers. It underlines the variety of forms which holdings can take; it is quite hopeless to seek a single "total quantity of data held". Although conversion factors among some types and formats can be conjured up, the result is little more than an arbitrary number.

Table 7.7 summarizes the quantities of data for particular programs (not centers). These approximations are all expressed in bits and can be summed, across all three levels, to 336 terabits (terabits) of data. However, dates of acquisition are sometimes vague, and some quantities are annual rates rather than total data holdings. Again the total is a virtually arbitrary number -- but large!

Table 7.7 APPROXIMATE DATA QUANTITIES FOR SELECTED GEOPHYSICAL PROGRAMS

Program	Collection Period	Data Type Level I	Level II	Level III
BOMEX	1969 (3 mos.)		$1x10^{10}$	
GARP				
GATE	1974 (4 mos.)	$1.2x10^{12}$	$3x10^{11}$	
FGGE	1978 (12 mos.)	$3x10^{14}$	$6x10^{10}$	$1x10^{11}$
SOLAR-TERRESTRIAL				
IMS (part)	1977-1978	$1.1x10^{13}$	$5.5x10^{12}$	
AFGWC	on going, annual	$7x10^{11}$		
SOLID EARTH				
Marine Science Reflections	1975-1978	$34x10^{11}$		
SDAC	permanent file	$14x10^{12}$		
SRO	future, annual	$1x10^{11}$		
OCEAN				
IDOE	long-term		$3x10^{10}$	
POLYMODE	1977-1978	$1x10^{11}$	$4.1x10^{9}$	
IGOSS	1972-present		$11.1x10^{7}$	
MESAWR (one of several categories only)	1973-present		$1x10^{9}$	
NEGOA	1975-present		$12x10^{8}$	
IYGL	1972-1973		$18.6x10^{10}$	
Total by Level		$3.3x10^{14}$	$6.1x10^{12}$	$1x10^{11}$
Grand Total			$336x10^{12}$ tera-bits	

SOURCE: The National Research Council, Geophysical Data Centers: Impact of Data Intensive Programs, 1976.

SECTION 8

PHYSICAL SCIENCES

Under the heading of physical sciences this report discusses numeric data programs in astronomy, physics and chemistry. Geodesy and meteorology are classified under environmental sciences. The programs of NASA deal with numeric data in both areas, but are discussed in Section 9, Environmental Sciences.

The major source for specifically astronomical data in the United States is the Naval Observatory. The most comprehensive program for evaluating and organizing physical and chemical constants is the National Standard Reference Data System.

8.1 Astronomy Data Programs

The crucial importance of astronomy to navigation has given the Navy a long-standing mission need for astronomical data.

The U.S. Naval Observatory provides a catalog (32) of machine-readable data which it holds and of which it can provide copies on request. The Observatory is prohibited from collecting fees for copies of data it provides. It estimates that its practice of a 3 for 1 exchange (a user requesting a copy of one magnetic tape, sends 3 blank tapes to the Observatory) approximately covers the cost of copying the data. The Observatory could not provide estimates of the size of its user population nor of its costs.

U.S. Naval Observatory files contain star catalogs, ephemerides, and observational data. According to our counts, based upon the contents of Circular #146, 223 separate files are available. Of these files, 113 are available as magnetic tape only. No measure of file size is given for tape-only files. Of the 110 card files, 96 contain a total of almost 1.5 million cards. For 14 files, only annual increments were available: a total of 34 thousand cards per year. The initial year was not given for any of these 14 files. There was no indication whether the sizes of these files had been growing at a constant rate or are expected to continue doing so.

In general the field of astronomy has become involved with machine-readable data bases only since the beginning of this decade. The Naval Observatory has international agreements for exchange of data with the Royal Greenwich Observatory (England) and Astronomisches Rechen-Institut (Germany). An international data locator service was formed in Paris in 1970 by the International Information Bureau on Astronomical Ephemerides. These activities appear to be in addition to those described under World Data Centers (Section 7).

Star catalogs are also available from the Smithsonian, both as a multi-volume set and, for some observatories, on-line. The Kitt Peak Observatory which uses the Smithsonian star catalogs on-line reports an orientation to machine-readable data only in the present decade, in contrast to the Naval Observatory's involvement "since computers were first used." While maintaining that it has no real data library, Kitt Peak has two data generating programs producing substantial quantities of data (See Table 8.1). The Solar Synoptic Program records solar magnetograms and spectroheliograms filling about 200 tapes per year. These data are passed on to NASA and NOAA in pictorial form.

Table 8.1 KITT PEAK NATIONAL OBSERVATORY DATA GENERATION: 1976

Program	Tracks per Tape	Bits per Tape	No. Tapes	Annual Increase in No. of Tapes	Annual Cost ($000)	Annual Requests
Solar Synoptic Program	9	2.5×10^8	225	200	100*	49
Solar Spectral Atlases	7	10^8	3	n.a.	n.a.	16

*Includes the entire operation of the telescope. Its primary function is creation of this file.

SOURCE: Milkey, R.W., Kitt Peak National Observatory, personal communication.

8.2 National Standard Reference Data System

Numerous international bodies are in various ways involved with scientific and technical numeric data compilation and dissemination. Thus, for example, a unit of the International Union of Pure and Applied Chemistry has for decades certified "best values" for atomic weights, and the International Bureau

of Weights and Measures maintains the fundamental units of the Système Internationale des Unités. The National Standard Reference Data System (NSRDS) is similarly involved with the determination of "best values" for data in the physical sciences.

Weissman (130) refers to the National Standard Reference Data System (NSRDS), centered upon the National Bureau of Standards (NBS), as a subset of Information Analysis Centers. The data with which this system is concerned are critically evaluated quantitative numerical values of the physical and chemical properties of well-defined substances.

A 1975 status report (67) for NSRDS lists 52 component centers, 43 of which are under direct program management of the NBS Office of Standard Reference Data (OSRD). The remaining nine centers are managed by another agency or by industry or academia, but also supply evaluated data to NSRDS (Table 8.2). The complete list of centers appears in Figure 8.1.

Table 8.2 NUMBER OF DATA CENTERS & PROJECTS OF NATIONAL STANDARD REFERENCE DATA SYSTEM

Location	Direct	Indirect	Total
NBS	28	1	29
National Labs*	0	2	2
Academic	13	4	17
Industry	2	2	4
Total	43	9	52

*National Laboratories affiliated with universities.

SOURCE: Rossmassler, Stephen, Critical Evaluation of Data in the Physical Sciences-A Status Report on the National Standard Reference Data System, 1975.

The NSRDS was formally initiated in 1963. The Numerical Data Advisory Board (NDAB) of the National Academy of Sciences (NAS) provides advisory services to NSRDS. In addition OSRD is an active participant in CODATA, thus fostering international exchange of data and standards.

The modus operandi for most NSRDS centers consists of a review of the literature in a specified area, extraction of the quantitative data, and

evaluation. Evaluation involves comparisons among various sources of data and a critical appraisal of the methodology used to collect the data and even of individual experiments. Most centers as such do not create new laboratory data. However, the centers are operated by experts in the respective fields, and they are always active in these fields. When it was found that substantial amounts of data in the published literature could not be evaluated due to insufficient detail on how authors made measurements, the scientists at the centers have sometimes been inspired to make original measurements under suitable conditions.

NSRDS activities, like those of other information analysis centers, are not limited to data compilation - bibliographies and literature reviews are also provided by many NSRDS centers. The main stream of its output appears in the quarterly Journal of Physical and Chemical Reference Data first published in 1972. This journal is an official publication of both the American Institute of Physics and the American Chemical Society as well as of NBS. The journal has averaged 1,150 pages and 59 articles annually; 1,250 subscribers at the end of 1974 included about 30 percent outside of the U.S.

Additional publication outlets are the NBS-NSRD Series, and the NBS Technical note series. Over 250,000 data documents were sold in the decade 1964-1974 (133). Total output in early 1975 contained 160 compilations (Table 8.3) totaling 28,000 pages, and including quantitative data on over 30,000 materials (134).

Table 8.3 NATIONAL STANDARD REFERENCE DATA SYSTEM PUBLICATIONS: 1975

Publication Type	Cumulative Number
Journal of Physical & Chemical Reference Data Articles	62
NSRDS Series Publications[1]	62
Berkeley Particle Data Group Publications	8
Russian Translations[2]	16
NBS-Technical Notes, Monographs and Compilations[1]	21
Books & Supplements to JPCRD	16
Magnetic Tapes[2]	2

[1] Some available NTIS

[2] Available NTIS

SOURCE: Rossmassler, Stephen, Critical Evaluation of Data in the Physical Sciences-A Status Report on the National Standard Reference Data System, 1975.

SECTION 9

ENVIRONMENTAL SCIENCES

Under the heading of environmental science fall the study of environmental quality, meteorology, geology, and geodesy.

Data on environmental quality (and safety) are collected and managed by a large number of government agencies. This widespread activity derives from the pervasive impact of environmental quality on the missions of so many agencies. Conversely, several agencies' missions require numeric data on many other topics, not all of which properly fall within environmental science.

The Environmental Protection Agency is one of many government organizations which are directly concerned with environmental quality. The other organizations discussed in this section, whose mission includes some environmental sciences, are the Federal Energy Administration, the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration. Numerous other government agencies are mentioned in connection with their mapping functions.

In 1973, a U.S. Congressional subcommittee[1] "expressed concern that the Federal government was not making maximum use of environmental data collected by diverse Federal organizations" (23). Lack of coordination and duplication of effort had previously been identified in a 1971 study (for the President's Office of Science and Technology by the Committee to Study Environmental Quality Information Programs (SEQUIP)) and by an EPA-sponsored symposium in 1972. No complete directory of environmental information then existed.

The General Accounting Office (GAO) was charged with identifying Federal environmental data systems and with exploring the establishment of a network of environmental data systems. The resulting survey identified 8 major departments and 10 independent agencies that collected and/or stored environmental data and accounted for a total of 320 separate systems.

---

[1] Subcommittee on Fisheries and Wildlife Conservation and the Environment, Committee on Merchant Marine and Fisheries, U.S. House of Representatives.

The "functional areas" covered were air, water, land use, pesticides, noise, radiation, and solid waste. In addition, the purpose of the environmental data collection effort was varied and was classed by GAO as legal (including legislative and regulatory), surveillance and monitoring, research and development, management and planning, socioeconomic. (For both functional area and purpose, there were fairly large additional "other" categories.) Many systems involved more than one functional area and/or more than one purpose and not all systems included numeric data (Table 9.1).

Table 9.1 NUMBER OF FEDERAL ENVIRONMENTAL DATA SYSTEMS BY FUNCTIONAL AREA: 1974

Functional Area	All Systems		Systems With Numeric Data	
	Number	Percent	Number	Percent
TOTAL	302	100	266	100
Water	175	55	145	55
Air	120	38	92	35
Land Use	105	33	89	33
Pesticides ...	62	19	43	16
Noise	27	8	18	7
Radiation	49	15	33	12
Solid Waste ..	45	14	30	11
Other	81	25	n.a.	n.a.

SOURCE: Comptroller General of the United States, Federal Environmental Data Systems, November, 1974.

Many but not all of the systems identified were "data oriented" rather than "information oriented." Respondents to GAO's question on this characteristic indicated that over half (56 percent) were "primarily" data oriented. Additional respondents indicated both orientations or "other", which included such things as canvassing, sampling, and evaluation.

For 266 data systems which GAO could classify by geographic scope (not all of which were primarily data-oriented systems) almost half (48 percent) were national, 30 percent were regional (U.S.), and 22 percent were international.

Four Departments or independent agencies accounted for more than two-thirds of the systems identified. The Environmental Protection Agency had 27

percent of the systems. The three next largest numbers of systems were in the Departments of Interior (19 percent), Commerce (12 percent), and Agriculture (10 percent).

A number of directories were identified by the GAO* but none was complete. Some directories listed environmental data systems within a specific Department. Some agencies had directories of computer systems, which included the environmental systems. Almost half (47 percent) of the managers of environmental data systems felt the need for a comprehensive directory.

Table 9.2 COLLECTION AND TRANSFER OF ENVIRONMENTAL DATA BY FEDERAL AGENCIES: 1974

Agency	Collect Data		Provide Data To Other Agencies
	Number	Percent	
USDA	33	10	17
DOC	39	12	31
EPA	87	27	22
DOI	61	19	37
AEC	23	7	14
Corps of Engineers	10	3	8
DHEW	18	6	7
NSF	10	3	5
TVA	11	3	9
Other	28	9	17
Total	302	100	167

SOURCE: Comptroller General of the United States, Federal Environmental Data Systems, November, 1974.

Almost three-fourths (72 percent) of the systems stored at least some of their data in computer media. Other media included microfilm, photographs, maps, charts, written reports, publications, and manual files. Over 260 different computer systems (and 12 computer manufacturers) were identified. Despite the apparent problems in data compatibility, almost one-third (32 percent) were involved in network exchange of data.

* Selected individual directories were mentioned, including; an EPA directory of its own environmental information systems, "Computer Systems in EPA", the Council on Environmental Quality's "Federal Environmental Monitoring Directory."

9.1 Environmental Protection Agency

In 1976 the Environmental Protection Agency (EPA) issued a directory of environmental information systems (33). This directory defies comparison with the list of 87 EPA systems identified earlier by GAO. Systems covered in the EPA directory are ones with computer costs generally greater than $20,000 per year and may be either operational or under development. In addition, the directory, for purposes of identifying scientific and technical data, includes too broad a scope. Functions range from payroll and project management through raw data files, predictive models, and abstracting and retrieval systems. A brief summary follows and indicates some of the problems in developing comparative statistics.

Of the 45 systems listed, 14 appear to be exclusively financial management or administrative systems and therefore deemed unlikely to contain environmental data. These systems are identified in Figure 9.1 which lists all 45 systems by name. Of the remaining 31 systems, four are document-oriented (three indexing and abstracting and one library management and control). Of the 27 data-oriented systems 18 were operational, and four of these are models or program packages.

The residual 14 operational, data-oriented systems contain scientific or engineering data primarily in the areas of water, air quality, and pesticides. All of these systems are geared in some degree to the agency's regulatory authority. Some of them are multi-agency efforts. Three of the four for which user data are available, each estimate 500-800 requests serviced per year. The other, with 400 users, logs 212 requests per day.

For 10 systems which monitor data regularly the number of input sources ranges from 825 locations to 200,000 (for STORET water sample data). While STORET claims inputs at 30 million data items (200,000 locations and 150 data items) no frequency of input information is provided. SAROAD estimates a file of 50 million "raw data values". The Pesticide Registry System lists its file size as 1/2 million records with updates "equivalent to 90 thousand cards" (1 card = 1 record?). The period covered by the 90,000 cards also is not specified. The Population Studies program works with a file of 300,000 records, each

containing an average of 75 data items. While another system with (only) 825 input sources, estimates data items per source at 450.

If all the numbers of "data items" mentioned by the fourteen data systems are totaled, they yield a sum of 116 million data items. How much meaning can be attached to this total is questionable, since in this case a "data item" may be anything from a "raw data value" to a record (which, it is safe to assume, in actuality contains multiple data elements). Thus, the above total number of data items is a lower bound on the true value. In addition, the values summed do not account in any way for periodic input, and there is no indication that any data are purged from the file. Thus the total is low even further to the extent of recent input and/or update for many of the systems included in the total. Three systems account for 88 percent of the 116 million: STORET, SAROAD, and Population Studies (Table 9.3).

The entire scientific and technical information (STI) complex of EPA systems (operational and developmental) accounts for $12.6 million of EPA's total annual systems costs (for systems in the Directory) of $16.4 million. Most of this, $11.0 million is in non-bibliographic systems. STI non-bibliographic systems being developed account for $1.8 million and operational systems for $9.2 million. In Table 9.4 costs for each system (except for the administrative systems which are shown as a group) are itemized as computer, personnel and contractor costs.

Computer costs for 26 data-oriented systems (no costs were available for one of them) average 29 percent and personnel costs, 25 percent. Contractor costs fall in between the two at 29 percent. The range of total costs is from $2.2 million for STORET to $17,000 for the "Form 67" system. Average cost for 26 of the data-oriented systems is $424,000. For operational data systems the average cost rises to $509,000, while the 4 operational models/programs average $516,000 per project.

9.2 Federal Energy Administration

The Federal Energy Administration (FEA) is a regulatory agency whose purpose "is to ensure that the supply of energy available...will continue to be

Table 9.3 FILE SIZE AND REQUEST COMPARISON OF 14 OPERATIONAL EPA DATA SYSTEMS: 1976

System	Information Supplied on Requests	Input Sources	Current Estimate of File Size: Data Items Per Input Source	Current Estimate of File Size: Total Records (000)
1. CDS	None	20,000	not given	20
2. ERSS	None	5,200	~ 21	109
3. ESPS	None	~48,000	not given	48
4. Form 67	cost as little as $10	800	450	371
5. PEMS	None	14,500	not given	14
6. PARCS	500/year (up sharply)	36,000	not given	36
7. Pesticide Registry	None	not given	not given	500
8. TADS	None	900	122	105
9. STORET	212/day; 400 users/year	200,000	150	30,000
10. SAROAD	800	4,000	not given	50,000
11. ECDBS	None	not given	32	960*
12. NES	None	800	not given	4,000
13. Population Studies	None	not given	75	22,500
14. NEDS	600/year	100,000	80	8,000
Total				116,487

* Anticipated number of records is 1 million.

SOURCE: U.S. Environmental Protection Agency, Environmental Information Systems Directory, January 1976.

Table 9.4 EPA ENVIRONMENTAL DATA SYSTEMS COSTS: 1976

(Thousands of dollars)

File* and Category	Total	Computer	Personnel	Contractor
Document Files Total	1,[illegible]	447	48	1,050
APTIC	61[illegible]	180	35	400
LIMS	121	58	13	50
SWIRS	359	109	–	250
Noise	450	100	–	350
Data Files	7,127	[illegible]	2,076	1,432
CDS	276	[illegible]	21	175
ECDBS	62	40	2	20
ERSS	125	25	51	50
ESPS	156	32	24	100
Form 67	17	16	1	–
PEMS	190	101	65	24
PARCS	892	[illegible]	317	75
Pest. Reg.	636	380	6	250
TADS	53	38	15	–
STORET	2,196	1,189	720	285
SAROAD	819	589	200	30
NES	105	65	40	–
Pop. Stud.	831	282	212	337
NEDS	769	283	400	86
Models & Program Packages	2,066	866	390	810
GLWQM	520	70	120	330
IPP	60	25	35	–
LDMS	57	22	35	–
SEAS	1,429	749	200	480
Data & Models Developmental	>1,820	>530	>305	>985
AEROS	n.a.	n.a.	n.a.	n.a.
EDS	157	16	16	125
INVWAS	195	149	26	20
Fuels D.B.	52	6	24	22
SPCCS	46	19	4	23
PRMS	121	51	20	50
Lab Autom.	645	150	145	350
Pred. Model (Fresh Water)	75	25	50	–
RAPS	529	114	20	395
Administrative & Management Systems (14)	3,894	2,382	1,099	353
Grand Total	>16,[illegible]92	>7,844	>3,918	<4,630

*For full names see Figure 9.1.

SOURCE: U.S. Environmental Protection Agency, Environmental Information Systems Directory, January 1976.

sufficient..." (126). It makes policy concerning availability, distribution, allocation, and utilization of energy and fuel; it is involved in economic forecasting and policy, foreign policy, resource development, conservation, and environmental impact assessment.

In the FEA Directory of Federal Energy Data Sources (35) 209 sources are listed and distributed among 14 subject fields (Table 9.5). Of the sources 26 percent are designated as data files. Some sources in the directory are major information centers ("Air Pollution Technical Information Center System"), others are single data files ("Other Continental Shelf Statistics"). Included as "energy data sources" are such series as BLS price indices and 1970 Census summary tapes. The Directory is subtitled, "Computer Products and Recurring Publications." For the recurring publications which account for more than half of the entries, it is not clear whether or not they are available in machine-readable as well as published formats.

9.3 National Oceanic and Atmospheric Administration Data Centers

Data base activities of the National Oceanic and Atmospheric Administration (NOAA) are concentrated in the Environmental Data Service (EDS). Five major facilities and a searching service account for the bulk of its activities. The following three national data centers: National Oceanographic Data Center (NODC), National Climatic Center (NCC), and National Geophysical & Solar Terrestrial Data Center (NGSDC) are supported by the functions of the Center for Experiment Design and Analysis (CEDDA) and the Environmental Science Information Center (ESIC).

CEDDA was initiated in 1969 to reduce, process, and validate Barbados Oceanographic and Meteorological Experiment (BOMEX) data. These data management and analysis activities have since been extended to the International Field Year of the Great Lakes (IFYGL) and the Global Atmospheric Research Project (GATE)*, for which CEDDA was also involved in provision of data acquisition systems prior to the beginning of field operations. CEDDA is responsible for placing the processed data in a permanent archive. For the three projects mentioned, the data are archived in the NCC.

*For quantitives of data resulting from BOMEX and GATE see Section 7.2 on WDCs.

Table 9.5 FEDERAL ENERGY ADMINISTRATION DATA SOURCES

Subject Field	Recurring Publications	Data File	Model, Simulations	Data Base Reference Services	Software	Total
Energy Conversion	1					1
Energy Cost, Economics, Financing	15	6				21
Energy Policy, Regulation	1		1			2
Energy Sources, Reserves	1	3		3		7
Energy Use, Supply & Demand	17	3	1			21
Electric Power, Heat Generation	11				1	12
Electric Power Distribution					9	9
Environmental Aspects	3	3	7	9	3	25
Extraction, Production	14	2				16
Fuel Properties	6	3				9
Fuel Transportion, Storage	10					10
Industrial Surveys (Suppliers)	12	12				24
Industrial Surveys (Users)	12	2				14
General*	15	19		1		35
Total	118	53	9	13	13	206*
% of Total	57	26	4	6	6	100

*Three additional entries in this category were directories and are not included in the tabulations.

SOURCE: Federal Energy Aministration, Directory of Federal Energy Data Sources, 1976.

ESIC acts as technical publisher and librarian for EDS. It deals with the organization of the literature and operates OASIS, the literature search system. A second search system which provides an index of data files is called ENDEX. In May 1974, EDS reported the ENDEX file size as 2,000 entries (132). In 1975, ENDEX contained information about almost 3,500 environmental data files (74). The files include those of NOAA activities outside EDS and of a number of Federal agencies. In August 1976, ENDEX described 5,500 data files (131). ESIC has set a goal of a comprehensive inventory of environmental data files by 1980.

Two of the three national data centers are located in Boulder (NCC is in Ashville, N.C.). The NGSDC consists of two major subcenters: Solar-Terrestrial Physics and Solid Earth Geophysics. These are further subdivided into a number of subcenters of their own. Each of the three national data centers is associated with a World Data Center (see Section 7.2).

The description in Figure 9.2 of the activities and responsibilities is excerpted from the Federal Register, April 16, 1974 (36). Additional data sets have been added since that time. The National Climatic Center is described as the largest climatic center in the world; the National Oceanographic Data Center is stated to house the world's largest usable collection of marine data.

Figures were obtained indicative of the size, activity, and growth of the EDS National Data Centers. Table 9.6 presents the numbers of user requests answered by the four centers (including ESIC), their origin, and the year-by-year changes from 1972 through 1975. The data are perturbed by a backlog of requests from within NOAA which was worked down during this period. Despite this fact, there was an overall increase of 43 percent in the annual number of requests answered during this period. The sharpest increase has been in requests from the general public, which almost tripled. Requests from academia more than doubled from 1972 to 1973, but declined 14 percent during the next two years.

Table 9.7 presents data for an 11-year period on the holdings of the National Geophysical and Solar Terrestrial Data Center and on the services rendered. It is immediately obvious that the figures for the numbers of requests

Table 9.6 ENVIRONMENTAL DATA SERVICE USER REQUEST STATISTICS: 1972-1975

	Number of User Requests[1] (000)								Percent Change			
	1972		1973		1974		1975		72-73	73-74	74-75	72-75
	No.	% of Total	No.	% of Total	No.	% of Total	No.	% of Total				
Center												
ESIC**, [1]	19.7	36	21.0	33	21.7	29	17.8	23	7	3	-18	-9
NCC	29.1	53	36.8	58	46.9	63	52.7	67	26	27	12	81
NGSDC	4.5	8	4.4	7	4.0	5	5.5	7	-3	-9	38	21
NODC	1.7	3	1.8	3	1.9	3	2.7	3	8	3	46	62
Category												
Foreign	2.5	5	3.1	5	3.7	5	4.0	5	22	19	9	59
Academic	5.9	11	12.1	19	11.0	15	10.4	13	106	-9	-5	77
Industry	16.0	29	13.6	21	19.0	26	20.1	26	-15	40	6	25
NOAA[1]	14.5	26	9.5	15	4.6	6	5.6	7	-35	-52	22	-62
Other Government	6.3	11	6.9	11	9.5	13	9.9	13	10	36	5	58
General Public	9.8	18	18.8	29	26.7	36	28.7	36	93	42	7	194
Total	55.0	100	64.0	100	74.4	100	78.8	100	16	16	6	43

[1] Approximately 30 percent were telephone requests in 1974.

[2] Does not include library and editing services; approximately 32,000 in 1974; increased by 38 percent in 1975.

[3] For the first three years shown, ESIC was answering a backlog of NOAA requests, resulting in abnormal decreases showing in data for both groups.

SOURCE: Telephone - Lewis Pitt, Special Projects, EDS, NOAA (11/76, unpublished data).

Table 9.7 SUMMARY OF RESPONSES TO KRI QUESTIONNAIRE: NATIONAL GEOPHYSICAL & SOLAR TERRESTRIAL DATA CENTER SOLID EARTH DATA

Fiscal Year	Holdings: Mag. Tapes[1] Number	Holdings: Mag. Tapes[1] Annual increase (%)	Holdings: Film Images-Seismograms Number (000,000)	Holdings: Film Images-Seismograms Annual increase (%)	Requests Filled[2] Total (000)	Requests Filled[2] U.S. (000)	Requests Filled[2] Foreign	No. Internal Users of Tapes	Sales of Services & Files[3] ($000)
1965 ..	50	n.a.	.5	n.a.	1.2	.90	300	10	100
1966 ..	70	40	.8	60	1.4	1.05	350	13	115
1967 ..	100	43	1.1	38	1.6	1.20	400	16	130
1968 ..	130	30	1.4	27	1.8	1.35	450	19	145
1969 ..	170	31	1.7	21	2.0	1.50	500	22	160
1970 ..	220	29	2.0	18	2.2	1.65	550	25	175
1971 ..	250	14	2.3	15	2.4	1.80	600	28	190
1972 ..	290	16	2.7	17	2.6	1.95	650	31	205
1973 ..	350	21	3.2	19	3.0	2.25	750	34	220
1974 ..	500	43	3.6	12	3.4	2.55	850	37	235
1975 ..	4,000	700	4.0	11	4.0	3.00	1,000	40	250

NOTE: Publications (data compilations and catalogs); 4,050 copies printed annually, all years.

[1] 70% Physical Science, 10% Computer Science, 10% Environment Science, 10% Engineering (no density given)

[2] 45% individual company; 45% institutions, 10% distribution organizations

[3] 80% U.S., 20% foreign

SOURCE: King Research, Inc.

filled in the years 1972-75 are substantially lower than presented in Table 9.6 for this data center. The figures furnished by NGSDC here cover only requests for Solid Earth Data. The growth in the number of requests filled is regular, and the proportion of foreign requests constant, as only 1965 and 1975 data were provided with the comment that linear growth occurred. The growth in the number of magnetic tapes shows no particular pattern, but does show a very sharp jump for the last year reported on. The seismogram film holdings show fairly steady arithmetic growth, 300,000 per year during the first six years and 400,000-500,000 thereafter.

Table 9.8 lists the size of magnetic tape holdings of five NOAA data centers both within and without EDS. (Tape holdings for NGSDC are only one-fourth of that shown in Table 9.7 for the preceeding year. It is not clear whether this is due to compression of data or a failure to provide up-to-date information.) Table 9.9 compares budget obligations for three periods for the National Meteorological Center, in the U.S. Weather Service, and the National Climatic Center, in EDS. Finally Table 9.10 presents the sizes of

Table 9.8 SELECTED TAPE VOLUMES IN EDS, 1976

NOAA Center	Number Tape Reels	Anticipated Rates of Increase
NGSDC	1,000	1,200/mo.: includes new seismic data inputs
NCC	75,000	*
NODC	8,000	-
Satellite Data Center	6,000	FY'78 plan calls for equivalent of 200,000/yr.
FGGE	11,000	-

*The number of automatic weather observations per station will increase from 3/day currently to 24/daty in 1978.

SOURCE: Newton Page, EDS, personal communication, March, 1976.

Table 9.9 BUDGET OBLIGATIONS NATIONAL METEOROLOGICAL CENTER AND NATIONAL CLIMATIC CENTER: 1972, 1974, 1976

	FY 1972 (000)	% of Total	FY 1974 (000)	% of Total	2 year increase %	FY 1976 (000)	% of Total	2 year increase %	4 year increase %
NMC	$6,529	64	$8,704	65	33	$9,700	60	11	49
NCC	3,711	36	4,657	35	25	6,520	40	40	76
Totals .	10,240	100	13,361	100	30	16,220	100	21	61

SOURCE: Ms. Ceasar, EDS, personal communication, December, 1976.

Table 9.10 PRESENT HOLDINGS OF SELECTED WDC-A AFFILIATED NATIONAL DATA CENTERS: 1975

Format	Medium	Center NCC	NGSDC	NSSDC	NODC
Manuscript	sheets	76 $x10^6$			
Microfilm of ms. .	100 ft. reels	86 $x10^6$		27,519	
Microfilm	cards	377 $x10^6$			
Magnetic tape	2400 ft. reels	77 $x10^3$	600	40,931	60
Radar film	100 ft. reels	19.8$x10^3$			
Satellite film ...	100 ft. reels	7 $x10^3$			
Negatives & film	sheets	175 $x10^3$		26,678	
Microfiche	sheets	58 $x10^3$		18,198	
Unpub. data tabs .	-	24.3$x10^3$	5.1$x10^3$ cu.ft.	188,612	
Issues data pubs.	-	194 $x10^3$			22,917
Punched cards	cards		1.6$x10^6$		
Linear film ..... (16mm to 9.5")	lin. ft.		12$x10^6$	1,452,217	
Observations	-				5.9$x10^6$
Stations	-				562,793

SOURCE: The National Research Council. Geophysical Data Centers Impact of Data Intensive Programs, 1976.

various types of holdings of the EDS centers. This table is a reformatting of Table 7.6 and serves primarily to illustrate the inadequacy of attempting to organize the data in such a way that holdings might be aggregated across centers or systems.

9.4 The National Space Science Data Center - NASA

Among the National Aeronatics and Space Administration (NASA) statutory directives is "to provide for the widest...dissemination of information concerning NASA's activities and their results" (127). The two major information activities of NASA are the Scientific and Technical Information Office which concentrates on technical literature, and the National Space Science Data Center (NSSDC), which, the name indicates, concentrates on data. (NSSDC also provides the facilities for the WDC-A [Rockets and Satellites], mentioned in Section 7.2.)

Interestingly, the Kruzas Encyclopedia (56), while listing the center, does not index it as a data collection and analysis center although this index contains the majority of agencies listed which hold numeric files. A site visit gave the impression that most of the center operations revolve around the dissemination and distribution of copies of holdings. However, NASA public information pamphlets emphasize an active role by NSSDC for the collection of data from space science investigations. A sampling of types of NSSDC documents was acquired during the site visit and included the following:

- The Orbiting Geophysical Observatories, OGO Program Summary (47), a catalog of experiments (which took place in the period 1964-1971) with a bibliography and abstracts of the associated literature for each experiment. It is not apparent that references include data tabulations.
- Interplanetary Magnetic Field Data Book (54), listings of 1963-1974 data, reduced (four to a page) computer listings and graphic reproductions (about 1" thick).
- Catalog of Particles and Fields Data 1966-1977 (55), a catalog of data sets available on magnetic tape, microfiche, microfilm reel or hard bound. OGO data (see above) are covered.

A number of data announcement publications are also produced:

Data Catalog of Satellite Experiments, (lists the entire collection at NSSDC)

Data Announcement Bulletin (updates to the Catalog above)

NSSDC Handbook of Correlative Data

Data Users' Note

NASA personnel estimate an increase in total data holdings of about 15 percent per year. Volumes of data held (1973 to 1975), by recording media, (Table 9.11) indicate that largest increases occur in microfiche and magnetic tape. The number of paper documents is evidently kept down by microfilming and purging. Only one format of photographic film was being accumulated in significant quantity by the end of the period reported on.

The National Space Science Data Center is one of the national centers listed in the 1974 Directory of Federally Supported Information Analysis Centers (79). The holdings as described in that document are 12,000 magnetic tapes, 2.4 million feet of microfilm and 1.3 million feet of photographic film, substantially less with respect to magnetic tapes and microfilm that the quantity of holdings tabulated as "year-end statistics" (99) by NASA for December 1973 or 1974. Photographic film holdings are given by both sources as about the same.

The 1974 Year-End Statistics (99), partially because it consists of 50 pages of tables with no accompanying text, appears to provide a great deal of conflicting information. In 13 tables dealing with "Request Statistics", 1974 total requests are variously given as 2,514; 1,060; 2,181; 1,709; 2,569; 2,995; 2,511. A footnote to the first table (with "total completed requests" at 2,514) offers the helpful advice that "the total number of individual requests during CY 1974 is 2,181." Using this as the total number of requests, it is revealed that 34 percent of requests are from outside the United States. The earlier statement that WDC-A (Rockets and Satellites) handles foreign requests and NSSDC, U.S. ones doesn't seem to fit this statistic. Requested data concerned over 2,000 spacecraft, 28,000 experiments, 2,000 data sets, and

Table 9.11 DATA HOLDINGS OF THE NATIONAL SPACE SCIENCE DATA CENTER: 1973-1975

Form	Volume			Change		Percent Change		
	1973[4]	1974[4]	1975[5]	73-74[5]	74-75[6]	73-74[5]	74-75[6]	73-75[6]
Sheets & bound volumes	222,974[1]	166,988[1]	188,612	55,986	21,624	-25.1	12.8	-1.5
Digital magnetic tapes, 1/2 in. x 2400 ft.	27,463	33,739	40,931[3]	6,276	7,192	22.9	21.3	48.0
Microfilm, 100-ft. reels	24,254	26,214[2]	27,519	1,960[2]	1,305	8.1	5.0	13.5
Photographic films:								
5-in. width, linear ft. ...	45,800	45,800	45,800	0	0	0.0	0.0	0.0
9-1/2-in. width, linear ft.	11,550	17,500	17,500	5,950	0	51.5	0.0	51.5
70-mm width, linear ft. ...	384,766	477,087	509,663	92,321	32,576	24.0	6.8	32.5
16-mm width, linear ft. ...	119,276	119,276	119,276	0	0	0.0	0.0	0.0
35-mm width, linear ft. ...	759,974	759,975	759,976	1	1	-	-	-
4 x 5 in., each	12,140	12,193	12,291	53	2	-	-	-
8 x 10 in., each	6,193	6,233	6,289	40	56	-	-	-
16 x 20 in., each	93	93	93	0	0	0.0	0.0	0.0
20 x 24 in., each	8,005	8,005	8,005	0	0	0.0	0.0	0.0
Microfiche	4,759	10,226	18,198	5,467	7,972	114.9	78.0	282.4

[1]Hardcopy is less due to filming of data and purging of hardcopy data.

[2]These amounts will be less when reels that were received, then spliced, are subtracted.

[3]Unspecified dimension.

NOTE: "-" indicates less than .05.

SOURCES:

[4]PMI Facilities Management Corp., National Space Science Data Center 1974 Year End Statistics, 1975.

[5]The National Research Council, Geophysical Data Centers: Impact of Data Intensive Programs, 1976.

[6]King Research, Inc.

7,000 sounding rocket launches. Table 9.12 lists the media used to satisfy some (1,709) of these requests and the quantities of each medium supplied.

Table 9.12 NSSDC REQUEST OUTPUT: 1974

Medium	Completed	Quantity	Output Unit
Digital magnetic tapes	94	1,253	2,400-ft. tapes
Punched cards	70	94,661	Cards
Computer printout	234	46,485	Pages
Microfilm (total)	235	1,829	100-ft. reels
Hardcopy	379	24,113	Pages
Microfiche (total)	360	18,669	Each
Photographic prints	217	8,146	Each
Contact prints	20	2,004	Feet
Film duplicates	75	2,061	Each
Film duplicates	25	12,390	Feet

SOURCE: PMI Facilities Management Corp., National Space Science Data Center 1974 Year End Statistics, 1975.

In 1974, NSSDC responded to over 1,000 requests for satellite data alone, as shown in Table 9.13; the 7-year total was almost 8,000 requests.

9.5 Cartographic and Geographic Data Programs

The data in this area are largely obtained from two reports (46, 92), augmented by personnel interviews. The principal activities covered tend to exclude those associated with the social sciences (the Census Bureau, for example). They deal with the broad area of land surveys and natural resources.

In 1975 a report was prepared for the U.S. Geological Survey (USGS) on digital geographic data handling programs (46) within the USGS. This report identified 54 activities and projects (see Figure 9.3) currently involving an estimated store of 366 gigabits of data. The additional quantity of data to be entered by 1980 was 999 gigabits, yielding a total data volume of 1.37 terabits in 1980. This represents an increase over the five-year period of over 270 percent.

Table 9.13 NUMBER OF REQUESTS FOR SATELLITE DATA SATISFIED BY NSSDC BY DATA SET SCIENTIFIC SUBDISCIPLINES

Subdiscipline	Requests 1968-1974	1968-1974 Average per Year	Requests 1974
Lunar science data			
Photography	3,686	527	446
Other lunar science	99	14	72
Planetary science data (exclusive of earth or moon)			
Atmospheres	4	1	0
Photography (surface)	867	124	83
Celestial mechanics	2	0	0
Earth and ocean physics data			
Geodetic observations	276	39	2
Other earth and ocean physics	203	29	10
Meteorology data			
IR photofacsimiles	488	70	32
IR digital	294	42	26
Other meteorology	4	1	0
Physics and astronomy data			
Magnetic fields	688	98	155
VLF and plasma waves	58	8	10
Energetic plasma (solar wind and plasma sheath)	399	57	69
Cosmic rays (solar and galactic)	247	35	53
Energetic magnetospheric particles (trapped and auroral)	139	20	24
Ionospheric sounder	229	33	17
Solar physics	166	24	15
Astronomy (RF, UV, and visible)	44	6	17
Astronomy (X ray and gamma ray)	17	2	10
Thermal atoms and ions	17	2	7
Thermal electron content	32	5	11
Other physics and astronomy	9	1	1
Ephemeris data	9	1	0
Other data	2	0	0
Total	7,979	1,040	1,060

SOURCE: PMI Facilities Management Corp., National Space Science Data Center 1974 Year End Statistics, 1975.

Although many of the projects identified in this report appear from the descriptions to be available only to USGS staff (i.e., are not directly offered to the general scientific and technical public) the products of these systems (maps, tables, charts, or listings), usually are publicly available. Operations of the USGS which involve large numeric data systems and which are available for other than staff use include:

Earth Resources Observation System (EROS)
Geographic Information System (GIS)
National Cartographic Information Center (NCIC)
National Water Data Exchange (NAWDEX)
National Water Data Storage and Retrieval System (WATSTORE)
Outer Continental Shelf Geological & Geophysical Open File
Resources & Land Investigation Program (RALI)

With the exception of WATSTORE and NAWDEX, the data systems described above by NSF (85) are not separately identified in the digital data handling document. Rather, selective activities or projects within these "systems" are described. The NSF Report provides these file size descriptions in 1975 for three of the USGS systems:

EROS - archives included:
740,000 items of ERTS & Skylab imagery
1.5 million items of NASA aircraft imagery
3.5 million items of conventional holdings

NAWDEX - an index to sources of water data from 19 Federal agencies and more than 300 non-Federal organizations

STORET - data from 130,000 sites
350,000 peak-flow observations
340,000 station years of daily streamflow values, levels, & quality observations
850,000 chemical analyses
annual input from 10,000 stations
1,300 lakes & reservoirs
4,300 water quality stations
4,100 temperature measurement sites
880 sediment stations
2,500 key wells
1,500 water resources investigations.

Some additional information was obtained* on WATSTORE, in part for distinguishing this system from EPA's STORET Water Quality System. WATSTORE includes three files:

Daily Values
Water Quality
Ground Water

Daily Values on stream flow are input from 10,000 stream stations. Each record consists of one year's accumulation of daily values for one station. Both Daily Values and Water Quality file data generally are passed to EPA's STORET as well as being maintained in the WATSTORE data base. The Ground Water data are evidently entered only into WATSTORE.

Although contributors to and users of WATSTORE were originally all within the Water Resources Division of USGS, by 1975 both populations had expanded to include other USGS Divisions and other Federal agencies. By 1976 non-Federal government agencies had been added to the list of users and contributors. One of the largest current users is the Army Corps of Engineers.

Total file size is currently (1976) almost 1.4 million records, with record lengths differing (Table 9.14). It is presently increasing at a rate of 6 1/2 percent per year. However, as long as data continue to be collected in the same manner from essentially the same number of loci, the files will grow arithmetically rather than by annual compounding.

The annual increases in file size shown in Table 9.14 were not included in the USGS Digital Geographic Data Handling report (46), although the current file sizes (for 1976) are in line with those stated by NSF for 1975. Descriptions of other systems indicate that increases in other expanding files may also not be included in the projections for USGS data handling in 1980.

Dollar expenditures are elusive because digital data programs tend to be part of the operating budgets of USGS divisions and "systems", "files", etc., are not separately identified as line items in agency budgets.

---

*Charles Showen, USGS, Water Resources Division, personal communication Dec. 1976.

Table 9.14 VOLUME OF DATA IN WATSTORE 1976, 1980

(Thousands of records)

WATSTORE Files	1976 Current Volume	Annual Increase	1975-76 Annual Rate of Increase %	1980 Estimated Volume
Daily Values	380	10	2.6	420
Water Quality	1,000	80	8.0	1,320
Ground Water	(.4)	(.055)	13.8	(.62)
Total	1,380	90	6.5	1,740

SOURCE: Charles Showen, USGS, Water Resources Branch, personal communication, December 1976.

The USGS operates the National Cartographic Information Center (NCIC). The Digital Geographic Data Handling report lists four separate programs within NCIC. The NCIC Newsletter (73) provided the following data on one of them, the Aerial Photography Summary Record System (see Table 9.15). The relationship between photography frames and records is not clear. From this information one cannot tell in what fashion, e.g., 629 records represents 35 percent of the 700,000 USDA/SCS frames.

NCIC provides one of the few instances in which states or private organizations are cited in such a tabulation. In addition to the organizations included in the table, NCIC has agreements (signed or under negotiation) with the following: NOAA, HUD, Army Corps of Engineers, EPA, and the National Archives and Records Service. NCIC has an even lengthier list of agencies that it considers to be prospective contributors to center holdings.

In 1973 the Office of Management and Budget (OMB) published a report on Federal Mapping, Charting, Geodesey and Surveying (92). The data are older than those in the USGS report, and overlap of activities is almost impossible to determine. The report does, however, address expenditure levels. These are listed in Table 9.16 separately for mapping, charting, and geodesy on the one hand and for supporting surveys and investigations on the other. Expenditures are shown for entire Departments and for major units within them. The latter values do not necessarily account for the entire Departmental totals.

Table 9.15 NATIONAL CARTOGRAPHIC INFORMATION CENTER: AERIAL PHOTOGRAPHY SUMMARY RECORD SYSTEM DATA BASE: JULY 1976

Agency	Est. total holdings (no. frames)	Approx. % holdings in APSRS	(No. of records) Status of photos* 1	2	3	Records (total)	Remarks
USDA							
ASCS	5,000,000	40%	33	49	5645	5727	Input by NCIC & ASCS
FS	800,000	45%	136	71	2918	3125	Input by FS> 1:40,000
SCS	700,000	35%	---	--	629	629	Input by SCS >1:40,000
DOD							
DMATC	35,000	--	---	2	224	226	Input by NCIC & Topo-Com
DIA	12,000,000	--	---	---	---	---	Fall '76 input
DOI							
USGS	5,000,000	90%	643	1208	34198	36049	EDC transfer of data base software
BLM	300,000	--	---	---	333	333	completed.
BOR	80,000	--	---	---	139	139	
NASA							
AMES	>1,000,000	95%	---	---	16750	16750	EDC transfer of data
JSC	> "	"	---	---	50035	50035	base software
MARSHALL	4,100	?	---	---	533	533	completed.
DOC							
NOS	550,000	10%	---	---	1690	1690	Latest date coverage of U.S. coastal areas.
TVA	175,000	--	---	---	---	---	Fall '76 input.
STATES							
Texas	?	--	---	---	---	---	Fall '76 input
PRIVATE							
MARK HURD	?	?	0	0	200	200	1:80,000 quadcentered 1976
KEYSTONE	?	?	0	0	24	24	1:80,000 quadcentered only of N.E. U.S.
TOTAL	25,644,100	--	812	1130	113,565	115,70	

*Symbols 1-planned,
2-in progress,
3-complete

NOTE: Abbreviations are as follows:
ASCS = Agricultural Stabilization & Conservation Service
FS = Forest Service
SCS = Soil Conservation Service
NOS = National Oceanographic Service
DMATC= Defense Mapping and Topological Command
DIA = Defense Intelligence Agency
USGS = Geological Survey
BLM = Bureau of Land Management
BOR = Bureau of Outdoor Recreation
EDC = Environmental Data Center

SOURCE: National Cartographic Information Center, Newsletter No. 5, Summer/Fall 1976.

Table 9.16 EXPENDITURES FOR MAPPING, CHARTING, GEODESY BY SELECTED DEPARTMENTS AND MAJOR CONTRIBUTING COMPONENTS: 1972

(Millions of dollars)

Department, Agency	Mapping, Charting & Geodesy		Related Activities*	
	Agency	Dep't Total	Agency	Dep't Total
DOD total		90.7		40.8
Navy Operations	29.1		8.5	
Office of Naval research	21.7		32.3	
Corp of Engineers, Army	22.9		-	
Defense Mapping agency	12.4		-	
DOI total		76.2		14.9
Geological Survey:				
Topographic Division	37.5		-	
Geologic Division	4.4		11.6	
Bureau of Land Management	11.2		-	
USDA total		20.5		23.6
Soil Conservation Service	6.5		21.4	
Forest Service	11.3		2.2	
DOC total		46.5		24.7
National Ocean Survey, NOAA	41.2		-	
National Marine Fisheries, NOAA			21.3	
DOT total		17.1		7.2
Federal Highway Admin.	11.0		-	
HUD total		17.2		- -
Independent Agencies total		36.7		30.7
NSF	20.3		29.8	
Federal total		304.8		142.0

*Surveys and investigations

SOURCE: Office of Management and Budget, Report of the Federal Mapping Task Force on Mapping, Charting, Geodesy and Surveying, July 1973.

Thirty-nine separate agencies expended 13,000 man-years in these efforts during FY 1972.

Table 9.17 shows the distribution of expenditures for mapping, charting, and geodesy according to type of activity. Land surveys and mapping accounted for 56 percent of the $304.8 million in expenditures; marine mapping, charting, and surveying account for 38 percent. The remainder was spent on aeronautical charting.

Table 9.17 FEDERAL EXPENDITURES FOR MAPPING, CHARTING, AND GEODETIC ACTIVITIES BY PURPOSE: FY 1972

(Millions of dollars)

Purpose	Expenditures			
	Individual	Subtotal	Total	Percent
Land Surveys and Mapping			171.3	56
Surveys		74.9		
Geodetic	9.1			
Earth Physics	7.3			
Geophysical	3.3			
Mapping Control	10.7			
Codastral	14.1			
Construction & Facility	30.3			
Mapping		96.4		
National Topographic Maps	36.3			
Special Base Maps	21.5			
Thematic & Other Maps	38.5			
Aeronautical Charting			16.5	5
Marine Mapping, Charting & Surveying			117.0	38
Systematic Mapping & Charting		50.4		
Nautical Charting	22.6			
Bathymetric Mapping	16.0			
Geophysical Mapping	11.8			
Scientific & Engineering Surveys		66.6		
Hydrographic	7.5			
Bathymetric	11.9			
Geophysical	47.2			
TOTAL			304.8	100

SOURCE: Office of Management and Budget, Report of the Federal Mapping Task Force on Mapping, Charting, Geodesy and Surveying, July 1973.

A number of other breakdowns are also provided by OMB. Table 9.18 shows expenditures for selected technical activities. The two largest expenditures were for cartography and for photogrametric processing, which accounted

for 12.1 percent and 5.8 percent respectively of the $304.8 million total for mapping, charting, and geodesy. Printing and distribution of some 50 million copies of products accounted for almost $14 million (4.5 percent).

Table 9.18 FEDERAL EXPENDITURES FOR SELECTED MAPPING, CHARTING, AND GEODESY TECHNICAL SERVICES: 1972

(Millions of dollars)

Service	No. of Agencies Having Service	Expenditures
Aerial Photography	Unspecified	$ 6.5
Photogrametric Processing	14	18.0
Cartography	18	37.5
Printing	18	9.5
Distribution	11	4.3
Data & Information Systems	19	9.6
Total		$145.5

SOURCE: Office of Management and Budget, Report of the Federal Mapping Task Force on Mapping, Charting, Geodesy and Surveying, July 1973.

Some examples of the relationship between data acquisition and data processing (reduction & analysis) are provided in Table 9.19 for three selected ocean data survey activities. Quite understandably, the acquisition costs, which include ships and equipment, account for the largest proportions, from 58 percent to 87 percent.

Table 9.19 DATA ACQUISITION AND DATA PROCESSING, SELECTED OCEAN DATA

(Thousands of dollars)

Activity	Data Acquisition Expenditure	Data Acquisition Percent of Total	Data Processing, Reduction	Total
Platform Related Surveying Activity	$60.1	58	$43.8	$103.9
Marine Geophysical/Geological Surveys and Mapping	66.9	78	18.8	85.7
Hydrographic Surveys	18.0	87	2.8	20.8

SOURCE: Office of Management and Budget, Report of the Federal Mapping Task Force on Mapping, Charting, Geodesy and Surveying, July 1973.

SECTION 10

LIFE SCIENCES

10.1 U. S. Department of Agriculture

A 1973 report (30) by a U. S. Department of Agriculture (USDA) task force on ADP/Research, in one of many references to numeric data files, describes a situation common to a number of agencies which have extensive staffs of researchers, but have not yet created an organized numeric data file function:

> Some deep rooted and sensitive feelings have developed over the years among agricultural scientists about the handling and distribution of data. In general, scientists have not thought of data and information in terms of what is frequently referred to as data and information bases-banks, etc. In general, they have not thought of these repositories as "open"data sources.

It is of course specifically "open" data sources that we are concerned with in this report. The Department of Agriculture has, in recent years, been attempting to identify those data files which lend themselves to being or could be converted to "open" data files.

Between 1971 and 1974, the USDA Office of Information Systems (OIS) oversaw the production of a series of data inventories. Six volumes were produced, each covering one or two of the agency's 10 missions. The seventh volume, a combined subject index, was never completed. No information was uncovered on the number of data files listed in the series, although the largest, Volume 6, (29) the only one examined, contains approximately 800 entries. If the smaller volumes were judged to average something on the order of 200 each, the total number of files inventoried, would be in the 1,500 to 2,000 range. However, less than 10 percent (judged from Volume 6 alone), are automated files, or were at the time of publication. Current estimates from USDA* places the number of automated "open" data bases at 41, of which "10-15" are bibliographic.

*Ronald DeClark, personal communication, 1976.

A member of the King Research staff recently (March, 1977) had need of data regarding the poultry industry. The aforementioned volume was in our library as a result of research for this report. It listed 5 files which appeared from the descriptions (see example of entries in Figure 10.1) to be possible sources of data. In spite of the fact that all 5 were listed as "published regularly," although manual files, we were unable to locate anyone in the Commodity Economics Division or anywhere else in the Economics Research Service or Statistical Reporting Service who was aware of the availability of any of these data.

During the same period of time that the Data Inventory was being produced, another USDA data base listing appeared. In 1973, Caponio and Bracken compiled Selected Food and Agriculture Data Bases in the U.S.A. (6) which contained 53 entries, including both "citation data" bases and "information data" bases and those that were combinations. The varieties of entries of data bases ranged across CAIN (Cataloging-Indexing) in the National Agricultural Library, Pacific Scientific Information Center (a "clearinghouse"), the Environmental Data Service and NODC (NOAA), World Wide Directory of Forest Tree Geneticists (names of people), and about 35 that could clearly be expected to be numeric data files such as the Census of Agriculture, Entomology Research Chemical File, Herbarium Data Base.

The OIS is currently compiling a new directory which is expected to be published in early 1977. Although this directory will include manual files as did the previous one, it will also include, for automated files, information on file sizes. There appears to be within the USDA, as in ERDA, a strong feeling that within the agency there are a large number of data bases which could be enjoying wider distribution and utilization, even within the agency itself, thus reducing duplication of effort.

The task force report (30) presents measures of the size "Research ADP" activities within USDA from 1970 to 1973. The report, however, does not define research ADP, and there is no indication of the extent to which numeric data activities are included. These data are shown in Table 10.1

Table 10.1 RESEARCH AUTOMATIC DATA PROCESSING: ESTIMATE OF COSTS, NUMBER OF COMPUTERS AND PERSONNEL IN USDA, 1970-73

Item	Year				% increase
	1970	1971	1972	1973	1970-73
Number of Computers	17	22	28	32	88
Number of Personnel	260	322	344	363	40
Capital Equipment Investment .	.7	.4	1.1	1.0	43
Salaries	1.8	2.4	3.7	4.1	128
Other In-house Expenses	1.5	2.0	1.7	2.5	67
Contractor	1.2	2.1	.7	1.1	-8
Net Obligations	5.2	6.9	7.2	8.7	67

NOTE: Costs in millions of dollars. These estimates are probably lower than actual expenditures since many costs related to ADP are not specifically identified in agency budgets.

SOURCE: United States Department of Agriculture, USDA Researchers' Needs for Data Processing, 1973.

10.2 Botanical Data

A brief review of botanical data banks was provided by T.S. Crovello* at the Fifth International CODATA Conference (26). The following outline is extracted from that paper. Numbers in parentheses state the numbers of records in some of the files. Crovello states that the strongest computer botanical data base creation effort has been in ecology and taxonomy.

A. Living Plants

- World Plant Germ Plasm Record Center: Washington State U.
- Plant Records Center
- Daffodil Data Bank
- Wild Oat Gene Pool
- American Type Culture Collection
- Flora Cruz Project
- EXIS ("the largest")

*Chairman, International Register of Computer Projects in Systematics (Taxonomy).

B. Preserved Botanical Specimen Data Banks

- Notre Dame (65,000 & 35,000)
- RAPIC
- Flora North America
- Flora of British Columbia
- Index Nominum Genericorum (50,000)
- Atlas of the Flora of the British Isles

C. Botanical Data Banks of Laboratory Information

- Arabidopsis
- Dayhoff

D. Data Banks of Field Observations

("Biomes of U.S." projects failed to establish data banks as they couldn't resolve problems.) $30-35 million was spent to support 5 integrated research programs.

E. Non-Botanical Data Banks of Interest to Botanists

- NCC - National Climatic Center
- LARS - Laboratory for Application of Remote Sensing (Purdue), crop production; plane or satellite spectrum reflection.

10.3 Medical Data Systems

We placed little emphasis on medical data systems as such, largely because such a large proportion are not research oriented, but rather oriented to record-keeping, patient management, and health care delivery. In general, these are included under Federal Statistical Programs (Section 11.1). However any consideration of numerical and/or computerized data systems should certainly recognize the existence of medical systems. They relate, at least indirectly and sometimes directly to research in the areas of epidemiology, environmental health, disease entities, and basic biology. At least three numeric data activities of the National Institutes of Health are presented below as examples of "research" data files.

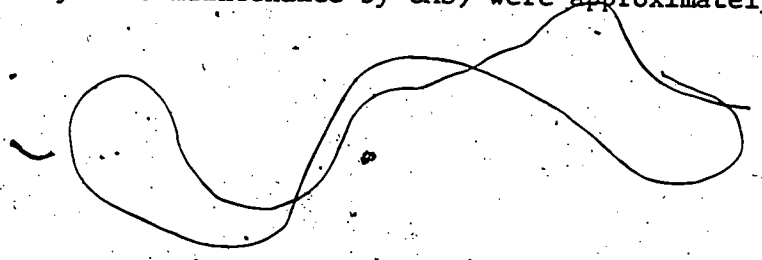
The National Cancer Institute's Surveillance, Epidemology, and End Results (NCI/SEER) data base currently contains information on about 130,000 reported cases of cancer*. New cases are added to the file at the rate of about 65,000 per year. Data are acquired through 11 local contractors and added to SEER files once a year. Contractor costs are currently $6.2 million. These costs include some activities other than data acquisition and analysis

*Mr. Geller, NCI, personal communication, December 1976.

(e.g., training) but do not include NCI's cost of maintaining the data base. Plans for the future include published output following the 1975 update, and the development of user tapes.

The Laboratory Animal Data Base is in an earlier stage of development. The project is funded by several different DHEW offices and administered by the National Library of Medicine. The file is scheduled to be on-line for a group of test users by mid-1977. It will be available to the public on a fee-for-service (subsidized) basis from Batelle Columbus Laboratores at a later date.* The file currently contains baseline data for 36,000 individual control animals. Thirteen strains of animals are currently covered, and the file will expand to 100 strains. The present 3-year contract is funded at $1.4 million.

NCI's Drug Research and Development numeric data activities include a pair of related data bases referred to as the Biological Information System and the Chemical Information System. The biologic data base consists of 16 reels of tape containing data from 4 million experiments. The Chemical Information System, covering compounds tested for anti-cancer activity, contains 230 mega-bytes** in a random-access file. The file currently includes 290,000 compounds and is expanding to include 15,000 new compounds per year. System operation is contracted to Chemical Abstracts Service (CAS) and has the CAS Chemical Compound Registry imbedded in it. Additional contractors are involved in analyses and testing of a sample of the compounds. Because about half of the compounds contained in the file are provided by industry and are highly confidential, the file may be used only by NCI and its contractors. Non-confidential data appear in published reports. Costs for the Chemical Information System (primarily file maintenance by CAS) were approximately $450,000 in FY 1976.



*Dr. Edward Greenstein, NCI, Biomedical Information Section, personal communication, December 1976.

**Sidney Richman, NCI, personal communication, December 1976.

SECTION 11

SOCIAL SCIENCES

The social sciences are currently experiencing substantial recognition as data-dependent and data-rich fields. Until the current decade, data issues in the social sciences (demography excepted perhaps) have been largely avoided by national and international organizations active in the coordination of numeric data. Social science data was really not considered - either because it was not "science", frequently non-quantitative, or simply that its boundaries were more difficult to define. The non-science assumption (compared to data in physics and chemistry for example), may have derived from the feeling that a large quantity of social science data was not generated by research experiments. Rather it resulted from tallys of numbers of events or things in the real world (numbers of financial transactions, or numbers of people for instance). A great deal of numeric data was collected by governments for the purpose of legislative and programmatic decision-making. Additional amounts of data were (and are) collected by industry also for management decision-making. Whatever the reason for the decision to collect such data, the fact remains that they are available for scientific analysis.

The combining of social science data with hard science data, as in (human) life and environmental sciences has helped to contribute to the current awareness that social science data must be considered important segments of scientific and technical data. As the social sciences continue to generate more and more data from experimental rather than emperical observations, this view is reinforced.

This section discusses two aspects of social science data, the principal statistical programs of the U.S. Government and a directory of social science data bases.

11.1 Federal Statistical Programs

The most prominent data collections in the social and behavioral sciences are perhaps in areas of demography and economics. Certainly a large

number of Governmental data activities are in these areas.

One of the problems in defining boundaries to Scientific and Technical Information (STI) is that in recent years concurrent with the greater inclusion of the social sciences in the definition of sciences, STI has changed from "information generated by scientists" to "information used by scientists". Many types of statistical information programs fall with in the latter scope, especially in two broad areas. One area covers business and financial data, the other Federal public-use statistical data programs. The two sometimes overlap. This section presents highlights of the Federal statistical programs.

The term "public use" (a term used primarily within the United States) applies to certain types of automated files, usually produced by governments and containing data which are legislatively accessible to the general public. In order to qualify as public use data, a data file (file series, or set of data files) has usually undergone at least two transformations. First, all personal (and sometimes organizational) identifiers have been removed. This process frequently requires aggregation to the level where no single cell contains a number smaller than (usually) five. This procedure is followed before Federal data is published in any form. In addition, to create public use files from agency files requires the documentation and structuring of a file so that a variety of access programs can be written. These programs may or may not be part of the package which is subsequently made available to the public. The complexities and expense of this latter effort are extreme and often prohibit the presentation of file copies to the public, although files may continue to be used internally and summaries published.

The subject fields included in public use data fall largely within the disciplines of social sciences, and to a lesser (but growing) extent, environmental sciences. These data are all products of government data collection efforts. They are characteristically associated with various censes, economic transactions or environmental phenomena. Associated with the "public use" tag are the file designs, programs and other supporting documentation which enable

the data files to be utilized by the general public. A number of organizational structures have evolved, both within the governmental sector and within the private and university sectors, to deal with the utilization of public-use files. The non-governmental organizations may provide access to both public files and commercially produced files.

11.1.1 Levels of Effort for Public Use Reports

In early 1975, the Office of Management and Budget (OMB) generated a "List of the 100 Active Public-Use Reports Having the Largest Man-Hour Burden" (91), sometimes refered to as "the hundred most burdonsome reports". The list actually includes 201 programs (which generate reports). It contains two separate lists of "100"; one for repetitive data collection programs (101 entries) and one for one-time programs. According to OMB*, the report-producing programs covered on this list account for 60 percent of the total Federal data collection effort.

Included in the lists are repetitive programs requiring 250,000 or more man-hours annually, and single-time programs using more than 7,500. The repetitive programs total over 79 million annual man-hours. The single-time programs as a group add just over 8 million annual man-hours yielding a grand total of 87 million man-hours per year as of FY 1975. Only six of the single-time programs require 250,000 or more annual man-hours each and together account for 5 million man-hours.

Number of programs by agency is shown in Table 11.1. The agencies having the largest number of programs are Department of Health Education and Welfare with 31 percent, and the Departments of Labor and Commerce with 11 percent each. Considering only the larger programs (250,000 or more man-hours annually), DHEW still holds the lead with about one third, DOL accounts for 15 percent and the Department of Agriculture has responsibility for 10 percent. Commerce has dropped to 4 percent. As these larger 107 projects, while representing only 53 percent of the total number of programs, account for over 95

*Joseph Duncan, OMB, personal communication.

Table 11.1 LARGE FEDERAL PUBLIC USE PROGRAMS BY AGENCY, SIZE[1] AND CLASS OF PROGRAM: 1975

Program Size[1]	Total No.	Total %	USDA	DOC	DOD	DHEW	HUD	Civil Service	DOL	DOT	Treasury	V.A.	All Other[2]
Total	201	100	18	22	5	63	9	6	23	13	5	12	25
Percent	100	-	9	11	2	31	4	3	11	6	2	6	12
>250,000	94	47	7	18	2	31	3	2	7	6	1	3	14
Percent	100	-	7	19	2	33	3	2	7	6	1	3	15
>250,000	107	53	11	4	3	32	6	4	16	7	4	9	11
Percent	100	-	10	4	3	30	6	4	15	7	4	8	10
Program class for programs >250,000 hrs.	107	100											
A. Applications	32	30	1		2	14	2	2	1	1	2	4	3
B. Program Evaluations	44	41	8		1	11	2	1	7	3	2	2	7
C. Statistical survey	11	10		4		2			4			1	
D. Other Management	12	11	2			3			2		2	2	1
E. Record Keeping	3	3					1		2				
F. Other	5	5				2	1	1		1			
"STI" subtotal (B-D, E)	72	67	10	4	1	18	3	2	13	6	2	5	8
Percent	100	-	14	6	1	25	4	3	18	8	3	7	11

[1] By number of man-hours expended annually.

[2] Includes Departments of Justice, State and Interior, EPA, GSA, NASA, SSS, NSF, SAODAP, ACTION, FRB, and U.S. International Trade Commission.

SOURCE: Office of Management and Budget. "List of 100 Active Public-Use Reports Having the Largest Man-Hour Burden." 1975.

percent of the labor involved, the remainder of this discussion addresses these programs only. Classes of programs are also shown in Table 11.1. Of all programs which require 250,000 or more man-hours annually, approximately one-third process data from applications (for example, applications for Medicaid or welfare benefits) or produce only record-keeping reports. These classes appear least likely to relate directly to STI. If these two classes are eliminated from consideration, the total number of programs is reduced to 72. In additional total man-hours are reduced by roughly half to 49 percent. The number of man-hours spent on the smaller subset of programs is now only 41 million annually.

In Table 11.2 man-hours for these 72 programs are distributed by agency. Mean level of effort per project is also shown by agency. These data indicate that the larger projects lie within HUD and Commerce.

The number of respondents per program is a less meaningful measure because "respondents" may be individuals, corporations, states, etc. There was an extremely broad range in man-hours per respondent among individual programs, from less than one tenth of an hour per respondent to over 10,000 hours per respondent! The latter was a one time study in which respondents were states; it is assumed that both the questionnaire and the analysis were lengthy and detailed. However, means by class of program (in Table 11.3) indicate surprisingly little variation (except for Class F, "Other") and ranged from 19 to 31 minutes per respondent.

Table 11.2 SELECTED FEDERAL PUBLIC USE PROGRAMS*
BY AGENCY: 1975

Agency	Man-Hours		Projects		Mean Man-Hours per Program	Man-Hour Index (72)
	Total (000)	%	No.	%		
USDA	4,720	11	10	14	118	0.2
DOC	4,791	12	4	6	1,198	2.1
DOD	425	1	1	1	425	0.7
DHEW	12,038	29	18	25	669	1.2
HUD	4,640	11	3	4	1,547	2.7
Civil Service	1,065	3	2	3	532	0.9
DOL	6,807	16	13	18	524	0.9
DTO	1,799	4	6	8	300	0.5
Treasury	630	2	2	3	315	0.5
VA	1,816	4	5	7	363	0.6
Justice	885	2	2	3	442	0.8
SSS	583	1	2	3	292	0.5
GSA	420	1	1	1	420	0.7
EPA	252	1	1	1	252	0.4
NASA	258	1	1	1	258	0.4
ACTION	250	1	1	1	250	0.4
Total	41,379	49	72	67	575	1.0
Application & Record-keeping Programs	37,746	51	35	33	1,078	1.9
Totals for Largest Programs	84,386	100	107	100	789	1.3

*Four classes only: Program evaluations, statistical surveys, other management and other.

SOURCE: Office of Management and Budget. "List of 100 Active Public-Use Reports Having the Largest Man-Hour Burden." 1975.

Table 11.3 MANHOURS EXPENDED BY CLASS, 107 LARGEST FEDERAL PROGRAMS: 1975

(all agencies)

Program Class	Annual Man-hrs. (000)	No. of Programs	Mean Man-hrs. per Program (000)	Man-hrs. per Respondent
A. Applications	39,509	32	1,235	.41
B. Program Evaluations ..	22,714	44	516	.40
C. Statistical Surveys ..	7,787	11	708	.45
D. Other Management	4,887	12	407	.32
E. Record Keeping	3,498	3	1,166	.51
F. Other	5,991	5	1,198	.92
Total	84,386	107	789	.40
Largest program	5,100	-	5,100	10,700.00
Smallest program	250	-	250	.09

SOURCE: Office of Management and Budget. "List of 100 Active Public-Use Reports Having the Largest Man-Hour Burden," 1975.

11.1.2 Obligations for Principal Statistical Programs

Federal obligations for principal statistical programs reveal that the top four agencies in total man-hours (calculated for 107 projects) were also the top four in terms of obligations in 1972, but only three were in the top four in 1975. Two other apparent discrepancies appear - both EPA and the Department of Justice have much larger obligations for statistical programs in 1975, than their position on the "100 most burdensome" list would seem to indicate. Table 11.4 displays annual obligations for 1972, 1975 and 1977 in millions of dollars for all agencies with annual obligations of $10 million or more estimated for 1977.

The distinction between 'current' and 'periodic' programs is not consistent with those used for the level of effort data presented in Section 11.1.1. If any comparison exists it is that "repetitive" programs in Section 11.1.1 above are not the "periodic" programs here.

Table 11.4 OBLIGATIONS FOR PRINCIPAL STATISTICAL PROGRAMS, SELECTED AGENCIES: 1972, 1975, 1977

(Millions of dollars)

Agency	1972 Current	1972 Periodic	1972 Total	1975 Current	1975 Periodic	1975 Total	1977 (estimate) Current	1977 (estimate) Periodic	1977 (estimate) Total	% Increase in total 1972-1977
USDA*	26.2	-	26.2	35.2	-	35.2	48.4	-	48.4	85
DOC*	36.4	28.1	64.5	58.7	22.7	81.4	67.7	47.4	115.1	78
DHEW*	73.5	-	73.5	108.2	-	108.2	132.6	-	132.6	80
DOL*	54.5	4.2	59.7	70.3	7.0	77.3	96.3	7.2	103.5	73
HUD	2.4	-	2.4	8.8	-	8.8	11.3	-	11.3	371
Treasury ...	10.7	-	10.7	15.6	-	15.6	15.8	-	15.8	48
DOI	6.0	-	6.0	16.2	-	16.2	18.5	-	18.5	208
Justice.....	8.2	-	8.2	46.0	-	46.0	38.8	-	38.0	363
DOT	7.0	-	7.0	15.7	-	15.7	23.1	-	23.1	220
EPA	17.2	-	17.2	26.0	-	26.0	21.4	-	21.4	24
FEA	-	-	-	9.3	-	9.3	10.4	-	10.4	n.a.
Sub Total ..	242.1	32.2	274.4	410.0	29.7	439.7	484.3	54.6	538.9	96
Total	9.6	-	9.6	18.1	-	18.1	19.9	-	19.9	107
Other agencies .	251.7	32.3	284.0	428.1	29.7	457.8	504.2	54.6	558.8	97

*The five major statistical services of the Federal Government are within these four agencies. They are; the Bureau of the Census (DOC), Statistical Reporting Services (USDA), The National Center for Educational Statistics (DHEW), The National Center for Health Statistics (DHEW), The Bureau of Labor Statistics (DOL).

SOURCE: Office of Management and Budget, "Principal Federal Statistical Programs" reprints from Special Analysis, Budget of the U.S. Government, January 1973, and January 1976.

Another breakdown is provided by OMB (93, 94) in which statistical obligations by subject areas are shown. Economic statistics account for over 50 percent for the three years shown in Table 11.5. Although the remainder were called Social and Demographic statistics in 1972, the emergence of Energy statistics as a separate category in 1975 and the inclusion of Environmental statistics within other social sciences raises some questions as to whether there is an evolving third category of Environment/Energy/Other non-social Sciences. However, this category has not yet made dramatic gains in terms of the percent of total dollar obligations.

Between 1972 (actual) and 1975 (estimated) obligations for all principal current Federal statistical programs will have increased by 100 percent. The largest increases were in the areas of Criminal Justice (363 percent), Price Statistics (117 percent), Construction and Housing (165 percent), National Income Accounts (148 percent) and Population (131 percent). In 1972, total obligations amounted to 284 million, expected to increase by 97 percent in 1977 to $559 million.

11.2 Other Social Science Data Sources

In 1973 Vivian S. Sessions* compiled a timely Directory of Data Bases in the Social & Behavioral Sciences (110). Sessions sought to include in this directory both U.S. and foreign sources which maintained mainly non-bibliographic machine-readable data bases. In the preface to this volume, Mira Rees states:

> There are some interinstitutional archives-noteably the Inter-University Consortium for Political Research at the University of Michigan and the National Opinion Research Center at the University of Chicago, and, in Britain, the Social Science Research Council's Data Archive at the University of Essex to add a European example; however, these archives by no means exhaust either the need for or the supply of collectable data that is required for todays' research in the academic community.

*Formerly Director for the Center for Advancement of Library Information Science of the City University of New York

Table 11.5 OBLIGATIONS FOR PRINCIPAL CURRENT STATISTICAL PROGRAMS BY BROAD SUBJECT AREAS: 1972, 1975, 1977

(Millions of dollars)

Subject Area	1972		1975		1977 estimate		% Increase
	$	%	$	%	$	%	72-77
Labor	57.9	23	66.0	15	83.2	16	44
Prices & Price Indices	11.4	4	15.1	4	24.7	5	117
Production & Distribution	57.7	23	86.3	20	110.4	22	91
Construction & Housing	7.4	3	15.8	4	19.6	4	165
National Income & Business Financial Account	16.5	7	35.2	8	41.0	8	148
Subtotal Economic	150.9	60	218.4	51	278.9	55	85
Health	45.3	18	76.5	18	91.2	18	101
Population	3.2	1	6.3	2	7.4	2	131
Education	15.4	6	16.6	4	21.4	4	39
Crime	8.2	3	45.2	11	38.0	8	363
Social Security & Welfare	11.5	5	19.9	5	22.9	4	99
Environment	17.2	7	30.9	7	26.8	5	56
Subtotal Demographic & Social	180.8	40	195.4	46	207.7	41	106
Energy	-	-	14.3	3	17.6	4	n.a.
Total	251.7	100	428.1	100	504.2	100	100

SOURCE: Office of Management and Budget, "Principal Federal Statistical Programs" reprints from Special Analysis, Budget of the U.S. Government, January 1973, and January 1976.

The Social Science...Data Archive referred to may be well known as the social science data archive to social scientists. However, additional social science (data) archives are listed in the Directory and interject some confusion. Similarly-named archives are located at the University of Iowa, University of Michigan and Yale University among others.

The introduction to the directory provides the frame of reference for inclusion and is not dissimilar to the frame of reference for this report:

> Organized collections of...individual bits of data constitute data files or data bases, the subject of this directory.
>
> The major thrust of this directory is the identification of the nonbibliographic data bases; it does, however, include a few highly specialized files that refer to research reports and similar documents.
>
> Although there are, undoubtably, many individual collections of research data in a variety of private and institutional settings, it is virtually impossible to gain knowledge of those that exist outside an organized center of some type.
>
> In the larger archives, there are so many discrete data bases, contributed by different investigators, that the archival centers often issue a codebook listing the titles of the files, and, sometimes, such information as the name of the investigator, types of variables, and the sizes of individual files. Since these codebooks are available, at least to members of archival centers, the holdings of the large archives are represented in this directory as listings of data base categories rather than as listings of individual file titles.
>
> With the exception of the 1970 Census, the federal data were not pursued [for the purposes of extensive followup] for the very reason that there are so many existing guides to federal statistical series.
>
> The volume is particularly strong in the areas of urban and regional planning because these are areas into which the federal government has put a great deal of experimental money.
>
> A...limiting factor was that many governmental data centers responded that their files were confidential, and many centers in the private sector responded that their files were proprietary.

Information for the Directory was obtained by mail out questionnaire and information in the Directory is given as stated by the respondents with

occasional comments by the editor. Information includes, but is not limited to; title of institution, center, file, subject field and time frame of data, data source, storage media, hardware, software, output media and products, and access. Not all information is provided for all files. No information is given for file size other than time frame of data.

Distinction between institution, center and file are provided in Figure 11.1.

Table 11.6 presents our count, for U.S. and foreign locations of the number of centers, institutions, and files which the directory covers. Numerous institutions, especially in the U.S., listed more than one data center as a source of social and behavioral science data bases. Numerous centers both in the U.S. and abroad maintained more than one file.

Table 11.6 NUMBER OF INSTITUTIONS, CENTERS & FILES COVERED IN THE DIRECTORY OF DATA BASES IN THE SOCIAL & BEHAVIORAL SCIENCES: 1973

Designation	Total	U.S.		Foreign	
		No.	%	No.	%
Institutions	455	327	72	128	28
Centers	685	547	80	138	20
Files (estimate)	1,220	976	80	244	20

SOURCE: Sessions, Vivian, Directory of Data Bases in the Social and Behavioral Sciences, 1973.

Approximately 72 percent of the institutions covered in this directory are in the U.S., while 80 percent of the centers and of the files are in the U.S. It was not possible from the information provided to estimate file sizes, nor to determine in what proportion of cases different centers or institutions were providing services based upon the same or similar data bases (as might be expected to occur especially with 1970 Census data files).

Subject fields having over 50 entries each in the subject index, (in which any file could be indexed under multiple fields) were:

Community Health
Demography
Economics
Education
Ethnic Group Studies and Anthropology
Geography
Housing
International Studies
Land Use
Manpower
Marketing
Political Science and Law
Population
Psychology and Psychiatry
Public Administration
Public Opinion
Regional Development
Social Welfare
Sociology
Transportation
Urban Development

SECTION 12

SUMMARY AND CONCLUSION

The present report on numeric data bases is indicative only. A direct survey of numeric data bases was not conducted as had been anticipated, because it became apparent that such a survey conducted within the resources allocated would not provide data which could be aggregated or analyzed statistically. The most striking feature of the results of the study reported on is the intractability of the information which has been published or which was provided by informants.

The available measures which in some way reflect the magnitude of numeric data operations include size of holdings, rate of growth of holdings, volume of services and products, number of individuals or organizations involved, and costs. The values of any of these measures could only rarely be intercompared among distinct operations because of differences in units of measure and frames of reference. Even within the same operation it was generally possible to obtain either no total magnitudes at all or only pro forma totals.

Groupings of operations which were used in various published reports differ among themselves, even within different editions of a single series of reports. It is often not clear how much of an operation is properly described as the numeric data activity, whether it is a field-based operation with a mission of collecting numeric data or a service function which includes numeric data management or processing as one service aspect among several. In addition the service function relating to numeric data may be well hidden if it only provides services to staff members.

Outright contradictions occur between supposedly reliable distinct sources. Personal communications from individuals involved in particular operations sometimes disagree with published statements, and different publications disagree with each other. It can be frustrating indeed to identify a wide discrepancy between the information contained in documents published at about the same time by the same organization and then to receive in explanation, "Well, I guess they were written by different people."

The numeric data bases and operations mentioned in this report certainly do not include all such resources, nor even almost all nor all significant ones. There are undoubtedly additional large data bases and very important smaller ones, not to mention numerous others of comparatively limited use. It is quickly evident that secondary sources do not suffice even to identify which known information activities comprise or consist of numeric data activities. Existing descriptions are ambiguous on this point.

Entries from both the directories of so-called data bases as well as more comprehensive directories of information resources are more frequently than not unspecific about the existence and availability of numeric and or other factual data. "Holdings" of an organization may specifically be stated as; data, statistical and demographic data, maps, charts, tables, technical data sheets, items of unevaluated data, data compilations, graphs, specimens, technical data, data forms, photos, unpublished data, tapes, cards, graphics. "Publications" may include; data, maps, charts, technical data, sheets, specifications, samples. Additional types of publications with even less indications of whether numeric data is contained include those for which publication titles are given and; technical reports, newsletters, pamphlets, books, journals, slides, microforms, circulars, annual reports, annual reports on data, and so forth. Descriptions of "areas of interest" are even less illucidating. Nor can the presence of electronic data processing equipment be assumed to indicate the handling of numeric data. Such equipment frequently may be primarily dedicated to word processing, type setting, accounting or other administrative purposes, bibliographic or other reference files, etc. In general, it is rare that an accurate understanding of whether any or all of these descriptors indicate actual involvement with numeric data can be acquired. Some organizations which do not provide any direct allusions to data in these directory entries, are found upon further investigation to be very active in handling of numeric data.

CODATA's own attempt to provide a directory of numeric data activities was published in 1969. Recent publications from CODATA and others stress the importance of providing an updated and more complete edition of the International Compendum of Numerical Data Projects (14). CODATA appears more conscious than

other groups of the extent of the difficulties of such a project. This awareness is one of the major reasons why an update has not been provided earlier. When it is done, it is expected to be accomplished in stages, starting with the most well-defined projects.

Even beyond all the considerable technical and coordination problems associated with collection, evaluation, and processing and storage of data are considered--and these take on seemingly insurmountable proportions--two other problem areas have been identified, and are expected to continue to plague present and future efforts; funding (insufficiency) and marketing. In some sense the latter is the more severe problem as there is some reason to anticipate that funding levels may increase as the services become more widely used. The marketing of services and products of data centers is seen as particularly inadequate - limiting use by scientists who could benefit as well as impacting negatively on problems of coordination, coverage (duplication and gaps) and income. The review of the literature in this area provides no consistent pattern, except that number of users appears to be small, frequently smaller than the discipline or topic should warrant. In effect, not only are there many more scientific areas which lend themselves to the accumulation of numeric files but there appears to be a much wider potential audience for existing data files than is currently being realized.

Although the area of standards and compatability of systems is one in which CODATA is heavily involved, this is considered to be a less severe, e.g. restricting, problem than the other two. It is, however, critically important that solutions are arrived at initially, rather than recognizing their importance after the fact.

These major problems and potential solutions can be summarized as follows:

Funding Levels:	solution; increased marketing
Marketing:	solution; advanced techniques in publicizing services, education in use of data services, directories
Standards and compatability of systems:	solution; greater concern for international and intranational coordination prior to as well as during data base creation

To the degree to which these problems have not been resolved, the immediate and continuing effect is inaccessibility to users and potential users (further exacurbating the funding problem). From the potential user's side, the difficulties can be stated as:

- lack of awareness
- restrictive cost (usually a minor impediment)
- non-compatability of data from various sources
- use restricted to certain groups of individuals
- lack of sophistication in the use of automated data

APPENDIX A: SURVEY PRE-TEST

A total of 22 centers were selected from the Kruzas Encylopedia (56), and sent a data collection form. The effort was, in effect, a pilot test to see if any useful information on file sizes and growth rates could be collected in this manner. In almost all cases it was necessary to make an initial telephone contact to determine the full (current) address and the individual to whom the questionnaire could be addressed. In some cases it was found that the center we originally intended to contact was less likely to have data files than another center within the same organization. Of the 22 questionnaires mailed, returns were received from 15. Three of these fifteen (two industrial and one archaeological survey) indicated that their organization did not have any numerical data files. It is felt that the questionnaire either never reached the proper office, or that the organization preferred not to provide any information about their files.

The selection of centers to contact was based upon subjective criteria and a desire to contact those that appeared, from the entries in the Kruzas directory, to be most likely to have numeric files. It was also desired to cover a range of scientific disciplines and types of organizations: Federal, state, university, private industry; large versus small. It was a purely shotgun approach with very few pellets! The range of fields of science (judging from the results) was not broad and included only physical, social, environmental and life sciences.

File Sizes

Many respondents found this question difficult to answer and considered the number of tapes an irrelevant number. One telephone inquiry elicited the following comment, "You wouldn't count the number of books in a library to estimate the size of the library, would you?" The affirmative response did little to mollify him. However, the problem is understandable. Density in terms of number of bits of data per tape varies (current estimates seem to indicate that densities run either 800 or 1,600 bits per inch) and has increased over the years. Even given bpi, there is no way of knowing how much of the tape is filled or in effect how many inches of tape contain data. Further, the data

bits on the tape consist of an assortment of information carrying bits (i.e., control characters, linkages, field and record identifiers, etc.) which are not part of the "data" as it would be listed in print form.

Respondents were given an opportunity to answer the "file size" question using some other measure than number of tapes. All other measures are frought with similar caveats. Seven respondents provided some indication of the size of their tape libraries. Two were small (< 5 tapes) but still showed 100 percent increases from 1970-1975. Four others showed increases of from approximately 250 percent to over 4,000 percent between 1970 and 1975 (the lower percentages were increases from 1971 or 1972, thus the low end of the range would be higher still if 1970 data had been provided). Little additional knowledge is gained from adjusting for bpi - the low end of the percent increase range becomes almost 1,000 percent. We tend at this point to be reduced to agreeing with the oft' stated response to our question about file size growth rates ("very fast"). The small amount of data obtained does not allow for much in terms of quantitative measure. The experience does, however, underline the importance of creating different types of data collection instruments for different classes of organizations and types of files.

Users

From the small amount of data acquired, there was no indication that either the magnitude of the number of users nor the annual rates of increase bear any relationship to file sizes.

Four of the non-Federal organizations have had static "user" populations. (For purposes of this discussion a broad definition of user is applied in which are included users of both services and products). Two provided no information at all on users. Only one indicated that number of users was relative to file size. In this case the only information on file size was "10,000 time series" (on-line) in 1969 growing to "several million" in 1975 - the same quantities were used for number of users. Another respondent (academic) indicated a doubling of the number of users from 1972 to 1975 (1,500 to 3,000 respectively), while the file size doubled from 1966 to 1975.

For the four Federal organization files, one sells only hard copy products, and all showed consistent growth in number of users. User populations have at least doubled for all four between 1970 and 1975.

Foreign sales were limited to the Federal organizations and amounted to as much as 25 percent only for the World Data Centers. Only one non-Federal organization provided any sales data, and indicated annual increases of over 50 percent in the early 1970's, but an almost 50 percent decrease from 1974 to 1975.

The exercise proved to be fruitless in terms of collecting useful data. It was clearly demonstrated that such data collection was difficult or impossible. The small amount of quantitative information obtained from the four Federal agencies is incorporated in the discussions elsewhere of these operations. These were:

Machine-Readable archives Section, National Archives
World Data Center A for Solar Terrestrial Physics (EDS, NOAA)
Solar Earth Data Service Division, (NGSDC, EDS, NOAA)
Water Resources Division, USGS.

MARKET FACTS, INC.

NUMERICAL DATA SYSTEMS SURVEY

COLLECTION FORM

Organization Name: _____

Organization Address: _____

_____

Telephone Number: ( ) _____ Extension: _____

Name of Individual Completing Questionnaire: _____

1a. Does your organization have "Numeric Data Files", i.e., files consisting of mainly machine-readable numeric data (non-bibliographic, non-narrative)?

Yes

No (GO TO QUES. 12)

1b. In what year did your organization establish numeric data files? _____

2. In order to measure growth rates, please state whatever measures you can, for whichever years you can (1965-1975), of the size of your data files.

Year	Number of Tapes and Approx. Average Density	Number of Card Decks and Approx. No. of Cards	Number of other: maps, photo images, etc. Please identify	Any Other "Size of file" Measurement You Wish to Provide
1965				
1966				
1967				
1968				
1969				
1970				
1971				
1972				
1973				
1974				
1975				

Is this FY or CY data? FY
CY

3. Who are your main contributors and/or the main sources of your data? Please list. e.g., NOAA, NASA, FDA, NTIS, individual researchers, universities, oil dealers, internal extractions of data from narrative material, etc.

4. Please circle the fields of science with which your files are mainly involved and a percentage estimate of your holdings in this area.

Fields of Science	Percent of Holdings
1. Physical Sciences Data	
2. Mathematics Data	
3. Computer Science & Engineering Data	
4. Environmental Sciences Data	
5. Engineering Data	
6. Life Sciences Data	
7. Psychology Data	
8. Social Sciences Data	
9. Other_____	

5. Does your organization honor requests for your data from outside users?

Yes

No (GO TO QUES. 6b)

6a. Please state the number of requests filled for as many years as possible between 1965-1975, for U.S. vs foreign requests, and individual (company) versus institution vs distribution organizations.

	Number of Requests Filled		Number of Requests Filled			
Year	Foreign	U.S.	Individual Company	Institution	Distribution Organizations*	Total Requests
1965						
1966						
1967						
1968						
1969						
1970						
1971						
1972						
1973						
1974						
1975						

*A distribution organization is considered an organization which would make your data files available to users.

Is this FY or CY? FY

6b. If tapes, card decks, etc. are used internally within your organization, please state number of tapes, decks, etc. and number of users or number of hours used, or whatever measurement you have for whichever years you have.

Year	Number of Tapes	Number of Card Decks	Number of Users, or Number of Hours Used or Whatever Measurement you Have
1965			
1966			
1967			
1968			
1969			
1970			
1971			
1972			
1973			
1974			
1975			

7. Please state the total dollar value of sales of services (use of card decks, maps, tapes, etc.), for as many years as possible between 1965-1975 by foreign versus U.S. sales, and by individual (company sales) versus institution vs Distribution Organization Sales.

	$ Amount of Sales		$ Amount of Sales			
Year	Foreign Sales	U.S. Sales	Individual or Company Sales	Institution Sales	Distribution Organization Sales	Total Sales
1965	$	$	$	$	$	$
1966	$	$	$	$	$	$
1967	$	$	$	$	$	$
1968	$	$	$	$	$	$
1969	$	$	$	$	$	$
1970	$	$	$	$	$	$
1971	$	$	$	$	$	$
1972	$	$	$	$	$	$
1973	$	$	$	$	$	$
1974	$	$	$	$	$	$
1975	$	$	$	$	$	$

8. Please state the total dollar value of sales of files (copies of tapes, card decks, etc.), for as many years as possible between 1965-1975 by foreign versus U.S. sales, and by individual (company sales) versus institution vs Distribution Organization Sales.

	$ Amount of Sales		$ Amount of Sales			
Year	Foreign Sales	U.S. Sales	Individual or Company Sales	Institution Sales	Distribution Organization Sales	Total Sales
1965	$	$	$	$	$	$
1966	$	$	$	$	$	$
1967	$	$	$	$	$	$
1968	$	$	$	$	$	$
1969	$	$	$	$	$	$
1970	$	$	$	$	$	$
1971	$	$	$	$	$	$
1972	$	$	$	$	$	$
1973	$	$	$	$	$	$
1974	$	$	$	$	$	$
1975	$	$	$	$	$	$

9. Do you have publications which are a result of your numeric data system. (e.g., published print on paper or microform products extracted from data base)? A catalog or index of the machine-readable files or a printout or hard-copy of a data tape or card deck are to be included here; however, do not include files which are bibliographies of literature.

Yes

No (GO TO QUES. 12)

10. Please state (for whichever years you can) the type of publication (index, catalog, hard-copy of data file), the issuance frequency of publication, the types and number of subscribers, number of copies sold, total dollar value of sales.

Year	Type of Publication (Index, Catalog, Data Compilation)	Frequency, (Weekly, Monthly, Annually, occasionally, etc.)	No. of Subscribers, Total No. of Copies Annually						Total Sales $
			Individual or Company		Institution		Distribution Organizations		
			# of Subscribers	Total # of Copies Annually	# of Subscribers	Total # of Copies Annually	# of Subscribers	Total # of Copies Annually	
1965									
1966									
1967									
1968									
1969									
1970									
1971									
1972									
1973									
1974									
1975									

11. The information you have provided will be of great assistance in measuring the growth of machine-readable numeric files over the last decade; in addition, this information will provide an indicator of the growth that is expected in this area in future years.

We thank you for your time in completing this data collection form.

12. PLEASE RETURN THIS FORM TO:

Market Facts, Inc.
6110 Executive Boulevard, Suite 835
Rockville, Maryland 20852

Attn: Ms. Kathleen McEvoy

APPENDIX B

FIGURES

Figure 6.1 FEDERALLY FUNDED RESEARCH AND DEVELOPMENT CENTERS, 1973-75

Name	Sponsoring agency	Organizational affiliation
	Administered by universities	
Ames Laboratory	Atomic Energy Commission	Iowa State University of Science and Technology
Applied Physics Laboratory	Department of the Navy	Johns Hopkins University
Applied Research Laboratory	Department of the Navy	Pennsylvania State University
Argonne National Laboratory	Atomic Energy Commission	University of Chicago and Argonne Universities Association
Brookhaven National Laboratory	Atomic Energy Commission	Associated Universities, Inc.
Cambridge Electron Accelerator	Atomic Energy Commission	Harvard University
Center for Naval Analysis	Department of the Navy	University of Rochester
Cerro Tololo Inter-American Observatory	National Science Foundation	Association of Universities for Research in Astronomy, Inc.
E.O. Lawrence Berkeley Laboratory	Atomic Energy Commission	University of California
E.O. Lawrence Livermore Laboratory	Atomic Energy Commission	University of California
Jet Propulsion Laboratory	National Aeronautics and Space Administration	California Institute of Technology
Kitt Peak National Observatory	National Science Foundation	Association of Universities for Research in Astronomy, Inc.
Lincoln Laboratory	Department of the Air Force	Massachusetts Institute of Technology
Los Alamos Scientific Laboratory	Atomic Energy Commission	University of California
Fermi National Accelerator Laboratory	Atomic Energy Commission	Universities Research Association, Inc.
National Astronomy and Ionosphere Center	National Science Foundation	Cornell University
National Center for Atmospheric Research	National Science Foundation	University Corporation for Atmospheric Research
National Radio Astronomy Observatory	National Science Foundation	Associated Universities, Inc.
Oak Ridge Associated Universities	Atomic Energy Commission	Oak Ridge Associated Universities
Plasma Physics Laboratory	Atomic Energy Commission	Princeton University
Space Radiation Effects Laboratory	National Aeronautics and Space Administration	College of William and Mary
Stanford Linear Accelerator Center	Atomic Energy Commission	Stanford University
	Administered by industrial firms	
Bettis Atomic Power Laboratory	Atomic Energy Commission	Westinghouse Electric Corporation
Hanford Engineering Development Laboratory	Atomic Energy Commission	Westinghouse-Hanford Corporation
Knolls Atomic Power Laboratory	Atomic Energy Commission	General Electric Company
Liquid Metal Engineering Center	Atomic Energy Commission	Rockwell International Corporation
Mound Laboratory	Atomic Energy Commission	Monsanto Research Corporation
National Reactor Testing Station	Atomic Energy Commission	Aerojet Nuclear Corporation
Oak Ridge National Laboratory	Atomic Energy Commission	Union Carbide Corporation
Sandia Laboratory	Atomic Energy Commission	Western Electric Company, Inc.-Sandia Corp.
Savannah River Laboratory	Atomic Energy Commission	E.I. du Pont de Nemours & Co., Inc.
	Administered by other nonprofit institutions	
Institute for Defense Analysis	Department of Defense	Institute for Defense Analysis
Research Analysis Corporation*	Department of the Army	Research Analysis Corporation
Aerospace Corporation	Department of the Air Force	Aerospace Corporation
Analytic Services, Inc.	Department of the Air Force	Analytic Services, Inc.
MITRE Corporation	Department of the Air Force	MITRE Corporation
RAND Corporation	Department of the Air Force	RAND Corporation
Atomic Bomb Casualty Commission	Atomic Energy Commission	National Academy of Sciences
Pacific Northwest Laboratory	Atomic Energy Commission	Battelle Memorial Institute

*Phased out 9/72.

SOURCE: National Science Board, National Science Foundation, Science Indicators, 1974, (p. 201).

Figure 6.2

INFORMATION ANALYSIS CENTERS IN THE OAK RIDGE NATIONAL LABORATORY: 1968

1. Accelerator Information Center - hardware design
2. Actinide Research Information Center - (new) properties
3. Atomic and Molecular Processes Information Center - collects & evaluates data
4. Biogeochemical Ecology Research Collection* - tabular probability estimates
5. Charged-Particle Cross Section Data Center - collects & evaluates data
6. Civil Defense Research Collection* - multi-disciplinary
7. Computer Handling of Reactor Data: Safety - design of nuclear power plants
8. Criticality Data Center - inputs to safety
9. Office of Saline Water Materials Information Center - desalinization processes
10. Engineering Data Collection* - reference file of (primarily ORNL) drawings & specifications
11. Information Center for Internal Exposure - data interpretation
12. Isotopes Information Center - industrial applications
13. Molten Salt Information Center
14. Nuclear Data Project - data compilation
15. Nuclear Desalinization Center
16. Nuclear Fuel Technology Information Center
17. Nuclear Safety Information Center*
18. Photographic Reference Collection* - ORNL internal
19. Radiation Shielding Information Center
20. Research Materials Information Center
21. Machining and Gauging Information Center

*"Collection" usually means a document collection, which may pre-stage full-fledged information center activity.

SOURCE: Kertesz, Francois. "The Information Center Concept" in Key Papers In Information Science, 1971.

Figure 6.3

FEDERALLY SUPPORTED INFORMATION ANALYSIS CENTERS: 1974

Center	Sponsor(s)	Administration/Location
Aerospace Safety Research and Data Institute	NASA	Lewis Research Center
Air Pollution Technical Information Center (APTIC)	EPA	Research Triangle Park
Alloy Data Center	NBS (NSRDS)	Metallurgy Div.
Analysis and Evaluation Group	AEC	Lawrence Livermore Lab
Argonne Code Center	AEC	Argonne National Lab
Atomic Energy Levels Data Center	NBS (NSRDS)	Institute for Basic Standards
Atomic Transition Probabilities Data Center	NBS (NSRDS)	Institute for Basic Standards
Bathythermograph (BT) Data Processing and Analysis Facility	NSF, DOD (ONR), NOAA (NODC), etc.	Scripps Inst. of Oceanography
Berkeley Particle Data Center	AEC, NBS (NSRDS), NSF	Lawrence Berkeley Lab
Biomedical Studies Group (BMS)	NLM (TIP), USFS, EPA	Oak Ridge Nat'l Lab
Brain Information Service	NIH (NINDS)	UCLA School of Medicine
Bureau of Mines. Assistant Director, Mineral Supply	DOI	Bureau of Mines, Arlington
Bureau of Mines--Mineral Supply. Alaska Field Operation Center	DOI	Bureau of Mines, Juneau
Bureau of Mines--Mineral Supply. Eastern Field Operation Center	DOI	Bureau of Mines, Pittsburg
Bureau of Mines--Mineral Supply. Western Field Operation Center	DOI	Bureau of Mines, Spokane
Center for Experiment Design and Data Analysis	NOAA (EDS)	D.C.
Chemical Kinetics Information Center	NBS (NSRDS), NAVY	Inst. for Materials Research (NBS)
Chemical Propulsion Information Agency (CPIA)	DOD (DSA, ARMY, NAVY, AF), NASA	Applied Physics Lab (John Hopkins)
Chemical Thermodynamics Data Center	NBS (NSRDS)	Institute for Materials Research
Controlled Fusion Atomic Data Center	AEC	Oak Ridge Nat'l Lab
Criticality Data Center	AEC	Union Carbide Corp. (ORNL)
Cryogenic Data Center	NBS (NSRDS), NASA, Amer. Gas Assoc.	Cryogenics Div. (NBS Boulder)

(Continued)

Figure 6-3 (cont.)
FEDERALLY SUPPORTED INFORMATION ANALYSIS CENTERS: 1974

Center	Sponsor(s)	Administration/Location
Crystal Data Center	NBS (NSRDS)	D.C.
Data Center for Atomic and Molecular Ionization Processes	NBS (NSRDS)	Institute for Basic Stardards
Data Center on Atomic Line Shapes and Shifts	NBS (NSRDS)	Optical Physics Div.
Data Collection and Processing Group	DOD (ONR), NSF, AEC, etc.	Scripps Inst. of Oceanography
Diatomic Molecule Spectra and Energy Levels Center	NBS (NSRDS)	Institute for Basic Standards
Diffusion in Metals Data Center	NBS (NSRDS)	Metallurgy Division
DOD Nuclear Information and Analysis Center (DASIAC)	DOD (DNA)	General Electric (Santa Barbara)
Earth Resources Observation System (EROS) Data Center	U.S.G.S.	Sioux Falls
Ecological Sciences Information Center	AEC	Oak Ridge Nat'l Lab
Electrolyte Data Center	NBS (NSRDS)	D.C.
Electromagnetic Metrology Information Center	NBS	Radio Standards Engineering Div.
Electronic Properties Information Center (EPIC)	DOD (DSA)	Thermophysical Properties Research Center (Purdue
Energy Information Center	NSF (RANN)	Oak Ridge National Lab
Environmental Information Analysis Center (EIAC)	AEC	Batelle Memorial Institute (Columbus)
Environmental Information Division	DOD (AF)	Air Training Command (Maxwell)
Environmental Mutagen Information Center	AEC, NCI, NIEHS	Oak Ridge Nat'l Lab
ERIC Clearinghouse for Junior Colleges	NIE (ERIC)	U.C.L.A.
ERIC Clearinghouse for Science, Mathematics, and Environmental Education	NIE (ERIC)	Ohio State Univ.
ERIC Clearinghouse for Social Studies/Social Science Education	NIE (ERIC)	Social Science Educational Consortium, Inc. (Boulder)

(Continued)

Figure 6.3 (cont.)
FEDERALLY SUPPORTED INFORMATION ANALYSIS CENTERS: 1974

Center	Sponsor(s)	Administration/Location
ERIC Clearinghouse in Career Education	NIE (ERIC)	Northern Illinois Univ.
ERIC Clearinghouse on Counseling and Personnel Services	NIE (ERIC)	University of Mich.
ERIC Clearinghouse on Early Childhood Education	NIE (ERIC)	University of Illinois
ERIC Clearinghouse on Educational Management	NIE (ERIC)	University of Oregon
ERIC Clearinghouse on Handicapped and Gifted Children	NIE (ERIC)	Council for Exceptional Children
ERIC Clearinghouse on Higher Education	NIE (ERIC)	The Geo. Washington Univ.
ERIC Clearinghouse on Information Resources	NIE (ERIC)	Stanford Univ.
ERIC Clearinghouse on Languages and Linguistics	NIE (ERIC)	Modern Language Assoc. of Amer.
ERIC Clearinghouse on Reading and Communication Skills	NIE (ERIC)	Nat'l Council of Teachers of Eng.
ERIC Clearinghouse on Rural Education and Small Schools	NIE (ERIC)	N. Mex. State Univ.
ERIC Clearinghouse on Teacher Education	NIE (ERIC)	Am. Assoc. of Coll. for Teachers of Eng.
ERIC Clearinghouse on Tests, Measurement and Evaluation	NIE (ERIC)	Evaluation Testing Serv.
ERIC Clearinghouse on the Disadvantaged	NIE (ERIC)	Columbia Univ.
Eutrophication Information Program	DOI (WRSIC), ARS, EPA	Univ. of Wisconsin
Gamma-Ray Spectrum Catalogue	AEC (NSRDS)	Nat'l Reactor Testing Station
Health and Safety Analysis Center	DOI	Mining Enforcement & Safety Admin.
High Pressure Data Center	NBS (NSRDS)	Brighom Young Univ.
Information Center for Hearing, Speech, and Disorders of Human Communication	NIH (NINDS)	Johns Hopkins Med. Inst.

(Continued)

Figure 6.3 (cont.)
FEDERALLY SUPPORTED INFORMATION ANALYSIS CENTERS: 1974

Center	Sponsor(s)	Administration/Location
Information Center for Internal Exposure	AEC	Oak Ridge Nat'l Lab
Infrared Information and Analysis Center (IRIA)	DOD, etc.	Environmental Research Inst. of Mich.
Institute of Polar Studies	NSF, AEC, U.S. Army	Ohio State Univ.
International Statistical Programs Center	AID	Bureau of the Census
Joint Institute for Laboratory Astrophysics Information Analysis Center	NBS (NSRDS)	Univ. of Colorado
LMFBR Fuel-Cladding Information Center	AEC	Hanford Engineering Devel. Lab
Machinability Data Center	DOD	Metcut Research Assoc. Inc.
Mechanical Properties Data Center	DOD	Belfour Stulen, Inc.
Metals and Ceramics Information Center (MCIC)	DOD	Batelle Memorial Inst. (Columbus)
Microwave Spectral Data Center	NBS (NSRDS)	Institute for Basic Standards
Molten Salts Data Center	NBS (NSRDS)	Rennselaer Polytechnic Inst.
National Center for Educational Statistics	DHEW	Office of Education
National Center for Health Statistics	DHEW (PHS)	Health Resources Admin.
National Clearinghouse for Mental Health Information	NIH (NIMH)	Div. of Scientific and Technical Info.
National Climatic Center (NCC)	NOAA (EDS)	Ashville
National Environmental Satellite Service	NOAA	D.C.
National Geophysical and Solar-Terrestrial Data Center	NOAA (EDS)	Boulder
National Meteorological Center (NMC)	NOAA	National Weather Service
National Neutron Cross Section Center	AEC (NSRDS)	Brookhaven Nat'l Lab

(Continued)

Figure 6.3 (cont.)
FEDERALLY SUPPORTED INFORMATION ANALYSIS CENTERS: 1974

Center	Sponsor(s)	Administration/Location
National Oceanographic Data Center	NOAA (EDS)	Boulder
National Space Science Data Center	NASA	Goddord Space Flight Center
Nondestructive Testing Information Analysis Center	DOD (Army)	Materials and Mechanics Center
Nuclear Data Project	AEC	Oak Ridge Nat'l Lab
Nuclear Safety Information Center	AEC	Oak Ridge Nat'l Lab
Phase Diagrams for Ceramists	NBS, American Ceramic Soc. Inc.	Inorganic Materials Div. (NBS)
Photonuclear Data Center	NBS (NSRDS)	Center for Radiation Research
Physical Data Group	AEC, DOD (DNA)	Lawrence Livermore Lab
Plastics Technical Evaluation Center (PLASTEC)	DOD (Army)	Piscatinny Arsenal
Poison Control Program	FDA	Bureau of Drugs
Primate Information Center	NIH (ARB)	Univ. of Washington
Program Analysis Branch, Drug Research and Development, Division of Cancer Treatment	NIH	National Cancer Inst.
Radiation Chemistry Data Center	NBS (NSRDS), AEC	Univ. of Notre Dame
Radiation Shielding Information Center	AEC, DOD (DNA)	Oak Ridge Nat'l Lab
Rare-Earth Information Center (RIC)	AEC, domestic and foreign Co's.	Iowa State Univ.
Reliability Analysis Center	DOD (AF)	Grifiss Air Force Base
Rock Properties Information Center (RPIC)	NSF (RANN)	Thermophysical Properties Research Center (Purdue)
Shock and Vibration Information Center	DOD, NASA	Naval Research Lab
Shock Wave Data Center	AEC (NSRDS)	Lawrence Livermore Lab
Suicide Prevention Center and the Institute for Studies of Self-Destructive Behavior	NIH (NIMH), Justice Dept., etc.	Los Angeles
Superconductive Materials Data Center	NBS (NSRDS)	General Electric R & D Center
Table of Isotopes Project	NBS (NSRDS), AEC	Lawrence Berkeley Lab

(Continued)

Figure 6.3 (cont.)

FEDERALLY SUPPORTED INFORMATION ANALYSIS CENTERS: 1974

Center	Sponsor(s)	Administration/Location
Tactical Technology Center (TACTEC)	DOD (ARPA)	Batelle Memorial Inst. (Columbus)
Technical Information Staff (HFX-25), Bureau of Radiological Health	FDA	Rockville
Texas A&M Thermodynamics Research Center (TRC)	NBS (NSRDS), API	Texas A&M
Thermophysical Properties Research Center (TPRC)	NBS (NSRDS), other Fed. & industry	Thermophysical Properties Research Center (Purdue)
Toxic Materials Information Center (TMIC)	NSF (RANN), AEC	Oak Ridge Nat'l Lab
Treaty Information Project	NSF	Univ. of Washington
USAF Environmental Technical Applications Center	USAF	Navy Yard Annex
X-Ray Attenuation Coefficient Information Center	NBS (NSRDS), DOD (DNA)	Center for Radiation Research (NBS)

SOURCE: King Research, Inc.: (Based on: Directory of Federally Supported Information Analysis Centers, National Referral Center, 1974.

Figure 8.1

NATIONAL STANDARD REFERENCE DATA SYSTEM CENTERS AND PROJECTS, APRIL 1975

Thermodynamic and Transport Properties

Chemical Thermodynamic [Data] Center[1]
Thermodynamic Data for Industrial and Municipal Incinerator Properties
Thermodynamic Data on Organic Compounds[1] [Thermodynamic Research Center] (Texas A&M)
Thermodynamic Properties of Polar Fluids
Cyrogenic Data Center[1] (NBS-Boulder)
PVT and Related Thermodynamic Properties of Ethylene
Thermodynamic Properties of Fluids in the Critical Region
Cyrogenic Fluid Mixture Properties (NBS-Boulder)
Fused Salt Electrochemistry[1] [Molten Salts Data Center] (RPI-Troy)
Aqueous Electrolyte Data Center[1]
High Pressure Data Center[1] (Brigham Young-Utah)
Fluid Transport Properties (NBS-Boulder)
Correlation of Thermophysical Property Data of Fluids (U.Md.-College Park)
Thermal Conductivity[1] [Thermophysical Properties Research Center] (Purdue-Indiana)
JANAF Thermochemical Tables (Dow-Michigan)
Contributions to the Data on Theoretical Metallurgy (Albany Metal. Res. Center-Oregon)
Thermochemistry for Steelmaking (MIT-Cambridge)

Atomic and Molecular Data

Adjustments to Fundamental Constants
Atomic Energy Levels Data Center[1]
[Atomic] Transition Probabilities Data Center[1]
Data Center on Atomic Line Shapes and Shifts[1]
Electromagnetic Cross Section Compilations[1] [X-ray Attenuation and Co-efficient Information Center]
Atomic Collision Cross Section Information Center (NBS-Boulder)
Data Center for Atomic and Molecular Ionization Processes[1]
Diatomic [Molecule] Spectra Data
Index to High Resolution Spectral Data
Microwave Spectral [Data Center] Tables[1]
Fundamental Vibration Frequencies of Molecules (U. of Tokyo)
NMR Data Compilation (Texas A&M)
Vibrational Force Field Constants for Polyethylene
AP144-TRC Selected Spectral Data (Texas A&M)

Chemical Kenetics Properties Data

Radiation Chemistry Data Center[1] (Notre Dame)
Chemical Kinetics Information Center[1]

See footnotes at end of figure (Continued)

Figure 8.1 (cont.) NATIONAL STANDARD REFERENCE DATA SYSTEM CENTERS AND PROJECTS, APRIL 1975

Solid State Properties

Superconductive Materials Data Center[1] (GE-Scenectady)
Cambridge Data Centre (Cambridge, England)
Data Compilation--Crystal Data[1] [Crystal Data Center]
Mossbauer Effect Data Index-MEDI (U.N.C.-Asheville)
Refractive Index Evaluation and Compilation
Alloy Data Center[1]
Diffusion in Metals Data Center[1]
Thermal Expansion Compilation[1,2]
Phase Diagrams for Ceramists[1,2]
Rare Earth Information Center[1,2] (Iowa State-Ames)

Nuclear Data

Table of Isotopes Project[1] (U.Cal.-Berkeley)
Berkeley Particle Data Center[1] (U.Cal.-Berkeley)
Compilation and Evaluation of Photonuclear Data[1] [Photonuclear Data Center]
Tables of Nuclear Spins and Moments
Nuclear Data Project[1,2] (Oak Ridge Natl. Lab.)
National Neutron Cross Section Center[1,2] (Brookhaven Natl. Lab.)
Physical Data Group[1,2] (U.Cal.-Livermore)

Mechanical Properties

Elastic Constant Data for Metal and Alloys (NBS-Boulder)
Central Surveys of Data Sources (Batelle-Columbus & GE-Schenectady)

---

[1]These centers also are listed as IAC's in 1974.

[2]These data centers are not directly under NSRDS program management, but supply relevant evaluated data.

NOTES: Entries in brackets [] indicate alternate titles found in other references to the same center.

Entries in parentheses () are locations if other than NBS, Gaithersburg.

SOURCE: Rossmassler, Stephen, Critical Evaluation of Data in the Physical Sciences-A Status Report on the National Standard Reference Data System, 1975.

Figure 9.1
ALPHABETIC LIST OF SYSTEMS IN THE ENVIRONMENTAL INFORMATION SYSTEM DIRECTORY: 1976

Acronym	System Name
AEROS	Aerometric & Emission Reporting System
APTIC	Air Pollution Technical Information Center
IPP	Air Quality Implementation Planning Program
CDS	Compliance Data System
CIS	*Contracts Information System
ECDBS	Emissions Control Data Base Syetem
EDS	EPA Energy Data System
ESPS	Epidemiological Studies Program System
ERSS	Establishment Registration Support System
FFF	*Federal Facility Budgetary Data System
-	Form 67 Retrieval System
-	*Formal Reporting System
-	Fuels Data Base System
GICS	*Grants Information & Control System
GLWQM	Great Lakes Water Quality Models
FMS	*Integrated Financial Management System
INVWAS	Inventory of Public Water Supplies
-	Lab Automation Project
LDMS	Laboratory Data Management System
LMS	Library Management System
NEDS	National Emissions Data System
NES	National Eutrophication Study
-	Noise File
PCS	*Permit Compliance System
PPS	*Personal Property System
DIPS	*Personnel/Payroll System
PARCS	Pesticides Analysis Retrieval & Control System
PEMS	Pesticides Enforcement Monitoring System
-	Pesticides Registration System
PRMS	Plans Review Management System
-	Population Studies System
-	Predictive Models for Fresh Water Ecosystems
RAPS	Regional Air Pollution Study
RMIS	*Resources Management Information System
SWIRS	Solid Waste Information Retrieval System
SPCCS	Spill Prevention Control & Counter Measure System
SRIS	*Standards & Regulations Information System
SOS	*State of the System Model
STORET	Storage & Retrieval for Water Quality Data
SAROAD	Storage & Retrieval for Aerometric Data
SEAS	Strategic Environmental Assessment System
-	*Survey of Needs for Municipal Waste Water Treatment Facilities
TADS	Technical Assistance Data System
TSSMS	*Time Sharing Services Management System
-	*Word Processing - Effluent Guidelines Division

*These systems generally deal with financial management and administration.

SOURCE: U.S. Environmental Protection Agency, Environmental Information Systems Directory, January 1976.

Figure 9.2
EXCERPTS FROM THE FEDERAL REGISTER, RULES AND REGULATIONS FOR NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION; ENVIRONMENTAL DATA

The Environmental Data Service is the Government's first major line organization created specifically to manage environmental data. EDS shall acquire, process, archive, analyze, and disseminate worldwide environmental (solid earth, marine, atmospheric, solar, and aeronomy) information, data, and products; guide applied research pertinent to the improvement of such services; provide relevant World Data Center facilities; coordinate national exchange activities in oceanic, climatological, geophysical, solar, and aeronomy data; and provide editorial, publishing, library, and related information services. In support of this EDS operates five centers: National Climatic Center, National Oceanographic Data Center, National Geophysical and Solar-Terrestrial Data Center, Center for Experiment Design and Data Analysis, and Environmental Science Information Center.

The National Climatic Center acquires, processes, archives, and disseminates climatological data and develops analytical and descriptive products to meet user requirements, and provides facilities for the World Data Center—A (Meteorology and Nuclear Radiation). It is the collection center and custodian of all United States weather records, the largest of the EDS Centers, and the largest climatic center in the world:

Climatic Data Available from NCC include:

(1) Hourly Surface Observations from Land Stations (ceiling, sky cover, visibility, precipitation or other weather phenomena, obstructions to vision, pressure, temperature, dew point, wind direction, wind speed, gustiness).

(2) Three-Hourly and Six-Hourly Surface Observations from Land Stations, Ocean Weather Stations, and Moving Ships (variable data content).

(3) Upper Air Observations (radiosondes, rawinsondes, rocketsondes, low-level soundings, pilot balloon winds, aircraft reports).

(4) Radar Observations (radar log sheets, radar scope photography).

(5) Satellite Data (vidicon pictures of earth and clouds, Earth Resources Technology Satellite (ERTS) imagery and other radiation data, derived products).

(6) Selected Maps and Charts (National Meteorological Center products).

(7) Derived and Summary Data (grid points, computer tabulations, digital summary data).

(8) Special Collections (Barbados Oceanographic and Meteorological Experiment meteorological data, Global Atmospheric Research Program basic data set, solar radiation data, many others).

The National Oceanographic Data Center acquires, processes, archives, and disseminates oceanographic data and develops analytical and descriptive products to meet user requirements, and provides facilities for the World Data Center—A Oceanographic Data Available from NODC include:

(1) Mechanical and expendable bathythermograph data in analog and digital form.

(2) Oceanographic station data for surface and serial depths, giving values of temperature, salinity, oxygen, inorganic phosphate, total phosphorus, nitrite-nitrogen, nitrate-nitrogen, silicate-silicon, and pH.

(3) Continuously recorded salinity-temperature-depth data in digital form.

(4) Surface current information obtained by using drift bottles or calculated from ship set and drift.

(5) Biological data, giving values of plankton standing crop, chlorophyll concentrations, and rates of primary productivity; also papers on marine biology.

The National Geophysical and Solar-Terrestrial Data Center acquires, processes, archives, evaluates, and disseminates solid earth and marine geophysical data as well as ionospheric, solar, and other space environment data; develops analytical, climatological and descriptive products to meet user requirements; and provides facilities for World Data Center—A (Geomagnetism, Gravity, Seismology, and Solar-Terrestrial Physics):

(a) Geophysical and solar-terrestrial data available from NGSDC include:

(1) *Marine geology and geophysics.* Bathymetric measurements; seismic reflection profiles; gravimetric measurements; geomagnetic total field measurements; and geological data, including data on heat flows, cores, samples, and sediments.

(2) *Solar-terrestrial physics.* Ionosphere data, including ionograms, frequency plots, riometer and field-strength strip charts, and tabulations; solar activity data; geomagnetic variation data, including magnetograms; auroral data; cosmic ray data; and airglow data.

(3) *Seismology.* Seismograms; accelerograms; digitized strong-motion accelerograms; earthquake data list (events since January 1900); earthquake data service, updates on a monthly basis.

(4) *Geomagnetic main field.* Magnetic survey data and secular-change data tables.

(Continued)

Figure 9.2 (cont.)
EXCERPTS FROM THE FEDERAL REGISTER, RULES AND REGULATIONS FOR NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION: ENVIRONMENTAL DATA

The Center for Experiment Design and Data Analysis provides service and support in data management and scientific analysis for large-scale environmental field research projects, and assists in the planning, design, and implementation of such projects to ensure that data needs are met:

(a) CEDDA is currently concerned with three major field projects:

(1) BOMEX—The Barbados Oceanographic and Meteorological Experiment. The complete set of data resulting from this project are available at the National Climatic Center.

(2) IFYGL—The International Field Year for the Great Lakes. Most of the data resulting from this project are available at the National Climatic Center.

(3) GATE—The Global Atmospheric Research Project (GARP) Atlantic Tropical Experiment. A set of basic data from this underway project is available from the National Climatic Center.

The Environmental Science Information Center develops policies for and provides editorial and publishing services to NOAA components; manages a central library system; provides functional guidance to other NOAA libraries; and develops and implements automated scientific information systems for NOAA and external use.

(a) ESIC issues a "NOAA Publications Announcement" several times a month.

(b) The NOAA libraries, run by ESIC, are open to the public for reference use.

(a) Since 1969, EDS has been building the Environmental Data Index (ENDEX). When fully operational (target date, 1978), ENDEX will provide convenient, rapid referral to existing NOAA, national, and global environmental science data files and sources, as well as documentation concerning their quality, quantity, and character. A complementary, literature-based system, Oceanic and Atmospheric Scientific Information System (OASIS), will provide a parallel subject-author-abstract referral service.

SOURCE: Federal Register, Vol. 39, No. 74, April 16, 1974.

Figure 9.3
1975 DIGITAL DATA HANDLING ACTIVITIES - USGS

Division/Office of USGS	Project Acronym	Project Name
Conservation Division	CLRIS	Coal Lease Reserve Information System
	RAS	Royalty Accounting System
	FRRE	Field Reservoir Reserve Estimate System
	GIAP	Geophysical Interpretive Aid System
	LPR	Lease Production and Revenue System
	OCSLAS	Outer Continental Shelf Lease Activity System
	OCS11	Outer Continental Shelf Order 11 System
	9-152	9-152 System
	PDB	Pipeline Data Base
	PMS	Pipeline Management System
	WHS	Well History File
	CORE	Computerized Onshore Record of Events
	IWRS	Individual Well Record System
Land Information Analysis Office	IDIS	Image Data Inquiry System
	DAL	Data Analysis Laboratory
	LUMAD	Land Use and Data Analysis Project
	CARETS	Central Atlantic Regional Ecological Test Site
	OPLUDE	Ozark Pilot Land Use Data Base
	CART/8	Cartographic On-Line Interactive Digitizing and Display Edit System
Geologic Division	DIP	Digital Image Processing
	LUNAR DATA	Global Synthesis of Large Lunar Data Arrays
	NCDS	National Coal Data System - Phase I
	EMS	Earthquake Strong Motion
	EMHDC	Earthquake Monitoring & Hypocenter Determination System
	LFGDS	Low Frequency Geophysical Data System
	ZCFS	Zip Code File System
	CCM	Computer Composit Mapping
	DLSD	Digital Landslide Susceptability Determination
	ACCMLUP	Application of Computer Cellular Mapping to Land Use Planning
	OGFDS	Oil & Gas Field Data System
	UIS	Uranium Information System
	OSIS	Oil Shale Information System
	WHCS	Well History Control System
	RADB	Radiometric Age Data Bank
	GA	Geometric Analysis
	GP	Gravity Projects
	RASS-II	Rock Analysis Storage System

(Continued)

Figure 9.3 (cont.)
1975 DIGITAL DATA HANDLING ACTIVITIES - USGS

Geologic Division (cont.)	CDS	Cruise Data System
	CRIB	Computerized Resources Information Bank
	GRASP	Geologic Retrieval and Synopsis Program
	IAEMOSRS	Interactive Alaska Economic Mineral Occurance Storage & Retrieval System
	CAMGGIM	Computer Assisted Methods to Generate Geologic Index Maps
	GDB	Geothermal Data Bank
	PA	Paleomagnetic Analysis
Topographic Division	DCDB	Digital Cartographic Data Base
	CGSR	Computer Generated Shaded Relief
(NCIC)	APSRS	Aerial Photography Summary Board System
(NCIC)	T-70X	Map and Chart System
(NCIC)	NCIC	EROS Explode
(NCIC)	APQF	Aerial Photography Quadrangle File
	CGN	Computerization of Geographic Names
Water Resources Division	WATSTORE	National Water Data Storage and Retrieval System
	SDRP	Satellite Data Relay Project
	NAWDEX	National Water Data Exchange

SOURCE: The International Geographical Union, Commission on Geographical Data Sensing and Processing, First Interim Report on Digital Geographic Data Handling Activities in the U.S. Geological Survey, March 1976.

Figure 10.1
SAMPLE ENTRIES FROM USDA DATA INVENTORY, 1974

2976-73
POULTRY & PRODUCTS: PRODUCTION.
Economic Research Service
Monthly summary; published regularly. Manual file.
15 99744
Broilers, Chickens, Eggs, Production, Ready-to-cook weight, Turkeys.
Abstract: Egg, chicken and turkey production from Statistical Reporting Service's annual reports. Broiler, other chicken and turkey production converted to ready-to-cook weight. Commodity Economics Division.

2977-73
POULTRY & PRODUCTS: SUPPLY & UTILIZATION.
Economic Research Service
Monthly summary; published regularly. Manual file.
15 99743
Chickens, Eggs, Exports, Imports, Military purchases, Stocks, Turkeys, Utilization.
Abstract: Data include factors making up total supplies available and disappearance of commodities for the year. Supply includes production, imports, and beginning and ending storage stocks. Disappearance includes exports and shipments to American territories, military use, USDA donations, and total and per-capita civilian consumption. In addition, the utilization of eggs include those going for hatchery purposes. Commodity Economics Division.

2978-73
POULTRY SLAUGHTERED UNDER FEDERAL INSPECTION.
Economic Research Service
Monthly summary; published regularly. Manual file.
15 99748
Broilers, Chickens, Poultry, Slaughter, Turkeys.
Abstract: Young chickens (primarily broilers) mature chickens, and turkeys slaughtered are reported by the Statistical Reporting Service. Number inspected, pounds liveweight, and pounds certified ready-to-cook weight. The certified ready-to-cook weight is used to estimate total slaughter by months and quarters during the year. Commodity Economics Division.

2979-73
PRELIMINARY FARM ALLOTMENTS-TOBACCO.
Agricultural Stabl Conserv Srv
Annual summary; available, not published. Manual file.
11 10918
Allotments (acreage), Tobacco.
Abstract: An acreage computed for each farm based on history of production during the immediately preceeding 5 years; summarized by kind of tobacco, by county, by state, and national totals. Used to apportion national acreage allotment to old farms. Tobacco and Peanut Division.

2980-73
PREMIUM AND INDEMNITY BY INSURANCE CLASS - ANALYSIS.
Federal Crop Insurance Corp
Annual point in time; published regularly. Integrated processing.
13 13409
Acreage, Counties, Crop insurance, Crop year, Crops.
Abstract: Magnetic tape file contains data by insurance classification within each county by crop, farming practice and crop year from 1948 to date. Includes premium, liability, indemnity, potential acres, insured acres, guarantee in bushels, pounds, etc., and premium rate in percent. Data Processing Branch, National Service Office.

SOURCE: U.S. Department of Agriculture, USDA Data Inventory, Volume 6, "Farm Income", 1974.

Figure 11.1

EXAMPLES OF DISTINCTIONS BETWEEN INSTITUTIONS, CENTERS AND FILES*

Entry	Designation
University of Colorado	INSTITUTION
Bureau of Governmental Research & Services	CENTER
Public Administration & Local Government	FILE
Census Summary Tape Processing Center	CENTER
Housing	FILE
Population	FILE
Colorado Business & Economic Data Bank	CENTER
Business Economic Data	FILE
Institute of Behavioral Science	CENTER
Davao Provence Urban: Rural Population Survey	FILE
Papago Indian Population Register	FILE
Travel Reference Center	CENTER
Travel, Tourism, & Recreation	FILE
. . . .	
Bureau of Labor Statistics (U.S.)	INSTITUTION
BLS Data Bank	CENTER
Area Wage Surveys	FILE
Consumer Price Index	FILE
Input-Output Matrix	FILE
Survey of Consumer Expenditures	FILE
Wholesale Price Index	FILE
Employment, Hours, & Earnings	FILE
Industry-Occupational Matrix	FILE
Job Openings & Labor Turnover	FILE
Labor Force	FILE

*Subtitles of files have been deleted.

SOURCE: Sessions, Vivian S., Directory of Data Bases In The Social And Behavioral Sciences, 1974.

APPENDIX C

BIBLIOGRAPHY

1. Ackoff, Russell L. Towards an Ideal Scientific Communication and Technology Transfer System (NSF-SIS 74-1041 AOI). Philadelphia, Pennsylvania: University of Pennsylvania, 1975.

2. Austin, Donald M., et al. An Overview of the LBL Socio-Economic-Environmental-Demographic Information System (LBL-3699). Berkeley, California: Lawrence Berkeley Laboratory, March 1975.

3. Becker, Joseph. A National Approach to Scientific and Technical Information in the United States. (NSF Contract C963). Los Angeles, California: Becker and Hayes, Inc., July 1976.

4. Borko, Harold. "The Analysis and Design of Information Systems." In: Key Papers in Information Science. Edited by Arthur W. Elias. Washington, D. C.: The American Society for Information Science, 1971.

5. Bulletin of the American Society for Information Science (special edition devoted to data). 1:7:3-40 (February 1975).

6. Caponio, Joseph F. and Marilyn C. Bracken. Selected Food and Agriculture Data Bases in the U.S.A. Washington, D. C.: U. S. Department of Agriculture, January 1973.

7. Carroll, Kenneth D. Survey of Scientific-Technical Tape Services (AIP ID 70-3). New York: American Institute of Physics, September 1970.

8. Catalog of Machine-Readable Records in the National Archives of the United States. Washington, D. C.: U. S. National Archives and Records Service, 1975 and 1976.

9. Chambaud, Serge. "Classification of Data Banks." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, July 1, 1976.

10. Committee on Data for Science and Technology, International Council of Scientific Unions. "Abstracts, Fifth International CODATA Conference." CODATA Bulletin, No. 18 (April 1976).

11. Committee on Data for Science and Technology, International Council of Scientific Unions. CODATA. Pamphlet. Paris, France: CODATA, 1976.

12. Committee on Data for Science and Technology, International Council of Scientific Unions. "Flagging and Tagging Data." CODATA Bulletin, No. 19 (June 1976).

13. Committee on Data for Science and Technology, International Council of Scientific Unions. "Geological Data Files." CODATA Bulletin, No. 8 (November 1972).

14. Committee on Data for Science and Technology, International Council of Scientific Unions. International Compendium of Numerical Data Projects. New York: Springer-Verlog, 1969.

15. Committee on Data for Science and Technology, International Council of Scientific Unions. "Man-Machine Communication in Scientific Data Handling." CODATA Bulletin, No. 15 (March 1975).

16. Committee on Data for Science and Technology, International Council of Scientific Unions. "The Presentation of Chemical Kinetics Data in the Primary Literature." CODATA Bulletin, No. 13 (December 1974).

17. Committee on Data for Science and Technology, International Council of Scientific Unions. "Proceedings of the Fourth International CODATA Conference on Generation, Compilation, Evaluation and Dissemination of Data for Science and Technology." CODATA Bulletin, No. 14 (June 1974).

18. Committee on Data for Science and Technology, International Council of Scientific Unions. "Proceedings of the Plenary Sessions, Fifth International CODATA Conference." CODATA Bulletin, No. 21 (October 1976).

19. Committee on Data for Science and Technology, International Council of Scientific Unions. "Study on the Problems of Accessibility and Dissemination of Data for Science and Technology." CODATA Bulletin, No. 16 (October 1975).

20. Committee on Scientific and Technical Information (COSATI). Directory of Federally Supported Information Analysis Centers. Washington, D. C.: COSATI, Federal Council for Science and Technology, April 1968.

21. Committee on Scientific and Technical Information (COSATI). Directory of Federally Supported Information Analysis Centers. Washington, D. C.: COSATI, Federal Council for Science and Technology, January 1970.

22. Committee on Scientific and Technical Information (COSATI). The Management of Information Analysis Centers. Washington, D. C.: COSATI, Federal Council for Science and Technology, January 1972.

23. Comptroller General of the United States. Federal Environmental Data Systems (B-177222). Washington, D. C.: U. S. General Accounting Office, November 1974.

24. Comptroller General of the United States. Improvements Still Needed in Federal Energy Data Collection, Analysis and Reporting. Washington, D. C.: U. S. General Accounting Office, June 1976.

25. Corridore, Michael C. Scientific and Technical Needs of Users or Potential Users of the DSA-Administered, DOD Information Analysis Centers. Alexandria, Virginia: U. S. Department of Defense, Defense Supply Agency, April 1976.

26. Crovello, Theodore J. "Botanical Data Banking." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 30, 1976.

27. Data Use and Access Laboratories. Directory of Products and Services. Arlington, Virginia: DuaLabs, September 1975.

28. David, Antoinette. "Useful Data for Engineers; Nature, Means of Access, Difficulties." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 28, 1976.

29. Department of Agriculture. USDA Data Inventory. Washington, D. C.: USDA, Office of Information Systems, 1974.

Volume I: Agricultural Exports, Foreign Agricultural Development.

Volume II: Rural Development, Environmental Improvement and Resource Development and Use.

Volume III: Support for Non-Federal Governments and Institutions, General Administration and Program Support.

Volume IV: Food and Nutrition, Consumer Services and Human Resource Development.

Volume V: Agricultural Production and Marketing Efficiency.

Volume VI: Farm Income.

30. Department of Agriculture. USDA Researchers' Needs for Data Processing. Washington, D. C.: U. S. Department of Agriculture, March 1973.

31. Duncombe, Raynor, et al. "International Cooperation of Centers for Ephemerides." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 30, 1976.

32. Elvove, Solomon. Astronomical Data in Machine Readable Form. (Circular No. 146). Washington, D. C.: U. S. Naval Observatory, April 1974.

33. Environmental Protection Agency. Environmental Information Systems Directory. Washington, D. C.: U. S. Environmental Protection Agency, January 1976.

34. Federal Council for Science and Technology. Directory of Federal Technology Transfer (NSF-75-402). Washington, D. C.: National Science Foundation, June 1975.

35. Federal Energy Administration, Directory of Federal Energy Data Sources, Computer Products and Recurring Publications. Washington, D.C.: Office of Policy and Analyses, 1976.

36. Federal Register, Volume 39, No. 74. Washington, D. C.: U. S. Government Printing Office, April 16, 1974.

37. Federation of Astronomical and Geophysical Services (FAGS). "Objectives and Activities." ICSU Information Bulletin, No. 27 (1972).

38. Feller, Barbara A. Directory of Federal Agency Education Data Tapes (NCES 76-206). Washington, D.C.: National Center for Education Statistics, U.S. Department of Health, Education, and Welfare, 1976.

39. Freeman, James E. and Van A. Wente. "Evaluation of Numerical Data Flagging and Tagging in a Real World Aerospace Environment." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 28, 1976.

40. Fried, John B. "Basis On-Line Retrieval and Analysis of Large Numeric Data Bases." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, July 1, 1976.

41. FuJaros, R. G. et al. Data Terminal Technology, Present and Future: Volume 2. Technology Forecasting and Assessment. Ottawa, Canada: Department of Communications/Communications Research Center, April 1975.

42. General Accounting Office. 1976 Congressional Sourcebook Series. Washington, D.C.: Office of Program Analysis, U.S. GAO, June 1976.

Volume 1: Recurring Reports to the Congress.

Volume 2: Federal Program Evaluations.

Volume 3: Federal Information Sources and Systems.

43. Gerson, Gordon M. "Why not a Federal Data Administrator?" Paper presented at ASIS Conference, October 1976. Reston, Virginia: Defense Communications Engineering Center, August 1976.

44. Hellen, Rachelle S. et al. "The Mass-Spectral Search System." Paper presented at the Fifth International CODATA Conference, Boulder, Colorado, July 1, 1976.

45. Herner, Saul and Matthew J. Vellocci, eds. Selected Federal Computer-Based Information Systems. Washington, D.C.: Information Resources Press, 1972.

46. International Geographical Union, Commission on Geographical Data Sensing and Processing. First Interim Report on Digital Geographic Data Handling Activities in the U.S. Geological Survey. Washington, D.C.: U.S. Geological Survey, Department of the Interior, Revised March 1976.

47. Jackson, John E. and James I. Vette. The Orbiting Geophysical Observatories Program Summary. Washington, D.C.: National Space Science Data Center, NASA, 1975.

48. Kehiaian, Henry V. "The International Data Series." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 28, 1976.

49. Kenny, Charles T. and Jerome T. Maddock. DDC 10 Year Requirements and Planning Study Survey Plan. Prepared for Defense Documentation Center. Philadelphia, Pennsylvania: Auerbach Associates, Inc., August 1975.

50. Kertesz, Francois. "The Information Center Concept." In: Key Papers in Information Science. Edited by Arthur W. Elias. Washington, D. C.: The American Society for Information Science, 1971.

51. King, D. W., D. D. McDonald, N. K. Roderer, B. L. Wood. Statistical Indicators of Scientific and Technical Communication (1960-1980); Volume I: A Summary Report (NSF C-878). Rockville, Md.: King Research, Inc., 1976.

52. King, D. W., F. W. Lancaster, D. D. McDonald, N. K. Roderer, B. L. Wood. Statistical Indicators of Scientific and Technical Communication (1960-1980); Volume II: A Research Report (NSF C-878). Rockville, Md.: King Research, Inc., 1976.

53. King, Donald, et al. Statistical Indicators of Scientific and Technical Communication: 1977 Edition (NSF C-878). Rockville, Md.: King Research, Inc., 1977.

54. King, Joseph H. Interplanetary Magnetic Field Data Book. Greenbelt, Maryland: National Space Science Data Center, April 1975.

55. King, Joseph H. and Margaret L. King. Catalog of Particles and Fields Data 1966-1973. Greenbelt, Maryland: National Space Science Data Center, June 1975.

56. Kruzas, Anthony T. Encyclopedia of Information Systems and Services, Second International Edition. Ann Arbor, Michigan: Edwards Brothers, 1974.

57. Lerner, Rita G. "Data Tagging in Physics." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 28, 1976.

58. Lide, David R., Jr. "Cooperation Between Professional Societies and a Government Agency: The Journal of Physical and Chemical Reference Data," IEEE Transactions on Professional Communication 18:3:127-129 (September 1975).

59. Lockheed Information Systems. Dialog Information Retrieval Service. Brochure. Palo Alto, California: Lockheed Information Systems, April 1976.

60. Luedke, James A., Cuadra, Carlos, ed. "Numeric Data Bases" in Annual Review of Information Science and Technology, Volume 12. Washington, D.C.: American Society for Information Science, (forthcoming) 1977.

61. Maddock, Jerome T., et al. DDC 10 Year Requirements and Planning Study; Volume I: Executive Summary. Prepared for Defense Documentation Center. Philadelphia, Pennsylvania: Auerbach Associates, Inc., June 1976.

62. Maddock, Jerome T., et al. DDC 10 Year Requirements and Planning Study; Volume II: Technical Discussion, Bibliography, and Glossary. Prepared for Defense Documentation Center. Philadelphia, Pennsylvania: Auerbach Associates, Inc., June 1976.

63. Mason, Robert M. Development of Cost Benefit Methodology for Scientific and Technical Information Communication and Application to Information Analysis Centers (NSF Contract SIS 75-12741). Atlanta, Georgia: Metrics, Inc., 1976.

64. Morse, Larry E. "Some Information Management Problems Raised by the International Nature of Systematic Biology Data." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 30, 1976.

65. National Academy of Sciences, Geophysics Research Board. An Assessment of the Impact of World Data Centers on Geophysics. Washington, D. C.: National Academy of Sciences, March 1975.

66. National Association for State Information Systems. 1975-1976 NASIS Report-Information Systems Technology in State Government. Lexington, Kentucky: NASIS, 1976.

67. National Bureau of Standards. Critical Evaluation of Data in the Physical Sciences - A Status Report on the National Standard Reference Data System. (NBS Technical Note No. 881). Washington, D. C.: U. S. Department of Commerce, April 1975.

68. National Bureau of Standards. Materials Information Programs: An Interagency Review of Federal Agency Activities on Technical Information about Materials (NBS Special Publication 463). Washington, D. C.: U. S. Department of Commerce, January 1977.

69. National Bureau of Standards. "A Mechanized Information Services Catalog." (NBS Technical Note 714). Washington, D. C.: U. S. Department of Commerce, February 1972.

70. National Bureau of Standards. NSRDS News. (April/May 1976 and other issues).

71. National Bureau of Standards. "Six Data Base Management Systems: Feature Analysis and User Experiences." (NBS Technical Note 887). Washington, D. C.: U. S. Department of Commerce, March 1975.

72. National Bureau of Standards. "A Technical Index of Information Systems." (NBS Technical Note 819). Washington, D. C.: U. S. Department of Commerce, March 1974.

73. National Cartographic Information Center. Newsletter No. 5. Reston, Virginia: U. S. Geological Survey, Summer/Fall 1976.

74. National Oceanic and Atmospheric Administration. Endex, Oasis, NOAA Environmental Data Key. Pamphlet. Washington, D. C.: U. S. Department of Commerce, 1975.

75. National Oceanic and Atmospheric Administration. The National Climatic Center. Pamphlet. Washington, D. C.: U. S. Department of Commerce, 1975.

76. National Oceanic and Atmospheric Administration. The National Geophysical and Solar-Terrestrial Data Center. Pamphlet. Washington, D. C.: U. S. Department of Commerce, 1975.

77. National Oceanic and Atmospheric Administration. The National Oceanographic Data Center. Pamphlet. Washington, D. C.: U. S. Department of Commerce, 1975.

78. National Oceanic and Atmospheric Administration Environmental Data Service. Pamphlet. Washington, D. C.: The Environmental Science Information Center, U. S. Department of Commerce, 1975.

79. National Referral Center. Directory of Federally Supported Information Analysis Centers. Washington, D. C.: Library of Congress, National Referral Center, 1974.

80. National Referral Center. A Directory of Information Resources in the United States. Washington, D. C.: Library of Congress, Science and Technology Division.

Biological Sciences, 1972.
Federal Government, 1974.
General Interest, 1969.
Physical Sciences, Engineering, 1971.
Social Sciences, Revised Edition, 1973.
Water, 1966.

81. National Research Council. Geophysical Data Centers: Impact of Data Intensive Programs. Washington, D. C.: National Academy of Sciences, 1976.

82. The National Science and Technology Policy, Organization, and Priorities Act of 1976. P.L. 94-282. (May 1976).

83. National Science Foundation. Federal Funds for Research Development and other Scientific Activities Fiscal Years 1970, 1971, 1972 (NSF 71-35). Washington, D. C.: National Science Foundation, 1971 (and subsequent years).

84. National Science Foundation. Federal Scientific and Technical Communications Activities: 1974 Progress Report (NSF-SIS 75-1). Washington, D. C.: Office of Science Information Service, National Science Foundation, June 1975.

85. National Science Foundation. Federal Scientific and Technical Communications Activities: 1975 Progress Report (NSF 76-25). Washington, D. C.: Division of Science Information, National Science Foundation, June 1976.

86. National Space Science Data Center and World Data Center A for Rockets and Satellites. NSSDC and WDC-A-R&S Document Availability and Distribution Services. Greenbelt, Maryland: National Aeronautics and Space Administration, May 1974.

87. National Technical Information Service. Directory of Computerized Data Files and Related Software. Washington, D. C.: U. S. Department of Commerce, March 1974.

88. National Technical Information Service. A Directory of Computerized Data Files, Software and Related Technical Reports. Springfield, Virginia: U. S. Department of Commerce, NTIS, 1976.

89. National Technical Information Service. Software & Data Files. Springfield, Virginia: U. S. Department of Commerce, December 1973.

90. National Technical Information Service. Special Technology Group Catalog, 1976. Springfield, Virginia: U. S. Department of Commerce, 1976.

91. Office of Management and Budget. "A List of the 100 Active Public-Use Reports Having the Largest Man-Hour Burden Repetitive" (As of March 31, 1975). Unpublished Report No. CA17. Washington, D. C.: OMB, December 1975.

92. Office of Management and Budget. Report of the Federal Mapping Task Force on Mapping, Charting, Geodesy and Surveying: July 1973. Washington, D. C.: U. S. Government Printing Office, 1973.

93. Office of Management and Budget. "Special Analysis F; Principal Federal Statistical Programs." Reprint from: Budget of the United States Government, 1974. Washington, D. C.: U. S. Government Printing Office, January 1973.

94. Office of Management and Budget. "Special Analysis G; Principal Federal Statistical Programs." Reprint from: Budget of the United States Government, 1977. Washington, D. C.: U. S. Government Printing Office, January 1976.

95. Olson, Richard and F. Glenn Goff. "Development and Application of a Regional Environmental Data Base for Southeastern United States." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 30, 1976.

96. Panel on World Data Centres, International Council of Scientific Unions. Third Consolidated Guide to International Data Exchange through the World Data Centres. Washington, D. C.: ICSU Panel on World Data Centres, December 1973.

97. Pearlstein, Sol. "The Information Center as a Link Between Basic and Applied Research." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 30, 1976.

98. Phillips. S. L., et al. National Geothermal Information Resource. Berkeley, California: Lawrence Berkeley Laboratories, 1975.

99. PMI Facilities Management Corporation. National Space Science Data Center 1974 Year-End Statistics (NAS 5-11985). Silver Spring, Md.: P.M.I., 1975.

100. Porat, Marc U. "The Information Sector: Definition & Measurement" Comments prepared for the American Association for the Advancement of Science. Boston, Massachusetts; February 23, 1976.

101. Reader, William A. and Jon E. Updike. A Study of Data Management Policies in Federal Scientific Data Centers. Washington, D. C.: National Archives and Records Service, June 1975.

102. Roderer, N. K. Statistical Indicators of Scientific and Technical Communication (1960-1980); Volume VII: Worldwide Indicators (NSF C-787). Rockville, Md.: King Research, Inc., 1977.

103. Rodgers, G. I. Energy Information Activities at the FEA. Washington, D. C.: Federal Energy Administration, May 1976.

104. Rosencrantz, Gerald J. "Machine-Readable Records of the Federal Government: The Archives Program for Preservation and Access." Review of Public Use Data 2:1:2-11 (January 1974).

105. Rowe, Judith. "The Use and Misuse of Government Produced Statistical Data Files." RQ (Reference & Adult Services Division) 14:3:201-203 (Spring 1975).

106. Rowe, Judith and Mary Ryan. "Library Service from Numerical Data Bases: The 1970 Census as a Paradigm." College and Research Libraries, pp. 7-15 (January 1974).

107. Science Communication, Inc. Study of Scientific and Technical Data Activities in the United States -- Volume I: Plan for Study and Implementation of National Data Systems Concepts (ARPA Order: 892). Prepared for the Committee on Scientific and Technical Information, (COSATI), Federal Council for Science and Technology. Washington, D. C.: Science Communications, Incorporation, December 1968.

108. Science Communication, Inc. Study of Scientific and Technical Data Activities in the United States -- Volume II: Preliminary Census of Scientific and Technical Data Activities - Parts A & B (ARPA Order: 892). Prepared for the Committee on Scientific and Technical Information, (COSATI), Federal Council for Science and Technology. Washington, D. C.: Science Communications, Incorporation, December 1968.

109. Science Communication, Inc. Study of Scientific and Technical Data Activities in the United States -- Volume II: Preliminary Census of Scientific and Technical Data Activities - Part C (ARPA Order: 892). Prepared for the Committee on Scientific and Technical Information, (COSATI), Federal Council for Science and Technology. Washington, D. C.: Science Communications, Incorporation, December 1968.

110. Sessions, Vivian S. (ed). Directory of Data Bases in the Social and Behavioral Sciences. New York: Science Associates/International, Inc., 1974.

111. Shapiro, Arthur, et al. An Overview of the National Space Science Data Center Standard Information Retrieval System (SIRS). Greenbelt, Maryland: National Space Science Data Center, NASA, May 1974.

112. Shapley, Alan S. "The U. S. Environmental Data Service--A Multidisciplinary Multipurpose National Service in the Geosciences with Relations to the ICSU World Data Center System." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 30, 1976.

113. Sherrod, John. "Review of Federal Legislation Establishing Numerical Data Bases." Paper presented at the Fifth International CODATA Conference. Boulder, Colorado, June 28, 1976.

114. Shuman, Jack N. "The Data Revolution: Its Policy and Implications." Bulletin of the American Society for Information Science. 1:7 (February 1975).

115. Shuman, Jack N. The Policy Implications of Interactive Computer Systems. Washington, D. C.: Georgetown University, 1975 (unpublished).

116. Skolnik, Herman, editor. Journal of Chemical Information and Computer Sciences. (Special issue devoted to large data bases) 15:1:1-64 (February 1975).

117. Tomberg, Alex (ed.). Data Bases in Europe. Second edition. London: ASLIB, 1976.

118. Tomberg, Alex. "Networks for Data Base Sharing," NEWSIDIC, No. 19 (Spring 1976).

119. Tomberg, Alex. "Scientific Information Networks." Paper presented at the Electronics Conference. Paris: March-April 1977.

120. Tressel, George W. and Brown, Patricia L. A Critical Review of Research Related to the Economics of the Scientific and Technical Information Industry. (SIS 74-10449 A01) Columbus, Ohio: Battelle Columbus Laboratories, March 1975.

121. United Nations Educational, Scientific and Cultural Organization, CODATA Task Group on Accessibility and Dissemination of Data. UNISIST: Study on the Problems of Accessibility and Dissemination of Data for Science and Technology. Paris: UNESCO, 1973.

122. United Nations Educational, Scientific and Cultural Organization. UNISIST: Study Report on the Feasibility of a World Science Information System. Paris: UNESCO, 1971.

123. U. S. Congress. House. Commission on Information and Facilities. Inventory of Information Resources and Services Available to the U. S. House of Representatives. 94th Congress, 2nd session, December 1976.

124. U. S. Congress. Senate. Committee on Commerce. Ocean Data Resources. 93rd Congress, 2nd session, March 1975. Pursuant to S.Res.222.

125. U. S. Geological Survey. "Request for Proposal to Write Technical Specifications for the National Cartographic Information Center" (RFP-5871). Reston, Virginia: U. S. Department of the Interior, December 1976.

126. United States Government Manual 1974-75. Washington, D. C.: Office of the Federal Register, National Archives and Records Service, 1974.

127. United States Government Manual 1976-77. Washington, D. C.: Office of the Federal Register, National Archives and Records Service, 1976.

128. Veazie, Walter H. and Thomas F. Connolly. The Marketing of Information Analysis Center Services and Products. Washington, D. C.: American Society for Information Science, June 1971.

129. Weinberg, Alvin M. "Scientific Communication." In: Key Papers in Information Science. Edited by Arthur W. Elias. Washington, D. C.: The American Society for Information Science, 1971.

130. Weisman, Herman M. Information Systems, Services, and Centers. New York: John Wiley and Sons, Inc., 1972.

ADDENDUM

131. LaParte, Leon. "Endex/Oasis - NOAA Environmental Data and Information Referral Service" in SSIE Science Newsletter, Vol. V #9. Smithsonian Sciences Information Exchange, Inc., August 1976.

132. Hardy, William and Patrick Hughes. "The Environmental Data Connection", in Environmental Data Services, May 1974.

133. Office of National Standard Reference Data. Good Science, Bad Data. Pamphlet. Washington, D. C.: U.S. Department of Commerce, National Bureau of Standards.

134. Office of National Standard Reference Data. The Other Face of the Measurement Data Base. Pamphlet. Washington, D. C.: U.S. Department of Commerce, National Bureau of Standards.