FROM THE ARCHIVES

Geophysical Data Centers: Impact of Data-Intensive Programs (1976)

Pages
37

Size
7 x 10

ISBN
0309335906

Geophysical Data Panel; Committee on Data Interchange and Data Centers; Geophysics Research Board; Assembly of Mathematical and Physical Sciences; National Research Council

Geophysical Data Centers: Impact of Data-Intensive Programs

Geophysical Data Panel
Committee on Data Interchange and Data Centers
Geophysics Research Board
Assembly of Mathematical and Physical Sciences

NATIONAL ACADEMY OF SCIENCES
Washington, D. C. 1976

NOTICE

The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the Councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the Committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The Geophysics Research Board is pleased to acknowledge the assistance of the National Science Foundation and National Oceanographic and Atmospheric Administration, Department of Commerce, whose continuing support of the Committee on Data Interchange and Data Centers has made this report possible.

Available from
Geophysics Research Board
National Research Council
2101 Constitution Avenue
Washington, D.C. 20418

GEOPHYSICAL DATA PANEL

Carl H. Savit, Western Geophysical Company, *Chairman*
Sidney A. Bowhill, University of Illinois
Lewis M. Branscomb, International Business Machines Corporation
Thomas O. Haig, University of Wisconsin
Carl Kisslinger, University of Colorado
Marvin Minsky, Massachusetts Institute of Technology
William J. Schmitz, Jr., Woods Hole Oceanographic Institution

Donald C. Shapero, *Executive Secretary*

INTRODUCTION

At its meeting in May 1974, the Geophysics Research Board (GRB) Committee on Data Interchange and Data Centers (CDIDC) recommended that a special study be undertaken regarding new demands that are being placed on World Data Centers (WDCs) and associated national data centers by large-scale geophysics programs. This recommendation was strongly supported by the GRB. Thomas Malone, chairman of the GRB, then asked Carl Savit to visit data centers in the United States and to prepare a letter report on his findings. The main text of that letter report follows:

> Geophysical data of the more traditional types falling within the responsibilities of the Centers are handled efficiently and economically. Data management is excellent from the point of view of filing and retrieval procedures, data editing, compilation, data conditioning, and dissemination. Response to requests appears to be uniquely well tailored to the needs and requirements of the user community.
>
> In part, the efficiency and effectiveness of the operation may be attributed to the virtual identity of the communities of users and suppliers of data. The data are submitted in a format that is readily convertible to useful forms by well-established techniques. Editing and taxonomic functions are within the realms of expertise of the technicians in the Centers and are primarily carried out manually or with the aid of simple analog devices.
>
> In contrast to this well-ordered and stable data flow, new forms of data are beginning to appear. These new forms are primarily characterized by being several orders of magnitude more voluminous in information content than the old types of data. They generally appear on magnetic tape in one of a variety of digital (or sometimes analog) formats. All steps performed on the old types of data will necessarily have to be performed in some fashion on the new types of data. These new data will require editing, formatting, condensation, summarization, excerpting, and other derivative processes.
>
> To effect these requirements, a profound change will have to be made in the organization, staffing, and funding of the World Data Centers. Access will be required to powerful and large-scale digital computers. Special-purpose processors will probably have to be designed, built, and procured. Staffs of programmers will have to be employed and trained, and scientific guidance will have to be provided to these programmers.

Storage facilities suitable to magnetic tapes will have to be provided. Printout facilities under the control of EDP equipment will have to be available, and inventories of specialized materials will have to be maintained.

All the above procedures will require very substantial initial funding and a considerable increase in continued funding levels. The approach to data-user and data-supplier costs and charges will have to be re-examined. The supplier community and the user community will certainly be, at least in part, disjunct. It is quite probable that a large fraction of the user community will be interested only in summarized, generalized, or abstracted data, and will be unable to pay any substantial part of the costs of obtaining such derived data from the massive files of raw data.

A number of other problems and caveats stemming from the advent of the newer types of data are visible to one extent or another at the present time. Every aspect of the new data forms and new data content that have already begun to appear imply an exponential rise in data quantity, with the concomitant implication of substantial impacts on administration, staffing, and funding.

Subsequently, the Geophysical Data Panel (GDP) of the CDIDC was formed, with Mr. Savit as chairman. The charge to the GDP was to amplify and substantiate the findings of the letter report.

The resulting report consists of two parts. In the first part the findings and recommendations of the Panel are set forth and discussed. The operation of some typical World Data Centers and associated national data centers is described, and the impact of new data forms and quantities on them is considered. The second part contains the results of a survey of digital data quantities from past, ongoing, and planned large-scale geophysics programs.

The growth of the task of data centers can be measured in several dimensions:

1. Budget for large-scale geophysics programs
2. Number of programs
3. Number of nations involved in programs
4. The extent and resolution of geographical coverage
5. The amount of data generated

The figures pertaining to data quantities in the second part of the report serve to emphasize the importance of the main issues to which this report is addressed:

1. The increasing expectations of benefit and growing investment by domestic and foreign government agencies and national and international scientific bodies in large-scale geophysics programs;

2. The urgency of ensuring the effective dissemination of the knowledge gained in such programs so that those benefits will be realized;

3. The central role of World Data Centers and their associated national data centers in accomplishing the dissemination of that knowledge.

PART I
PANEL REPORT

Findings and Recommendations

1. *The facilities of national and world data centers must be expanded and improved if they are to continue to carry out their assigned functions.*

2. *The planners of large-scale geophysics programs are making increasing use of advanced sensor and computer technology to meet scientific objectives. The data sets so generated are orders of magnitude larger than those of programs using traditional methods of observation. This growth can be expected to continue, creating increasing demands to store data and to provide data products.*

3. *Dealing with these data sets will create a need for large-scale computers, special-purpose processors, and information storage and retrieval systems at data centers. Such systems require sophisticated knowledge and procedures to permit access to specific data sets. Staff to provide programming services and scientific advice will be necessary both in the storage and access phases.*

4. *Summarized, generalized, or abstracted data will be required by much of the user community. The costs of deriving such data products from raw data must be defrayed by means that will be compatible with the resources of both the generator of the data and the user community.*

5. *In the decisions leading to acquisition of large data sets in major geophysical programs, a strategy should be formulated for data handling and archiving. Cost tradeoffs between acquisition and handling and archiving considerations should be an integral part of planning. The strategy should include provisions for any thinning or compression of data. The starting point in planning should be consideration of the ultimate beneficiaries of the information. Program proposals should include the strategy for and costs of archiving data and making them available to users.*

6. *General criteria should be formulated for data acquisition, taking into account, among other things, format, data quality, and handling requirements.*

7. *Decisions to thin or discard geophysical data may be a necessary consequence of limited resources. Such decisions entail an obligation on the part of data centers to*

(a) *give notice in a timely way to the international scientific community of the intent to destroy data so as not to foreclose action to preserve the data;*

(b) *consider the relative economics of retention, thinning, and compression of data;*

(c) *preserve geophysically significant data samples or subsets;*

(d) *give priority to retention of data that are not reproducible.*

8. *Regarding the operation of the WDC system:*

(a) *The principles and practices of operation as set down in the* Guide to International Data Exchange through the World Data Centres *should be adhered to.*

(b) *World Data Centers should be consulted in the planning of programs that will lead to new demands on the WDC system.*

9. *Both technologically developed and less developed countries have found the World Data Centers to be a valuable resource. Future planning regarding World Data Centers should be directed toward increasing their responsiveness to the needs of less developed countries.*

Some Typical Data Centers

The data centers discussed below are treated in more detail in a CDIDC report.[1] The Appendix lists present holdings of the data centers, including traditional data forms. Conversion factors for information content of a few representative forms are given in that Appendix. Services related to traditional forms continue to be important.

Services related to digital data are considered below. Each of the centers discussed consists of a World Data Center and an associated national data center. Most of the resources of the centers are devoted to the operation of the national data centers. The World Data Centers serve special functions as described in the *Guide to International Data Exchange through the World Data Centres* and depend heavily on the facilities and resources of the associated national data center.

The National Geophysical and Solar Terrestrial Data Center (NGSDC) has on file approximately 600 reels of magnetic tape containing digital data. A CDC-6600 computer is used for batch computation. An XDS-940 time-sharing system is also available. Computers are used on a part-time basis only. Approximately 20 percent of the budget is devoted to computer services.

The National Climatic Center (NCC) has on file approximately 77,000 reels of magnetic tape containing digital data. A UNIVAC series 70/45 computer is used full time, and four other computers (Raytheon 703, Daconic 3630, Hewlett-Packard 9830, and Wang 700) are used on a part-time basis. NCC, in cooperation with the Air Force, is in the process of procurement of a new computer. Approximately 25 percent of the budget is devoted to computer services. Data from the Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment (GATE) are available from the Computer Products Branch of the National Climatic Center and are archived by WDC-A for Meteorology, which is collocated with NCC. Data from the First GARP Global Experiment (FGGE) will be similarly handled.

The National Oceanographic Data Center (NODC) has on file approximately 60 reels of magnetic tape containing digital data. An IBM 360/65 computer is used on a part-time basis. Approximately 20 percent of the budget is devoted to computer services.

The National Space Science Data Center (NSSDC) has on file approximately 41,000 reels of magnetic tape containing digital data. An IBM 7094 is used full time, and IBM 360/75 and 360/91 computers are used on a part-time basis. Approximately 15 percent of the budget is devoted to computer use.

Each of these centers also has large files of data recorded in conventional formats, such as photographic film, punched cards, and printed data sheets. Each center is experienced in meeting requests for data in these forms. Generally speaking, most of their resources are devoted to meeting such requests, a task that they discharge effectively (see Appendix).

New Demands on Data Centers

LESS DEVELOPED COUNTRIES

Although the technologically developed countries are the biggest users of the WDCs, both in terms of volume of data and numbers of requests, users of the WDCs, including regular users, are to be found in virtually all the countries that have participated in the international geophysical programs. Moreover, responses to the survey that led to the report *An Assessment of the Impact of the World Data Centers on Geophysics*[1] indicated that the World Data Centers have been particularly important to scientists in the less developed countries because of the lack of national data or scientific centers in those countries. In the future, the World Data Centers are likely to take on a sharply increased importance in connection with the less developed countries. The beginnings of this are already evident. Plans for the Global Environmental Monitoring System call for use of the WDCs. The proposal has been made (Law of the Sea Conference) that data obtained on oceanographic vessels in the territorial waters of other countries be made available through the World Data Centers. The World Data Centers are apparently accepted throughout the world as an effective even-handed mechanism. Some of the future programs are likely to be of specific interest to the less developed countries. Future planning regarding the WDCs should give careful attention to efforts to make them more responsive to the needs of the less developed countries in connection with basic science and practical problems. In particular, there will be a need for more highly processed data products designed to meet the needs of users lacking high-technology resources.

DATA VOLUMES

The data centers may be expected to encounter difficulties in meeting requests for data products derived from the kind of data-intensive programs described in Part II. The most apparent source of difficulty is the contrast between the volume of digital data generated by these programs and the volumes with which the data centers are accustomed to dealing. To take an example, only one (there will be five) of the synchronous meteorological satellites (SMS) in the First GARP Global Experiment (FGGE) will generate about 5×10^{13} bits of digitally encoded visual information over the 18-month FGGE period. To record these data on conventional 2400-foot reels of 1600 BPI computer tape would require 200,000 to 400,000 reels of tape. World Data Centers and the associated national data centers generally are able to deal with relatively small numbers of tape reels of this type as discussed above. But they clearly will not be able to deal effectively with a large number of programs each of which may generate volumes of digital data several orders of magnitude larger than those to which they are accustomed without some major adjustments in organization, staffing, and funding. Several other aspects of data handling may also present problems.

ASPECTS OF DATA HANDLING

RETENTION

It is not clear that it is practical to attempt to retain raw data volumes of the magnitude discussed above. Assuming that the data can be compressed with the aid of new technology, access time becomes a limiting factor. If we assume a processing speed of about 10,000,000 bits per second, it would require 75 days to process the output of one FGGE SMS. With present technology, then, it would be difficult to make such data bases available for the exclusive use of any but a small set of well-funded users. This difficulty might be partly overcome by scheduling times when specific subsets of extremely large data bases would be put on-line for simultaneous access by a large number of users. It should be noted that commercial facilities do exist that store and provide access to even larger magnetic tape quantities.

Computer technology is advancing rapidly, and what may appear today to be an almost unmanageable problem may seem much less formidable in the near future. There is strong evidence that performance of computers is growing exponentially.[2] Should data volumes also grow exponentially, then at any given future time, the volume of newly acquired data will dwarf that of old data. Should such circumstances obtain, discarding old data will have a negligible effect on required storage capacity. Therefore, if economic factors compel a decision to discard data, this decision will necessarily have to be directed primarily at the new data sets.

The acquisition and preservation of data from large geophysics programs should be systematized and accomplished in a coherent way. General criteria should be formulated for the acquisition of data taking into account, among other things, format and data quality. Some provision should be made for dealing with the various formats peculiar to each discipline, but some limits will have to be imposed on the variety of formats with which the data centers will be required to deal. These objectives were met very effectively in the planning of GATE.

However undesirable it may be to thin or discard data, limited resources may nonetheless force a decision to do so. Such a decision should be undertaken only with the greatest circumspection. The international scientific community should be notified of the intent to destroy data so that the opportunity to take action to preserve them will not be lost and so that the community will have the opportunity to make provisions for retaining data of special significance. Examples of such might include data associated with unique geophysical events or phenomena, unprocessed data in which reduction processes destroy information, or data that cannot be taken again simply because geophysical events do not repeat themselves. A decision to discard data can be costly in the sense that the expense of storing data is generally less than that of acquiring it.

INDEXING

Data are organized and indexed according to the objectives or exigencies of the program generating them. For example, for experiments covering a large area, it is often necessary to employ an array of field recording equipment generating a set of tapes each of which contains data in chronological order associated with a particular location. It may be desirable to reorganize these tapes into a form in which all data coded according to location and corresponding to a given time are grouped together. In the most general sense, a datum corresponds to a point in a multidimensional coordinate and parameter space. Different indexing schemes allow one to examine different sequences of subspaces. It is advantageous to anticipate indexing schemes that will be needed by secondary users and to put the entire data set on line to generate in one operation special data sets correspondingly organized. Anticipating such needs is important, since the largest aggregate demand for data comes from secondary users. Data center staff must not only have mastery of disciplinary fields but must also maintain close contact with the data-using communities.

LEVELS OF PROCESSING

The degree of processing of data can be characterized in the following way:

I. Raw data
II. Data processed to yield physical parameters
III. Data interpolated according to a standardized array of points in space and time

This characterization was developed in the GATE program, and we use it generally. The higher the level of processing, the more compact is the data set. One criterion that has been proposed for data retention and that would reduce the data volume flowing into data centers considerably would be to retain only the level III data set. As can be seen from the Appendix, for the FGGE, the level III data set is about three orders of magnitude smaller than the level I data set. Obviously, however, such a compression must involve considerable loss of information. For example, the processing of level I spectral data to produce temperature profiles is not a reversible process.

But spectral data have diverse uses. Thus, information of potential use might be lost if only the level III data set were retained. At least some subsets of raw data should, in general, be retained, not only for their information content *per se* but also to facilitate tracing of processing. Use for secondary purposes of data from scientific programs will in general require reprocessing. Such reprocessing requires trained personnel and a hierarchy of computing capabilities.

STORAGE

The amounts of data that may be expected to be retained in data centers are such that, to provide prompt access, a hierarchy of mass storage systems will be required. Ideally, the storage systems would have sufficient capacity and flexibility so that specially processed data sets could be produced automatically.

With advanced technology, storage of even the FGGE level I data set would not be particularly difficult. A volumetric compression factor of almost 200 is possible with videotape data storage. Use of videotape would reduce the number of tapes necessary to store the FGGE geostationary satellite data to about 5000 tapes for five satellites for the entire FGGE period. For purposes of comparison, the National Climatic Center has available for purchase 5986 SMS-1 tapes containing visual (VIS) and infrared (IR) imagery on a conventional 1600 BPI tape format. Thus videotape storage should make storage of the FGGE level I data set no more difficult than that of the GATE SMS VIS and IR imagery. However, as mentioned before, access is limited by processing rates for FGGE level I data. It would be straightforward for data centers to make available individual videotapes, but a user requiring a special subset of the data set would have to have access to the entire set. If the electronic data-processing staff and facilities necessary to prepare special data sets are not provided by data centers, some users would have to purchase the entire data set, most of which they might not need, and process it themselves to accomplish their objectives. That is already the case for the GATE VIS and IR data sets recorded on magnetic tape. Should this trend continue, data centers will increasingly become repositories for data that are effectively inaccessible to the majority of potential users.

An alternative to storage in data centers is to rely on informal arrangements with outside centers (e.g., universities) to maintain data sets. Data centers would then provide referral services. Without formal arrangements, data centers may lose control over the data sets, and it can be anticipated that as programs in outside centers are phased out, data sets would be discarded.

ACCESSIBILITY

If data centers are to ensure the preservation of information and availability of data products from large programs in accessible forms, advisory services and equipment will have to be provided to assist users in extracting needed data products. For many users, interactive facilities are desirable in order to allow optimization of the data products that they receive. Planning of data handling must encompass archiving and retrieval. Data-acquisition systems and format specifications should be designed not only to make possible efficient processing of data for the purpose of extracting and organizing physical parameters but also to maximize accessibility of information to users. Realization of the benefits of large-scale geophysics programs depends on the effectiveness with which information is shared, which in turn depends on the ability of data centers to disseminate information. Program planning must take into account not only experimental design and data analysis but also archiving and retrieval. The entire path of information flow, beginning with data acquisition in the experiment, including data analysis, and ending with archiving and retrieval, must be considered with the overall goal of providing maximum access to information. Careful experimental design and data analysis are necessary but not sufficient to accomplish that goal.

COSTS

Data centers will need additional support, either direct or indirect, for new activities and services associated with increased demands on them. Direct funding must provide part of the additional support, but it may be unrealistic to expect these demands to be met entirely by increases in internal budgets of data centers and charges to users. Therefore, in the interest of ensuring the dissemination of information, a significant part of the cost of depositing data in data centers in an accessible form must be provided for in the budget of the program that generates those data. Only marginal costs can be realistically charged to the user, and it is idle to suppose that, without additional funding from the initial data-gathering program, the data centers will be able effectively to provide access to the data. The amount involved may be 5 to 20 percent of the cost of the research. Experimenters should provide for these costs in the same way that they budget for publication costs. Funding agencies should examine proposals in the light of these guidelines, and if an experimenter has failed to budget to provide appropriate access to the information that he intends to gather, he should be advised to examine the tradeoffs between data acquisition and data access more realistically. Program planners should not be allowed to avoid responsibility for the costs of data storage and retrieval. If too great a fraction of the cost of obtaining access to information is forced on users, they will simply ignore the data in the data centers, and the expected benefits of large-scale programs will not be fully realized.

REFERENCES

1. *An Assessment of the Impact of World Data Centers on Geophysics*, Report of the CDIDC of the GRB (National Academy of Sciences, Washington, D.C., 1975).
2. C. Starr and R. Rudman, "Parameters of Technological Growth," Science *182*, 358 (1973).

PART II
SURVEY OF DATA QUANTITIES FOR GEOPHYSICAL PROGRAMS

INTRODUCTION

In Part II, estimated data quantities in bits for a sampling of programs from the four main branches of geophysics are tabulated. The nature of the program is described briefly. Where possible, separate quantities are stated for raw data (Level I), data processed to yield physical parameters (Level II), and data interpolated according to a standardized array of points in space and time (Level III). It must be borne in mind that these data quantities are *approximate*. For programs planned to take place at future dates, projected estimates of data quantities must be used. In some cases approximate conversions from number of tapes to number of bits on the basis of recording format have been made. Such conversions do not take into account the presence of record gaps in which no information is recorded and tend to overestimate data quantities by roughly a factor of 2.

ATMOSPHERE

BOMEX

One objective of BOMEX was to serve as a pilot experiment to develop the data-acquisition and data-handling capabilities that proposed large-area experiments, planned as part of the Global Atmospheric Research Program, would require. The central scientific objective was to evaluate the sea-air interaction process. The observations were completed in three months in 1969.

Measurement Category	Data Quantity in Bits Level II
1. Ship Operations and Navigation Data Set	2×10^8
2. Boom Surface Meteorological Data Set	7×10^8
3. Marine Meteorological Observations and Surface--Pressure--Marine Micro-barogram Data Set	2×10^8
4. Rawinsonde and Radiometer-sonde Data Set	2×10^9
5. Boundary Layer Instrument Package (BLIP) Data Set	2×10^9
6. Salinity-Temperature-Depth (STD) Data Set	4×10^9
7. Surface Radar Data Set	3×10^8
8. Aircraft Data Set	2×10^9
9. Dropsonde Data Set	4×10^8
TOTAL	1×10^{10}

GARP

GARP is a program for studying atmospheric physical processes governing transient, large-scale fluctuations and the statistical properties of general circulation. The objectives are to improve the accuracy and range of weather forecasting and to develop a better understanding of the physical basis of climate. The program has two components: numerical modeling and experimental studies to test the validity of these models. Two of the experimental subprograms on which data volume estimates are available are described below. There are several other experimental subprograms and related projects.

GATE

GATE was the first major international experiment of GARP. Its main objective was to clarify the relationship between convective activity and synoptic-scale systems in tropical regions sufficiently to allow accurate representation of the effects of convection in numerical prediction programs. It took place over the tropical Atlantic Ocean and adjacent land areas in a period of four months in 1974.

GATE provided a data quantity of approximately 2×10^{12} bits of Level I (raw) data. Of this amount, approximately 1.4×10^{12} bits were obtained by the United States (roughly half of which are retained in the National Holdings at NCC) and less than 1×10^{12} bits by non-U.S. sources.

Approximately 2×10^{11} bits of nationally processed and validated (Level II) data are being produced by the ten GATE National Processing Centers (about half U.S. and half non-U.S.), for subsequent archiving in World Data Centers in the United States and the Soviet Union.

Approximately 10^{11} bits of internationally processed and validated (Level II) data are being produced by the five GATE Subprogram Data Centers (about 4×10^{10} by the Convection Subprogram Data Center in the United States and 5×10^{10} by the other SDC's) for subsequent archiving in World Data Centers in the United States and the Soviet Union.

Measurement Category	Data Quantity in Bits	
	Level I	Level II
Data from GATE National Processing Centers		
1. WWW Land Stations	(Included in SSDC Below)	
2. A-Scale Ships		1.8×10^{9}
3. A/B-Scale Ships		3.8×10^{10}
4. B-Scale Ships		1.4×10^{11}
5. Roving Oceanographic Ships and Oceanographic Buoys	(Included in OSDC below)	
6. GATE Aircraft		1.8×10^{10}
7. Satellites (not including high-resolution radiance data)		2.2×10^{9}
8. Commercial Aircraft and Voluntary Ships	(Included in SSDC below)	
Data from GATE Subprogram Data Centers		
9. Boundary-Layer Subprogram Data Center (BSDC)		1.5×10^{10}
10. Convection Subprogram Data Center (CSDC)		3.9×10^{10}
11. Oceanographic Subprogram Data Center (OSDC)		2.6×10^{10}
12. Radiation Subprogram Data Center (RSDC)		1.8×10^{9}
13. Synoptic-Scale Subprogram Data Center (SSDC)		6.5×10^{9}
Selected Data Associated with National Holdings		
14. ATS-3 VIS Satellite Imagery	1.8×10^{11}	
15. SMS IR and VIS Satellite Imagery	8.8×10^{11}	
16. U.S. Ship Systems (not including upper air and radar)	1.2×10^{11}	
17. GARP Level II Data Set (NMC Global Synoptic Data Set)		9.4×10^{9}
18. NESS Sea-Surface Temperature Analysis		2.0×10^{9}
TOTAL	1.2×10^{12}	3.0×10^{11}

FGGE

The FGGE will attempt to find the limits of deterministic forecasting of atmospheric phenomena. It will test the accuracy with which existing global models of the atmosphere simulate climate. The buildup phase will begin on September 1, 1977. Data collection will start on January 1, 1978. The operational phase will begin on September 1, 1978, and will last until accepted data have been obtained for 12 consecutive months. It will provide a global data set more complete than any in the history of meteorology.

Measurement Category	Data Quantity in Bits Level I	Level II	Level III
1. SMS satellite imagery	3×10^{14}		
2. Polar-orbiting satellite imagery	3×10^{11}		
3. Raw radiance data	3×10^{11}		
4. Surface-based WWW system			
Land stations		2×10^{9}	
Ships		4×10^{8}	
Aircraft		1×10^{8}	
Radio		2×10^{8}	
Radiowind		2×10^{8}	
5. Space-based WWW system			
Polar-orbiting satellite		5×10^{10}	
Sea-surface temperatures		2×10^{8}	
Geostationary satellites		5×10^{8}	
6. Special observing systems			
Carrier balloons plus ships		4×10^{8}	
Constant-level balloons		3×10^{8}	
Buoys		4×10^{7}	
7. Experimental satellites (based on Nimbus-G equipment)			
Temperature profiles		2×10^{9}	
Radiation balance		1×10^{8}	
Precipitation over oceans, soil mositure, sea-ice extent and melting	6×10^{10}	1×10^{8}	
8. All categories			1×10^{11}
TOTAL	3×10^{14}	6×10^{10}	1×10^{11}

SOLAR-TERRESTRIAL

IMS

The IMS is an international cooperative enterprise to be conducted in 1976-1978. Its main objective is to elucidate the physical behavior of plasma in the earth's magnetic field. Its operational basis consists of a plan of coordinated observations from spacecraft, ground-based facilities, aircraft, balloons, and research rockets. Listed below are data quantities for a few specific IMS programs and estimates for the total ground and space data.

Measurement Category	Data Quantity in Bits	
	Level I	Level II
1. Specific Ground-Based Arrays:		
Magnetometer Network		2×10^{10}
AFS MAGAF Network[a]	10^{10}	
Meridian Scanning Photometer (one year)	5×10^{10}	10^{9}
2. Estimated Total Ground Arrays	$\sim 10^{12}$	$\sim 5 \times 10^{11}$
3. Specific Spacecraft Experiments:		
Magnetic Measurements	3×10^{11}	
4. Estimated Total Spacecraft Data	$\sim 10^{13}$	$\sim 5 \times 10^{12}$

[a]Data from MAGAF Network may be contributed to IMS.

AFGWC

The AFGWC processes several kinds of geophysical data received from DMSP Satellites. Quantities are for a one-year period.

Measurement Category	Data Quantity in Bits Level I
1. Auroral Imagery	7×10^{11}
2. Precipitating Electron Data	2×10^{9}

Solid Earth

MARINE SEISMIC REFLECTION DATA

Large quantitites of marine seismic reflection data will become available in 1975-1978. These multichannel CDP data are collected by USGS and other institutions.

Measurement Category	Data Quantity in Bits Level I
1. Atlantic CDP	4×10^{11}
2. Alaska/Pacific CDP	3×10^{12}

SDAC

The SDAC has a large number of digital tapes on file. Some digital seismic data collected from the ALPA and the Montana LASA are on permanent file.

Measurement Category	Data Quantity in Bits Level I
1. Short-Period Data (LASA)	2×10^{12}
2. Long-Period Data (LASA + ALPA)	2×10^{12}
3. All Digital Tapes	10^{13}

SRO

The USGS is developing a Seismic Research Observatory system using digital data recording. About 13 stations are planned for installation throughout the world. The data quantity is a rough estimate for one year of operation of the entire system.

Measurement Category	Data Quantity in Bits Level I
1. SRO System	10^{11}

Ocean

IDOE

The IDOE is a long-term international cooperative program of ocean research and exploration intended to contribute to preservation of the quality of the ocean environment and to determine potential resources of the ocean floor. U.S. intention to participate was announced in 1969. In 1972, further objectives of providing knowledge of biological processes necessary to manage living marine resources and improving mechanisms for worldwide exchange of marine data were added. There are four subprograms: Environmental Quality, Environmental Forecasting, Seabed Assessment, and Living Resources. Each subprogram consists of several ongoing projects. The estimates below combine the data quantities for all these projects with the exception of POLYMODE, which is listed separately.

Measurement Category	Data Quantity in Bits Level II
1. All Categories, 1970-1975	$\sim 10^{10}$
2. All Categories, 1975-1980	$\sim 2 \times 10^{10}$

POLYMODE

POLYMODE is a large-scale midocean dynamics experiment based on results of the Polygon (U.S.S.R.) and MODE-I (U.S.A. and U.K.) experiments. It will consist of an intensive initial field program lasting at least one year, with the objective of advancing understanding of mesocale midoceanic eddies and their role in general ocean circulation. Accomplishing this objective is one step toward the development of a physically correct ocean model that will make possible oceanological forecasting and the construction of a model coupling ocean and atmospheric dynamics. The intensive data collection period is to be 1977-1978.

Measurement Category	Data Quantity in Bits	
	Level I	Level II
1. Floats (U.S.)	10^{11}	2×10^{7}
2. Current Meters (U.S.)		$10^{9} \pm 2$
3. Density (U.S.)		2×10^{9}
4. Other		10^{8}
5. Moored Array (U.S.S.R.)		$10^{9} \pm 2$

IGOSS

The IGOSS is a system of real-time services for oceanography, including regional analysis and predictions. Data quantity is from 1972 to the present period. It is expanding to include international sightings of oil slicks and tar balls and to analyze ocean-water samples for hydrocarbons.

Measurement Category	Data Quantity in Bits Level II
1. Bathythermographs	10^{8}
2. Oil Slick Sightings (to 1980)	10^{7}
3. Tar Ball Sightings (to 1980)	10^{6}
4. Hydrocarbon Analyses (to 1980)	10^{4}

MESA NEW YORK BIGHT

The objective of the MESA New York Bight project is to establish baselines describing physical, biological, and chemical processes in the bight and to develop understanding of these processes in order to aid in assessment of the impact of future management decisions. Data quantity is from 1973 to the present on only one of the several measurement categories for this project.

Measurement Category	Data Quantity in Bits Level II
1. Current Meters	10^{9}

NEGOA

The NEGOA is a program of marine environmental data acquisition and analysis along sections of the Alaskan outer continental shelf now under consideration for leasing. Data quantity is from 1975 to the present.

Measurement Category	Data Quantity in Bits Level II
1. Current Meters	10^9
2. Density	10^8
3. Zooplankton	10^8

IFYGL

The IFYGL, an intensive field study of Lake Ontario and its drainage basin, was a joint Canadian and U.S. contribution to the International Hydrological Decade (1965-1974). It was a coordinated program of physical, chemical, and biological research, with the objective of improving understanding of Lake Ontario and its basin. Data were collected from April 1972 through March 1973, with some data systems operating through July 1973.

Measurement Category	Data Quantity in Bits Level II
1. U.S. Nonsatellite Data	1.7×10^{11}
2. U.S. Satellite Data	1.0×10^{10}
3. Canadian Data	6.2×10^9

Abbreviations

AFGWC	Air Force Global Weather Central
AFS	Air Force Cambridge Research Laboratories
ALPA	Alaska Long Period Array
ARPA	Advanced Research Projects Agency
ATS	Applications Technology Satellite
BOMEX	Barbados Oceanographic and Meteorolgical Experiment
CDP	Common Depth Point
DMSP	Defense Meteorological Satellite Program
FGGE	First GARP Global Experiment
GARP	Global Atmospheric Research Program
GATE	GARP Atlantic Tropical Experiment
IDOE	International Decade of Ocean Exploration
IFYGL	International Field Year for the Great Lakes
IGOSS	Integrated Global Ocean Station System
IMS	International Magnetospheric Study
IR	infrared
LASA	Large Aperture Seismic Array
MAGAF	Air Force Ground Based Magnetometer Network
MESA	Marine Ecosystems Analysis
MODE	Mid Ocean Dynamics Experiment
NCC	National Climatic Center
NEGOA	Northeast Gulf of Alaska
NESS	National Environmental Satellite Service
NGSDC	National Geophysical and Solar-Terrestrial Data Center
NMC	National Meteorological Center
NODC	National Oceanographic Data Center
NSSDC	National Space Science Data Center
SDAC	Seismic Data Analysis Center (Va.)
SMS	Synchronous Meteorological Satellite
SRO	Seismic Research Observatory
USGS	U.S. Geological Survey
VIS	visual
WWW	World Weather Watch

APPENDIX: PRESENT HOLDINGS OF DATA CENTERS*

CONVERSIONS

Form of Record	Number of Bits
2400-foot, 9-track, 1600 BPI magnetic tape reel	3.7×10^8
SMS satellite picture	2×10^9
Traditional ionogram	2×10^7 (unassessed) 3×10^4 (assessed echoes)

NCC

Form	Quantity
Sheets of original ms and autographic records	76,000,000
100-ft reels of microfilm of original ms and autographic records	86,000
Microfilmed images of punched cards	377,000,000
2400-ft reels of magnetic tape (includes approximately 3000 reels of satellite data)	77,000
100-ft reels of radar film (35- and 16-mm)	19,800
100-ft reels of satellite film	7,000
10" × 10" negatives of satellite film	175,000
Microfiche of original records and data publications	58,000
Unpublished data tabulations and summaries	24,300
Back issues of single climatological data publications (does not include shelf stock)	194,000

*Of four major national geophysical data centers and the associated branches of World Data Center A.

NGSDC

Form	Quantity
35-mm film	1.2×10^7 ft
Magnetic tape	600 reels
Punched cards	1.6×10^6
Paper data sheets, prints, or publications	5.1×10^3 cu ft

NSSDC

Form	Quantity
Microfilm, 100-foot reels	27,519
Microfiche	18,198
Digital magnetic tapes	40,931
Photographic films	
5 inch width, linear feet	45,800
9.5 inch, linear feet	17,500
70 mm, linear feet	509,663
16 mm, linear feet	119,276
35 mm, linear feet	759,976
4" × 5" each	12,291
8" × 10" each	6,289
16" × 20" each	93
20" × 24" each	8,005
Hardcopy, sheets	188,612

NODC

Form	Quantity
Station data geosort	549,793 stations
BT data digitized geosort	782,204 observations
XBT data digitized geosort	159,182 observations
BT analog prints	820,000 observations
Biological data, digital	13,000 stations
Biological papers stored for retrieval	22,917
Surface current data	4,175,324 observations
Total number of digital magnetic tapes	60