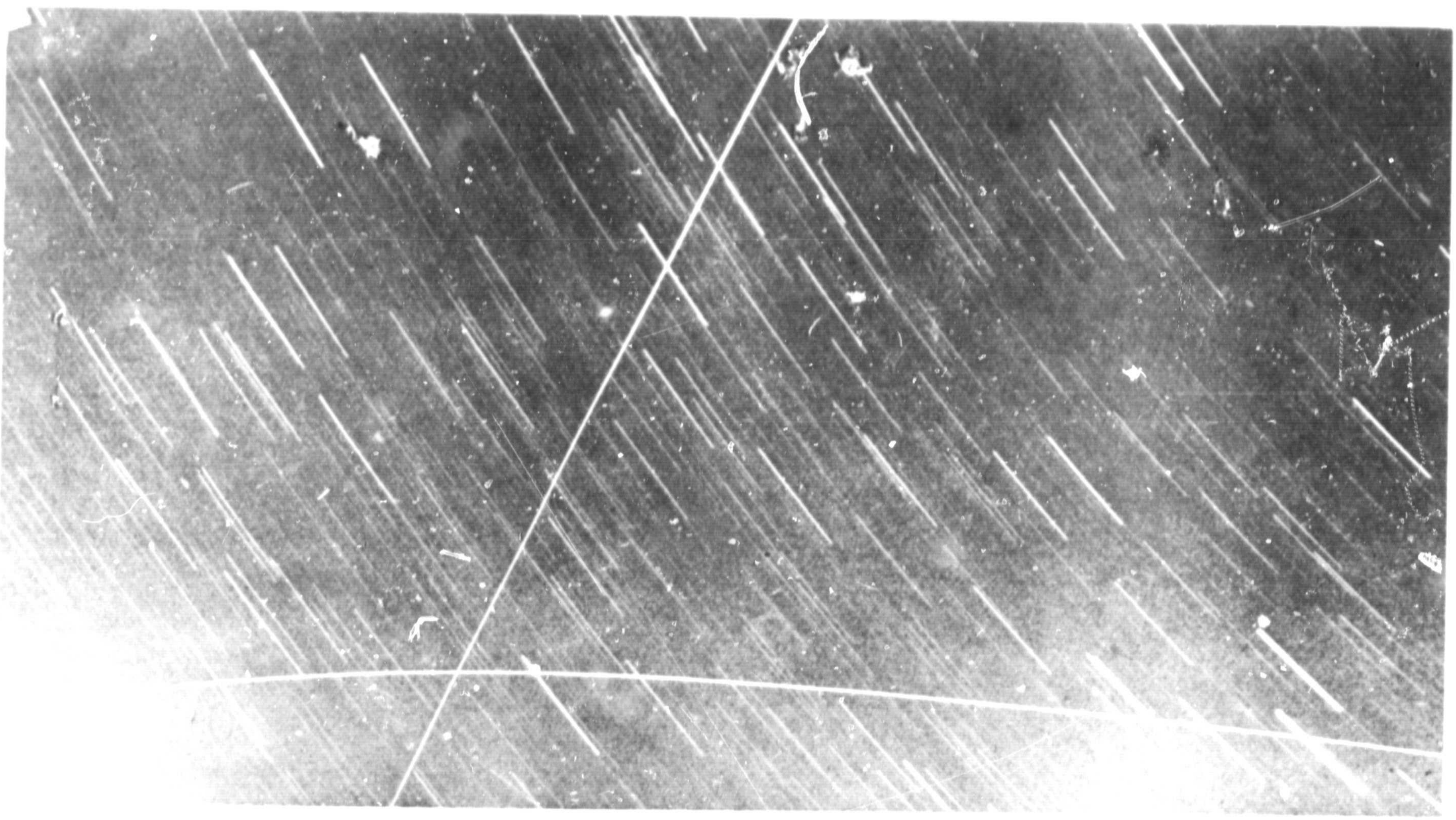


THE OPERATION OF THE NATIONAL SPACE SCIENCE DATA CENTER

NSSDC 67-41

OCTOBER 1967

NATIONAL SPACE SCIENCE DATA CENTER

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • GODDARD SPACE FLIGHT CENTER, GREENBELT, MD.

THE OPERATION OF THE

NATIONAL SPACE SCIENCE DATA CENTER

NSSDC 67-41

OCTOBER 1967

FOREWORD

The need to prepare and preserve space science data for use and further study is vital to many scientists and engineers. To help fill this need, NASA's National Space Science Data Center collects data from all space science flight experiments. At the present time, scientific satellites alone, exclusive of picture-taking, are generating 8×10^{11} bits per year. The goal of the Data Center effort is maximum dissemination and use of these data.

This document has been prepared to inform interested parties about (a) the forms in which data are collected, (b) the present activities and use of the Data Center, and (c) some of our future plans.

James I. Vette

Dr. James I. Vette
Director
National Space Science Data Center

PRECEDING PAGE BLANK NOT FILMED.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

THE OPERATION OF THE NATIONAL SPACE SCIENCE DATA CENTER

BACKGROUND

Before describing the operation of the National Space Science Data Center, it may be useful to take a look at the data involved. It has become apparent in the last few years that satellite measurements produce a tremendous amount of data to be processed and analyzed. The need for and use of this data by groups other than the scientists and engineers involved in the specific experiments are very great. Our understanding of the complex natural phenomena observed in space is far from complete. Consequently, the preservation of the basic data for further study in the future is vital to many scientists. Some other uses for this data will be presented later. The fact that satellite experiments are expensive and require 3-5 years between instrument design and return of data means that the information acquired by these space techniques is an expensive commodity. It is imperative, then, that means be established to promote maximum dissemination and utilization of the data.

To get on common ground in describing these data, it should be pointed out that there are several different categories in which the various space projects can be placed. Scientific experiments measure those quantities which will lead to a general understanding of the natural phenomena. On the other hand, engineering measurements are vital to the advancement of techniques and procedures which lead to more sophisticated, reliable, and useful spacecraft and systems. The applications satellites for communication, navigation, weather, and other operational needs form a significant portion of the present effort. Then there are the biomedical experiments which are conducted as part of the manned program and are important for achieving suitable ecological systems.

As can be seen, the great diversity of information that comes from satellites requires a dissemination system that can be responsive to the user community. In this context, the National Space Science Data Center, or NSSDC for short, is concerned with most of the scientific data obtained from space probes, satellites, sounding rockets, stratospheric balloons, and high-altitude aircraft. It is responsible for the acquisition, organization, storage, retrieval, announcement, and dissemination of this information. NSSDC was established by NASA in 1965 as an extension of a local data center activity at the Goddard Space Flight Center. By restricting the activities to the scientific class of space data, the proper lines of communication with the generators as well as the users of the data can be maintained. Encompassing too diverse a group would complicate staffing and management of the operation. The Data Center is a central facility within this restricted framework. The scope of operations is national in the sense that the space science data generated by all agencies and research groups are collected. The resulting data bank is available to all requesters.

Bearing in mind the goal of maximum dissemination and utilization of these data, this paper describes the general flow of space science data, discusses the activities of NSSDC, and examines some Data Center problems and their possible solutions.

THE FLOW OF SPACE SCIENCE INFORMATION

First, it may be instructive to examine the origin and overall flow of this kind of data and information in order to clarify the role that the Data Center is fulfilling. A good starting point is with an orbiting spacecraft as a typical generator of space science data and then following the sequential process usually required to translate this data into useful information which can be understood by others. The diagrams given in Figures 1 and 2 illustrate the steps involved.

All of the data obtained by various instruments are telemetered to the receiving station in a coded form and are recorded on analog magnetic tape. In some cases specialized equipment at the receiving stations converts the data directly to a meaningful form. This is true, for example, in the Surveyor Program where the primary pictures are made from the data as they are received at the ground station. It should be noted that the raw data on the analog tapes include the outputs from many different experimental sensors as well as information about the performance of spacecraft systems. Examples would include power supplies, on-board clocks, command modes, transmitters, solar panels, spin rate, and, in some cases, measurements by aspect or orientation devices. At this point the data may be degraded by telemetry noise, improper satellite system or experiment performance, or by ground station operations.

The analog tapes are then sent to a central processing center where they are converted in most cases into digital tapes. They are quality checked and the noise is removed where possible. It is here the spacecraft performance and time measurements are decoded and converted to standard units. The resultant edited tape is often called the master digital data tape. In special cases the scientific satellite programs generate data which are made available to the scientific community immediately; many of the picture-taking satellites such as Ranger, Surveyor, and Lunar Orbiter fall in this category. However, the majority of experiments are performed by a principal investigator, who has sole responsibility for the design, construction, and calibration of the instrument and for the initial interpretation of the data. For a period of several years he controls the availability of the data to others and can have exclusive rights to it.

FIGURE 1
DIAGRAM OF SATELLITE DATA FLOW FROM ORBITING SPACECRAFT THROUGH CENTRAL PROCESSING FACILITIES

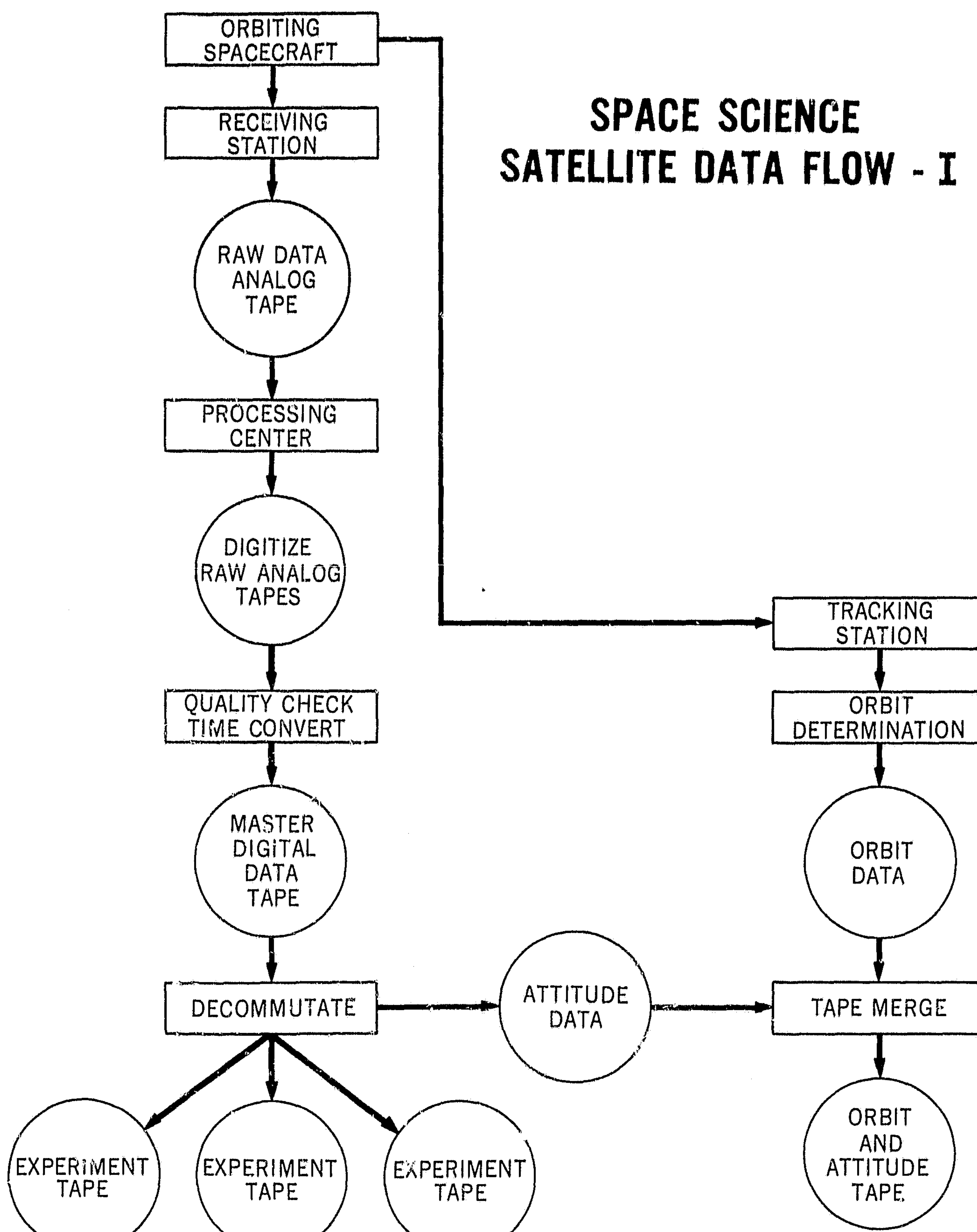
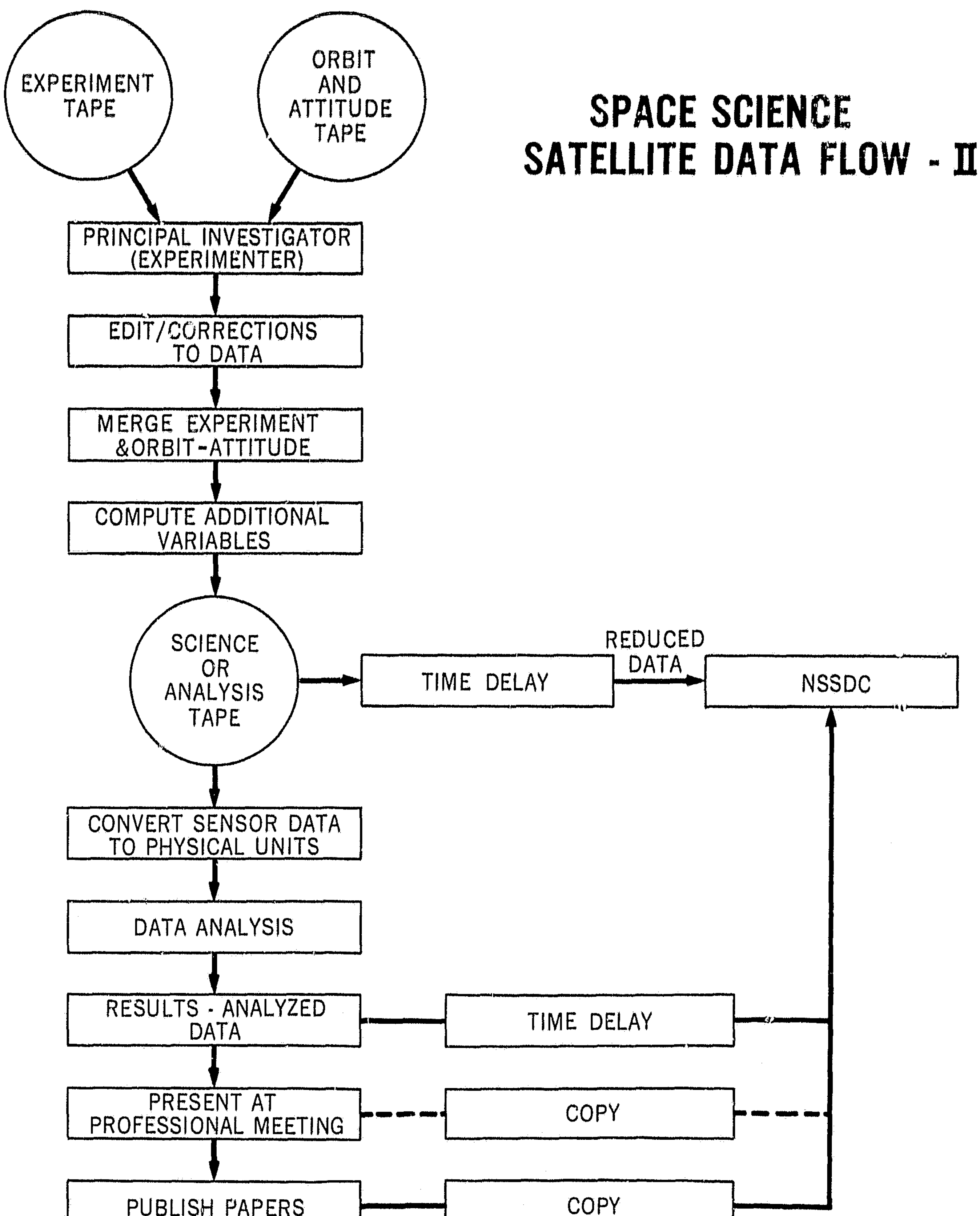


FIGURE 2
DIAGRAM OF SATELLITE DATA FLOW FROM EXPERIMENTER THROUGH ANALYSIS TO FINAL PUBLICATION
The phases where data is collected by NSSDC are shown.



In these cases the information pertaining to each individual experiment is stripped off the master tape onto separate tapes which are sent directly to the appropriate principal investigator. These experimenters are located in universities, government laboratories, NASA field centers, and in private industry. During this decommutation of the master tape, certain items of information such as spacecraft attitude are placed on a separate tape which will be merged with the orbit data as soon as the position of the satellite can be determined from the tracking data. This orbit-attitude tape is also sent to each experimenter.

The data processing at the principal investigator's facility involves editing out those periods in which the data are useless because of telemetry noise or malfunctions. It also involves applying correction factors to convert sensor response to standard conditions of temperature, voltage, amplifier gain, etc. This corrected data is merged with the satellite ephemeris and attitude data; generally, additional variables derived from positional information and used in data analysis are computed at this time. The resultant information usually appears on a digital tape commonly called a science or analysis tape. Hopefully at this point none of the basic information content of the experiment has been destroyed. The data in this form, usually called reduced data, are what NSSDC attempts to collect. Other scientists active in the specialized fields, given the detailed characteristics of the instrument, then can independently interpret the data for their own purposes. As pointed out earlier there is a time delay of 2-3 years before this data is made available to the Data Center. However, for programs such as the picture-taking satellites, the photographs are available several months after launch. The negatives are considered the reduced data for this class of experiments.

In order to proceed to the data analysis phase the experimenter must convert the detector or instrument response to a physically meaningful quantity. This normally includes a calculation using the calibrated instrument response and some assumptions about the various phenomena being observed. Frequently, models of the physical processes are used; some of these later prove to be incorrect. It is really at this point that the interpretation of the results commences. Consequently, for certain purposes, it is important to obtain the data for future users prior to this step. This point should be explained more completely. The data analysis process is one that varies considerably from experiment to experiment and from laboratory to laboratory. The results of this analysis represent the interpretation of the experiment by the principal investigator and his co-workers and display the scientific meaning from their point of view. These results are communicated to colleagues and other interested parties through oral presentations at professional meetings and through publication in scientific journals.

These results, which are called analyzed data, are very usef... to many other scientists and engineers since they represent the judgment of outstanding experts in the field. Therefore, the Data Center includes this type of information in its collection efforts. These data may include charts, graphs, photographs, and tables; examples of these appear in published works but usually there are too many to be published in their entirety.

ACTIVITIES OF THE NATIONAL SPACE SCIENCE DATA CENTER

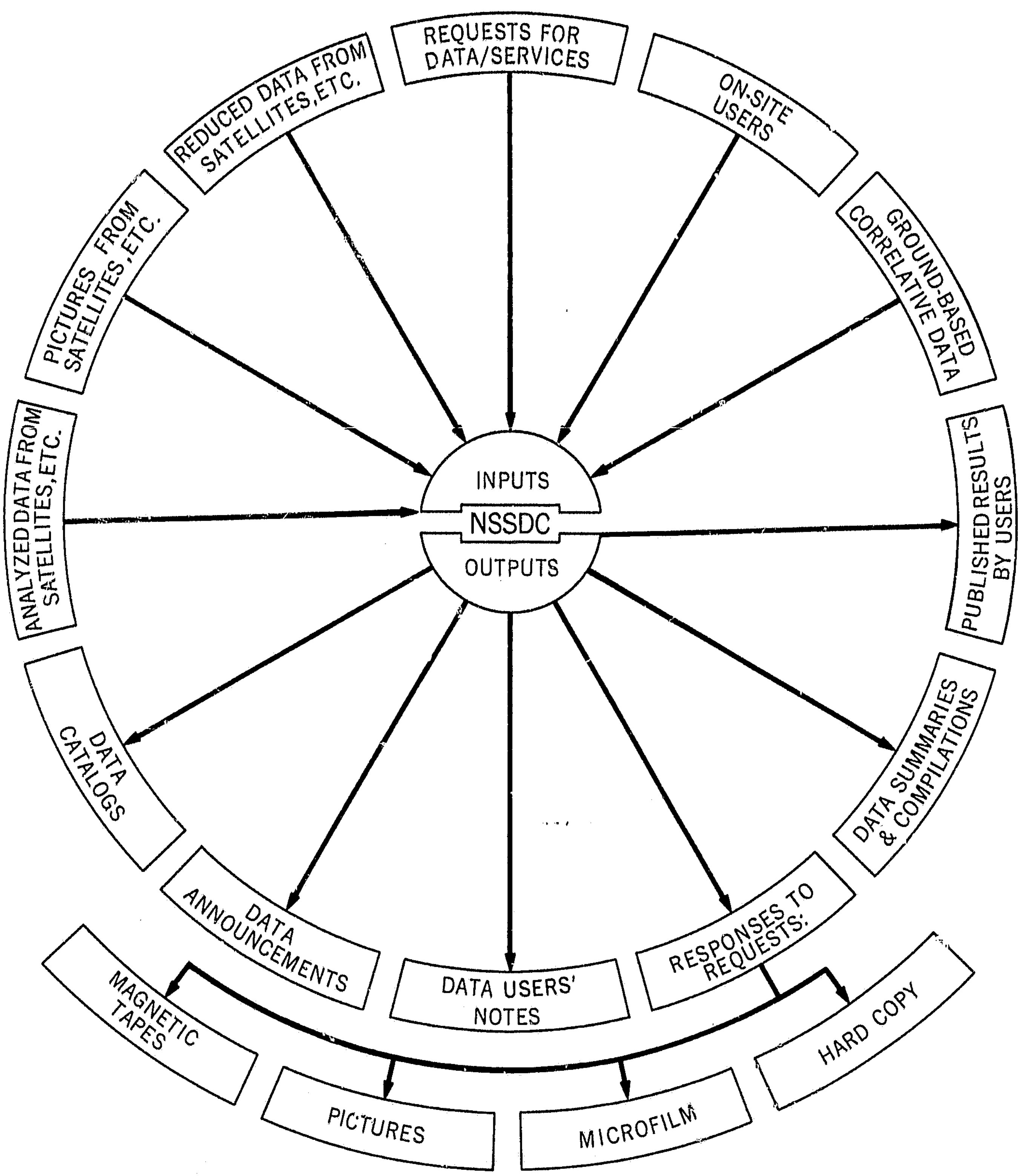
Now that it has been shown where the Data Center fits into the general scheme of space science data, the NSSDC operation merits a closer look. It should be apparent by now that the acquisition of data in the proper forms requires people who are familiar with the instruments, data processing, and the interpretations of the data. Thus, professional people who have specialized in the various space science disciplines are necessary to the successful operation of the Data Center. Furthermore, the proper documentation describing the instrument quantitatively and giving the results of calibrations must be obtained or the data may well be useless to other scientists and users.

This total acquisition is critical, as pointed out before, because the data obtained by space science experiments are extremely useful to others engaged in space activities. The engineer and system planner require a knowledge of the space environment in order to design the spacecraft subsystems and to perform the trade-off calculations necessary in obtaining the most efficient total operation. Most of the knowledge of the space environment comes directly from scientific measurements. However, the reduced or analyzed data is not in the proper form to be utilized directly by other than space science specialists. In fact, it is necessary to study the analyzed data (and in some cases to perform additional analyses with the reduced data) from many experiments in order to obtain a fairly complete description of the space environment. This translation of reduced or analyzed data into data summaries or compilations is a natural professional activity for those Data Center personnel involved in the data acquisition. Some of this is being done now, but there will be more activity as the staff and professional competence grow.

To be specific about the activities of the Data Center, the most important inputs and outputs are shown in Figure 3. Besides the data obtained by space experiments, which have already been described, the results of many ground-based measurements are useful in the interpretation of flight results. For example, primary and secondary effects of the molecules, atoms, ions, electrons, and photons detected in space have been measured by ground-based instruments for many years; whereas our present ability to place sensors in orbit has extended the capability of measuring r tural phenomena. The correlations of the

FIGURE 3
A SUMMARY OF THE ACTIVITIES AND PRODUCTS
OF THE NATIONAL SPACE SCIENCE DATA CENTER

ACTIVITIES OF THE NATIONAL SPACE SCIENCE DATA CENTER



two types of measurements greatly increase our understanding of the complex, stochastic phenomena encountered. Thus, the raw materials for the Data Center are comprised of space science data and ground-based correlative data; this latter information is generally obtained from other data centers, particularly those run by ESSA,* which specializes in its acquisition and distribution.

To enhance the dissemination of the available data, catalogs which contain a cumulative listing are published semiannually. In cases where recently acquired data sets have a wide interest, a special data announcement is published as soon as possible. Based upon documentation provided by the experimenter, by his data processing people, and by the published articles, a Data Users' Note is written by the NSSDC staff and reviewed by the principal investigators. This document provides the key information for the use of the data by other scientists. The Data Center has facilities for visiting scientists to perform studies on-site or to select those sets to be sent to their institutions for further study.

Requests for data through catalog and on-site selection result in the production of magnetic tapes, cards, pictures, microfilm, or copies of written, graphical, and tabular material. To demonstrate the growth of the use of NSSDC and to indicate the present level of activity, the number of completed requests as a function of time are shown in Figure 4. A breakdown of the requests during the past 6 months into various types is shown in Table 1. It can be seen that lunar and terrestrial photographs constitute a large portion of the present output. It should be remarked that correlative data are only distributed to NASA personnel since other data centers serve as the national distribution points.

To show the request output in still another way, Table 2 lists the materials which have been supplied during the period 1 June - 18 August 1967. The backlog or requests in process are also given. The practice to date has been to supply small requests free of charge if the use of the data is for non-commercial purposes. Costs of reproduction are charged for all other requests.

THE USE OF THE DIGITAL COMPUTER

Although the Data Center has been in existence for 2-1/2 years, extensive use of the digital computer has been undertaken only recently. Much of the data collected in the beginning was not in digital form, and a manual operation sufficed. At the present, several different computer-oriented projects are being implemented.

*Environmental Science Services Administration.

FIGURE 4
THE GROWTH OF NSSDC USERS
The availability of Lunar pictures produced the rapid increase in early 1967.

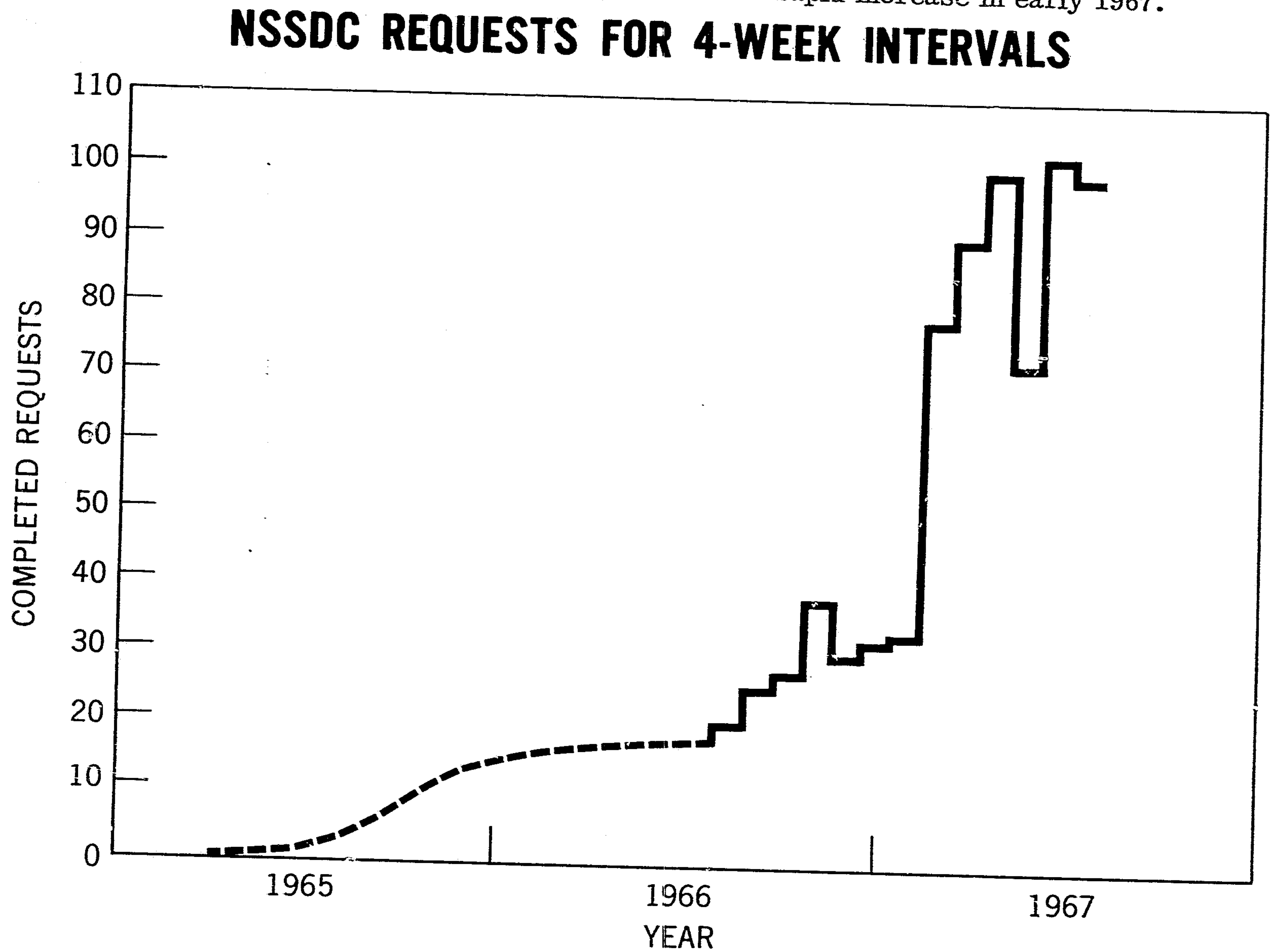


TABLE 1

NSSDC FRACTIONAL BREAKDOWN OF COMPLETED REQUESTS

TYPE OF REQUEST	PERCENTAGE OF TOTAL NUMBER
REDUCED DATA	19.1
CORRELATIVE DATA	19.6
CLOUD & EARTH PICTURES	20.9
LUNAR PICTURES	40.4

TABLE 2

NSSDC REQUEST OUTPUTS

MEDIUM	UNIT	REQUESTS COMPLETED BETWEEN 6/1/67 - 8/18/67	REQUESTS IN PROCESS
DIGITAL MAGNETIC TAPE	2400' REEL	98	449
PUNCHED CARD	CARD	40,176	1,394
COMPUTER PRINTOUT	SHEET	7,850	1,224
MICROFILM	REEL	294	985
HARD COPY	PAGE	4,853	3,400
PHOTO - LUNAR ORBITER SURVEYOR GEMINI NIMBUS	POSITIVE, NEGATIVE, OR TRANS-PARENCY	956 215 386 700	6,800 19,000 277 661

A request accounting, status, and history file is being developed to handle requests more efficiently. This system will provide an immediate status for any request and estimate the time necessary for completion. It will generate standard reports on the number of requests received and processed, generate automatic action reminders, identify the work queues, and maintain a request log. Statistical studies will be made on this request file to determine the frequency of demand for given data sets, the types of outputs requested, and the identification of the various users.

Quite naturally, an information storage and retrieval system is also being developed. All of the information deposited in the Data Center will be entered into this system. Appropriate search terms such as investigator, satellite (or other vehicle), instrument type, time period, and other keywords will be used. A good working knowledge of the various scientific disciplines and the forms of data requests are necessary to provide effective inputs to this system.

One of the most challenging problems that faces the Data Center operation is that of handling the vast array of magnetic tapes that will be supplied in the future. In order to bring this problem into proper focus, an estimate of the amount of scientific data generated by satellites as a function of time is shown in Figure 5.

The data that has been collected so far mainly covers the time period prior to 1965. The exceptions to this are the lunar and terrestrial pictures; all of those photographs taken by Gemini, Surveyor (I & III), and Lunar Orbiter (I, II, III, and IV) are available. It can be seen that large amounts of data will be handled within the next few years. Best estimates indicate that within three years 10,000 reels of magnetic tape will be deposited annually. The present holdings at NSSDC and estimated annual inputs are listed in Table 3.

Because the main sources of data for NSSDC are spread throughout this country and abroad, a wide variety of tape formats produced on various computers are anticipated. Roughly 500-1000 space science experiments are handled by 50-100 groups throughout the United States (a few dozen experiments have been performed by foreign scientists in conjunction with NASA). Unfortunately, magnetic tape is not a good long-term archival medium compared to photographs or printed matter. Although the quantitative lifetime of reliable storage on tape is not known, two to three years is all that is expected for high reliability.

Since tape damage or loss of data can occur during usage, it is necessary to maintain a spare copy. Consequently, a digital data base consisting of 30,000 tapes implies storage facilities for 60,000 tapes and the overhead expense of generating 20-30,000 tapes annually. It is obvious that any system of this

FIGURE 5

YEARLY GENERATION OF SPACE SCIENCE DATA BY SATELLITES

These totals do not include the data from pictures.

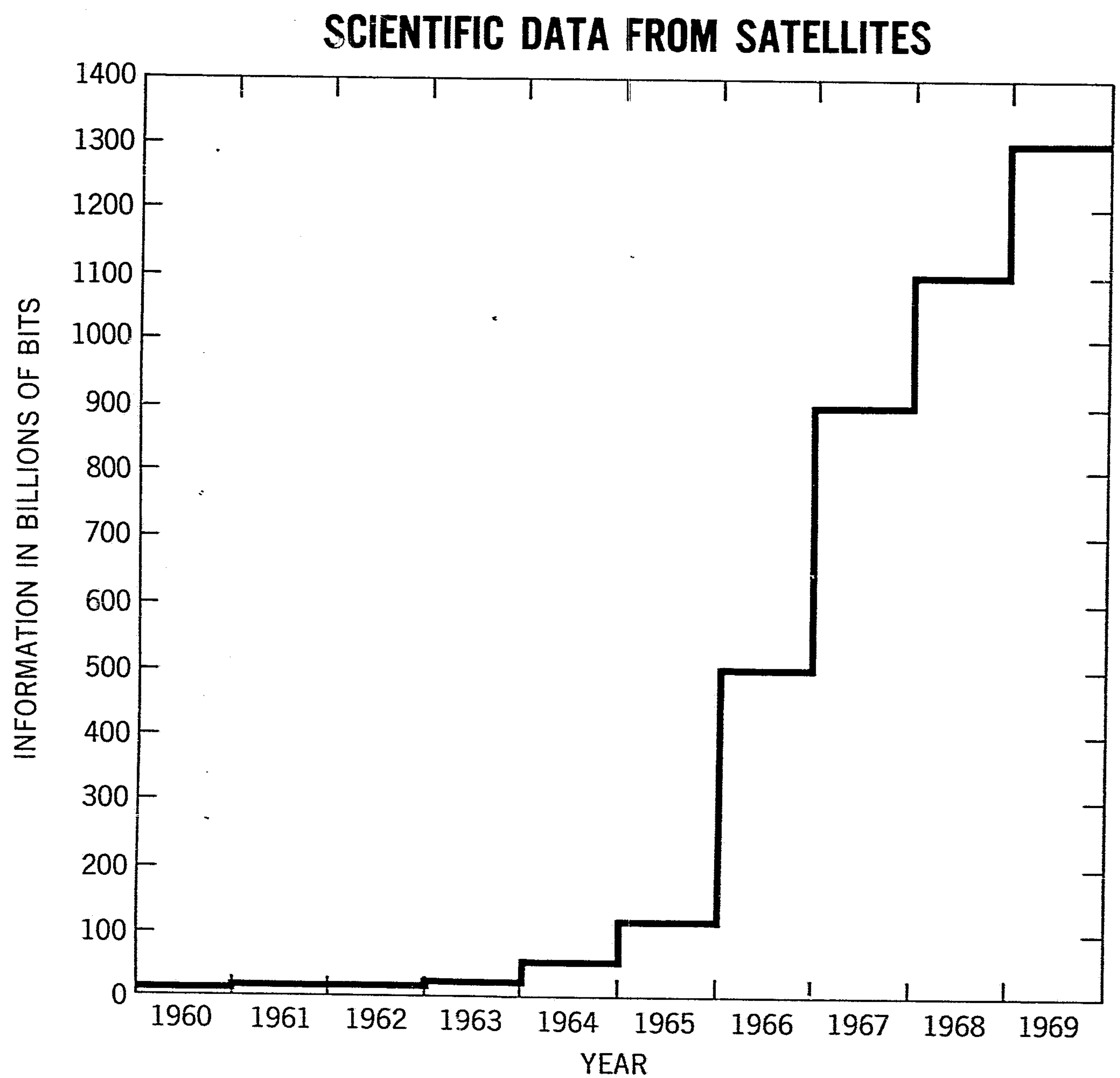


TABLE 3
VOLUME OF DATA AT NSSDC
Present (August 1967) and Projected, Annually (January 1970)

MEDIUM	ON HAND	INCOMING
SHEETS AND BOUND VOLUMES, SHEETS	175,000	100,000
ROLL CHARTS, LINEAR FEET	360,000	150,000
DIGITAL MAGNETIC TAPES, 1/2"x2400'	291	10,000
ANALOG MAGNETIC TAPES, 1/2"x2400'	1,035	0
MICROFILM, 100-FT ROLLS	7,800	2,000
PHOTOGRAPHIC FILMS:		
9-1/2" WIDTH, LINEAR FEET	14,000	4,000
70-mm WIDTH, LINEAR FEET	33,200	12,000
4- x 5-INCH, EACH	2,100	1,000
16- x 20-INCH, EACH	93	--
20- x 24-INCH, EACH	2,200	800
PHOTOGRAPHIC PRINTS:		
8- x 10-INCH	600	500
11- x 14-INCH	200	300
16- x 20-INCH	93	--
20- x 24-INCH	2,200	800

magnitude must pack the data efficiently to minimize the number of tapes. To do this, it is necessary to convert all incoming tapes to a standard internal format. Although this requires the development of computer software to process every incoming tape, the advantages are numerous. A quality check of the incoming data can be made. Errors or omissions can be discussed with the supplier of the data; hopefully these can be cleared up immediately instead of trying to obtain the information several years later from the principal investigator. As a minimum, the discrepancies can be flagged. All maintenance and system quality control programs can be written to handle a standard set of tapes rather than a heterogenous mixture. Requests for data on magnetic tape can be processed by reformatting the tape so it can be handled efficiently at the requester's facility. Work on such a standard format is in progress but it will be 6-12 months before the system will be operational.

With the amount of information increasing at a great rate, thought also must be given to retiring from the active base or compressing that data which is not being used. Within the next 5 years higher density storage techniques for digital data should be readily available. Several different processes are under development which will result in volume reductions of 100 up to 40,000 over standard

556 characters-per-inch magnetic tape. Reductions of this size will help to solve the storage problem. The photographic or metal film used in these processes also provide archival media which require much less maintenance than magnetic tape.

Besides the technological improvements, it should be possible to compress the actual data. First, a study of the data utilization through the request history file will provide insight into which items of information can be removed from the active data base. A preliminary study has indicated that the derived variables, which accompany the basic data, average about 10 times as many bits as the sensor measurements. Since these variables are very important in most analyses, they are not removed when the data is first deposited at NSSDC. However, for data that receives very little use, the removal of the variables would produce a large reduction in volume. A user of the data in this form could recalculate those variables from the original information.

There is another area of possible data compression. Most of the space science data can be broken up into two different types. One is the ambient, quiet time, background information and the other is the disturbed time, event information. There are many different events (many are related) such as magnetic storms, solar flares, and solar proton, electron, and X-ray events. These and other events are of extreme interest. However, the background information constitutes the majority of the scientific information obtained since the beginning of the space age. As newer instruments become available which have better time, energy, or spectral resolution, the fine details of the older background data will not be needed. Time averages of this background data can then be made over hours or even days. This averaged data will still be useful in determining the long-term changes which are apparent in the solar-geophysical phenomena.

Thus, there appear to be several ways to cope with the data explosion in this area of interest. By keeping the active data base to a reasonable size, the maintenance costs will be reduced. This will free resources for application to the storage and analysis of newer data.

SUMMARY

The National Space Science Data Center provides the means for analysis and dissemination of space science data beyond that provided by the primary experimenters and their co-workers. Because of the large amounts being generated and analyzed, it is necessary to be selective in the data that constitutes the active base if it is to be put to optimum use in the future. In this context, the computer is a useful tool for management and administrative control of the operation; it is mandatory for the processing of magnetic tapes and the implementation of information storage and retrieval systems.

The Data Center has to have professionals in the space science disciplines to: (a) acquire the data in the proper form from the experimenters, (b) maintain the appropriate interface with the scientific community, (c) provide the proper entries for the storage and retrieval system, and (d) translate the data into forms useful to the engineering and management communities. Although scientific societies have recognized the importance of the compression and exchange of information through data or information centers, more space science professionals are needed in this activity now; many more will be needed in the future.