X-601-69-416
PREPRINT

NASA TM X-63739

DATA INFORMATION SYSTEM AT THE NATIONAL SPACE SCIENCE DATA CENTER*

NICK KARLOW
JAMES I. VETTE

SEPTEMBER 1969

GSFC

GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

*To be presented at the ASIS 32nd Annual Convention, San Francisco, California, October 1-4, 1969.

X-601-69-416

DATA INFORMATION SYSTEM AT THE NATIONAL SPACE SCIENCE DATA CENTER

Nick Karlow
James I. Vette

National Space Science Data Center

September 1969

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

ABSTRACT

The National Space Science Data Center (NSSDC) developed an integrated information system to support its data handling activities. This paper presents a brief discussion of the information system, which is comprised of four main subsystems: the Automated Internal Management File, the Machine-Oriented Data System, the Technical Reference File, and the Request Accounting Status and History File. This system satisfies current operational requirements; however, it must grow to meet future needs. Considerations include: new and better software and hardware, orderly retirement of data with and without loss of information content, and experimenter on-line interactions to data held at NSSDC.

CONTENTS

ILLUSTRATIONS

DATA INFORMATION SYSTEM AT THE NATIONAL SPACE SCIENCE DATA CENTER

INTRODUCTION

Satellite measurements produce a tremendous amount of data to be processed and analyzed. For example, during 1968 there were over one trillion bits of data transmitted to the earth. The nature of scientific measurements requires that a much longer active life be provided for the data, as opposed to engineering, biomedical, and applications measurements. Large volumes of data are necessary to obtain patterns, interactions, and relations. As new ideas develop in understanding the phenomena, scientists will want to analyze further much of the existing data. This, in preference to repeating a measurement, is necessitated in space science because it now takes approximately 5 years between experiment conception and receipt of data for analysis.

For some of the earth and environmental sciences, satellite experiments represented only a new technique of measurement, not the real beginning of new scientific disciplines. However, those disciplines which comprise space science were really developed as a result of satellite and rocket experiments. It was clear a new data center needed to be established to handle this function. As a result, the National Space Science Data Center (NSSDC) was founded in 1965 with the primary function of providing the means for the dissemination and analysis of space science data beyond that provided by the original experimenter. The Data Center is responsible for the active collection, organization, storage, announcement, retrieval, dissemination, and exchange of space science data.

To satisfy the needs of the various user groups (space scientists, scientists in related fields, engineers/systems planners, and educational activities), specialists in the various space science disciplines, systems analysis, computer programming, data processing, technical writing, publication, and reproduction are required. The complexity of the job to be done has required the adoption of a total systems approach and the automation of the Data Center. A second reason for automation is the large volume of data which requires the handling of numerous magnetic tapes, cards, pictures, microfilm, and copies of written, graphical and tabular materials.

NSSDC has grown considerably in the last 3 years. There is a well-established flow of data and information into and out of the Center. In order to effect this transfer of knowledge and keep track of the large number of transactions which occur, an information system has been developed. Because most of the known literature on operational information systems primarily deals with documents, it is the purpose of this paper to present an overall view of the Data Center as a system, trace the flow of information, and discuss some of the changes envisioned for the future. Hopefully, it will prove useful to people like yourselves who are interested in information transfer and who require an understanding of the detailed processes involved in setting up, operating, and developing an effective data center.

FLOW OF DATA INFORMATION INTO THE SYSTEM

It was within this framework that the Data Center planned and developed its current integrated information system. To oversimplify the mission of NSSDC, one must first arrange for obtaining the space science data and understand the form/format of incoming data. Once the data begin to arrive, there must be a central source of information concerning these data. This need is satisfied by the subsystem called Automated Internal Management (AIM). Upon arrival, one must process the machine-sensible data, prepare it for retrieval, and be able to handle special types of data in different forms and formats - this is done through the Machine-Oriented Data System (MODS). (The steps for processing non machine-sensible data are generally analogous.) In their work, the professionals at NSSDC must have ready access to the documentation relating to appropriate satellites, experiments, and data - the Technical Reference File (TRF) serves this purpose. Finally, statistics must be kept on the processing and use of data, and management must have a variety of reports in this area - this is greatly facilitated by the use of the Request Accounting Status and History (RASH) file. These subsystems are supported by five additional special-purpose files: Computer Program Status Report, Magnetic Tape Unit Record, Computer Program File, Rocket File, and Distribution File.

To obtain a better understanding of the NSSDC information system and to get a broad picture of what happens during this process, it will be helpful to follow the path of information flow from the experimenter to the system. This information flow is summarized in Figure 1. First, a space data scientist, also referred to as an acquisition agent, is assigned to each satellite/experiment/data set, as appropriate. He then obtains advance prelaunch information from such sources as the satellite project office, news releases, bulletins, reports, and personal contacts. This and subsequent information is entered into a working acquisition file, and, at this time, an AIM entry is generated. The agent establishes contact with the experimenter and his data processing personnel to arrange for bringing in data and related documentation. It should be pointed out that long periods of time are normally involved between this first stage and getting the actual data into the NSSDC system. This usually takes two or more years after launch.

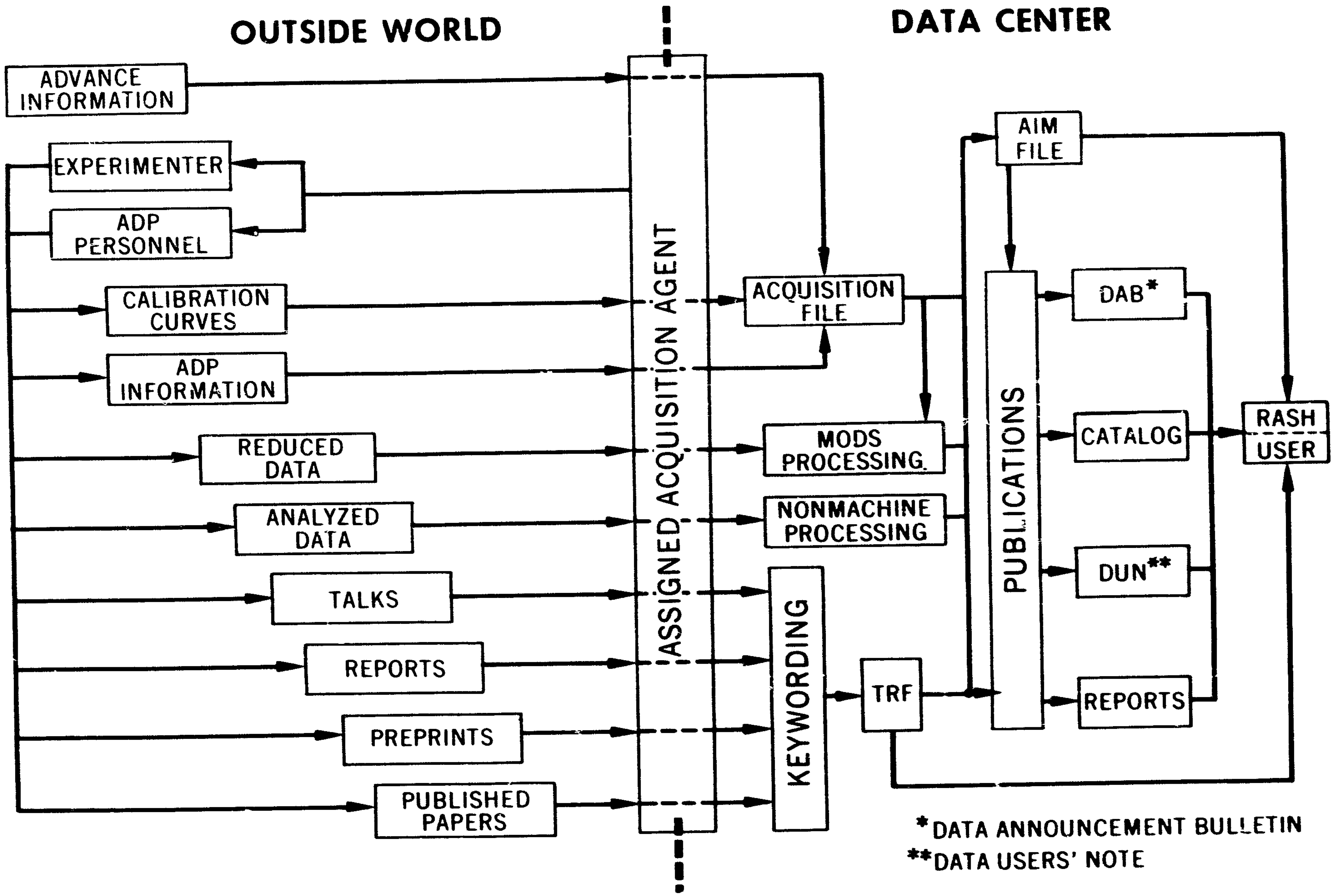

Figure 1. Flow of Information from the Experimenter into the NSSDC System

Once the preliminaries are over, the acquisition agent remains in constant contact, through visits or phone, with the experimenter and his data processing staff to help solve problems relating to the submission of data and documentation. Thereafter, the data and information come in almost on an automatic basis, except where special problems arise. The first items that should arrive at the Data Center are usually calibration curves, unpublished information, instrument descriptions, and data processing documentation. These are analyzed and selectively entered into the acquisition file and TRF, and notices are placed into the AIM file for subsequent use in processing incoming data and preparing publications.

Next, the reduced data,* consisting mainly of magnetic tapes, arrive. At this time, the acquisition agent, together with a programmer as required, verify and analyze this data, prepare duplicates of the tapes, and prepare data set catalogs (indexing) using the MODS subsystem to accomplish these tasks. At this point, tape reformatting often has to be accomplished. The agent then feeds appropriate information into special-purpose files such as the tape and program status files. The AIM entry is brought up to date. (These subsystems will be discussed later.)

The analyzed data,** normally made up of plots, graphs, and tables, arrive quite a bit later. Of course, for older experiments that are not yet in NSSDC, analyzed and reduced data may arrive in any order. The acquisition agent again goes through a similar processing cycle as in the case of machine-sensible data. Data are verified, analyzed for information content, logged, indexed, copied or microfilmed, and the information entered into the AIM information subsystem. (1)

The working relationship with the experimenter is beneficial for the information transfer in other ways. Through this association and contacts at professional meetings, the acquisition agent receives copies of appropriate talks, reports, preprints, and published papers, as well as gaining a deeper understanding of the experiment and the implications of the data. (These items are supplemented by a thorough screening of the current literature.) Each of these documents is carefully analyzed, keyworded (by the acquisition agent), and entered into the TRF. Appropriate information is extracted for entry into the AIM.

---

*Defined as corrected sensor data merged with satellite position data still retaining the basic information content of experiment.

**Defined as investigator-interpreted experiment data displaying scientific meaning from his point of view.

It should be clear by now that the acquisition agent spends a great amount of his time studying and working with the data to put all the necessary information together so that it may be useful to others. Using the information from the AIM and TRF subsystems and special-purpose files, the acquisition agent and publications staff prepare, as necessary, Data Announcement Bulletins and entries for the Catalog of Satellite and Rocket Experiments. This does not necessarily mean that all data from a particular satellite experiment have arrived or have been completely processed. Many other contacts and correspondence with the experimenter may still be necessary. The preparation of a Data Users' Note concerning a particular experiment normally occurs after the final stage of data acquisition and processing. This document shows where the supporting information is, in what forms the data are available, what literature of previous work relating to the experiment is available, and offers a key to the use of the data.

The information flow is not complete without a mention of the RASH subsystem. The acquisition and request agents work through the RASH subsystem in satisfying users' requests on a daily basis. These requests may involve copies of data or publications, logical searches of the information files, or may even require further detailed data analysis on the part of the acquisition agent to help solve a particular problem.

THE NSSDC INFORMATION SYSTEM

Thus, the NSSDC information system used to handle this flow of information is comprised of four main subsystems and five special-purposes files. (2) It is integrated at the operating level through the capability of logical searches of the AIM, RASH, and TRF subsystems. For example, AIM search items include: satellite, launch date, silent date, experiment group, rank (priority for acquisition), and acquisition agent. Standard satellite names, numbers, and discipline keywords are used.

Automated Internal Management (AIM)

AIM, as the centralized source of information, is the hub around which the other subsystems revolve. It is built upon detailed descriptions of the data, experiment, and spacecraft, along with the status of acquisition activity. The purposes served by the AIM subsystem are detailed in Figure 2. In addition to supplying this type of information, AIM has the capability to perform information retrieval on a more general scale.

The contents of the AIM file are organized into a hierarchical structure. The most significant level is the spacecraft. Information which relates to the

AIM AUTOMATED INTERNAL MANAGEMENT

- LOGICAL SEARCHES TO ANSWER QUERIES
- WORKLOAD/VOLUME OF EXPECTED DATA
- ACTION REMINDERS

- ACQUISITION MANAGEMENT REPORT
  - SPACECRAFT/EXPERIMENTS/DATA SETS
  - RESPONSIBLE AGENT
  - PRIORITY
  - STAGE OF ACQUISITION
  - HOURS EXPENDED
  - CURRENT ACTIVITY
  - NEXT CONTACT
- FILE INDEX
  - LISTING OF SPACECRAFT/EXPERIMENTS/DATA SETS

Figure 2. Uses of AIM

spacecraft and is implicitly true of all experiments on that spacecraft is included here. The second level relates to the experiment. All experiment identification, detector descriptions, and commentary about a single experiment are contained in this section. The third level deals with a single data set.* These levels are generally tied together in the following manner, depending on identification of experiment and availability of data: the satellite-level entry will be followed by all experiment-level entries which pertain to that spacecraft; similarly, the data set-level entries are associated with the experiment. This concept can be perhaps better visualized by examining the typical AIM File Index entry shown in Figure 3. Point 1 shows the satellite entry, point 2 the experiment level, and point 3 the data set entry, with pertinent information following across the remainder of the line.

*Defined as a body of data reduced by one group of investigators in one specific way in a form, format, or organization which uniquely describes it. It can be a unit of machine-sensible or nonmachine-sensible data which can contain one to several hundred magnetic tapes, rolls of film, etc.

1 / 2 / 3	4	5	6	7	8	9	10	11	12	13
58-005A	EXPLORER 4	B	BPS	A	EXPLORER 4 1958 EPSILON,	07/26/58	US ARMY	805A00 46 1 4 DE		,0913
58-005A-01	EXPLORER 4	B	JIV	A	EXPL 4 CHARGED PARTICLE DETECTORS		VAN ALLEN	805A10 134 12 4 B DE	0813	,0904
58-005A-01A	EXPLORER 4	B	JIV	A	EXP 4 UCSD TAPES, STATION ORDERED		MCILWAIN	805A1A 106 12 4 B DE	0719	,0904
58-005A-01B	EXPLORER 4	B	JIV	A	EXP 4 TIME ORDERED TAPE NSSDC		MCILWAIN	805A1B 23 12 DE	0820	,0904

1 SPACECRAFT IDENTIFICATION
2 EXPERIMENT IDENTIFICATION
3 DATA SET IDENTIFICATION
4 SPACECRAFT NAME
5 DISCIPLINE GROUP (INTERNAL CODE)
6 ACQUISITION AGENT INITIALS
7 PRIORITY
8 SPACECRAFT/EXPERIMENT/DATA SET
9 DATE OF LAUNCH
10 AGENCY/INVESTIGATOR
11 INTERNAL CODES
12 DATE OF LATEST ENTRY
13 MANAGEMENT INFORMATION UPDATE

Figure 3. Typical AIM File Index Entry

Within each of these levels in a full AIM entry, there are specified categories of information concerning personnel, objectives, telemetry, instrumentation, acquisition information, experiment performance, data set contents, etc.

As noted earlier, AIM is also used for providing management information. Based on the same levels just discussed, the following information is provided concerning spacecraft, experiment, and data set.

- NSSDC Identification
- Descriptive Information
- Acquisition Agent
- NSSDC Rank (priority assigned to data set)
- Total Acquisition Man-hours Used
- Man-hours Used Last Month
- Last Visit
- Last Contact
- Next Scheduled Contact
- Acquisition Stage
- Task Wait (processing stopped/reason)
- Data Set Form/Size (type/units)
- Date of Data Arrival
- Date of Completion

To sum up the AIM subsystem in terms of output, it is used to produce many reports, some of which are:

- Acquisition Management Report
- Acquisition Notebook
- AIM File Index

- Action Reminders
- Satellite List
- AIM Full Printout

Machine-Oriented Data System (MODS)

To be responsive to the users who request data in digital form, as well as to those who provide the original data, NSSDC must have the flexibility to accept the data in any format and to provide it in any format. Since both the giver and taker will have restrictions imposed by their existing computer hardware and software, the Data Center facility must provide the "impedance matching." MODS is used for processing the data into the NSSDC computerized data base, for data set analysis, generation of data set catalogs, tape reformatting (when the interchange of information is inhibited by the diversity of hardware), and production of allied reports. Perhaps the best way to examine the composition of this subsystem is to follow incoming machine-sensible data sets through their processing cycle and then look at the tape reformatting process.

Processing Incoming Data Sets—All magnetic tapes received by NSSDC are first entered into the storage records by filling out the proper forms and assigning a unique accession number. At this point, an acquisition agent, to be assisted by a programmer as necessary, is assigned to the data set for preliminary analysis.

The joint objectives of the acquisition agent and programmer in the preliminary processing are: (1) ability to read the entire tape in its logical format; (2) ability to list out any function or special record; (3) ability to detect errors (logical and physical); and (4) verification of the acquisition agent's understanding of the contents.

During this preliminary processing, the programmer writes all the necessary routines to manipulate the data and reformat it, if necessary. These programs are entered into the Computer Program File. The preliminary analysis stage is completed when NSSDC has the ability to use and interpret all data in the data set. This may require additional contacts with the experimenters.

At this time, the acquisition agent and the programmer define the format of the Data Set Catalog, the functions of which are to:

Provide an index to the contents of the data set

Provide a series of error checks

Calculate bounds or distributions of functions

Provide a useful tool to the request agent for identifying data

Provide a coarse description of the information in the data set

After the requirements are defined, the programmer writes a program to produce the Data Set Catalog. This routine should also produce a copy of the original tape or a reformatted version. After checking the program and turning it over to the computer people, the rest of the tapes in the data set are processed.

Tape Reformatting—The processing of normal machine-sensible data is well taken care of by using the procedures just outlined. However, experience has shown that people will not use data if it takes a lot of time and effort to convert it to a format which allows for direct entry into their own computer. Consequently, one of the major problems confronting the Data Center is the processing of magnetic tapes produced by different computers and operating systems. This presents two main difficulties: the physical problem of reading tapes which cannot be used with the normal hardware available to NSSDC and the logical problem of interpreting data where word size, word format, data arrangement, etc., cannot be easily handled with standard software. The approach used is to achieve the desired flexibility by producing software which is bit-oriented rather than character- or word-oriented. The Data Center currently has routines available to read tapes generated by a number of operating systems,* as well as BCD (binary coded decimal) tapes with arbitrary and variable record sizes, physically formatted binary tapes, and FORTRAN generated tapes. For achieving compatibility with systems using 9-track tape, NSSDC uses other computers at GSFC. The hub of the MODS subsystem is a package called PIFT (Package for Information Formatting and Transformation) which will accomplish the functions just outlined and at the same time will produce densely packed machine- and media-independent data sets that may be accessed in the man-machine mode. A basic building block of this subsystem is a bit manipulator program which has recently been completed.

Technical Reference File (TRF)

The Data Center professionals must have internal documentation support and a tool for satisfying the bibliographic needs of space science data users. This is why the TRF comes into the system. It provides access to documents used for evaluating and verifying acquired data and for publishing catalogs, Data

*BESYS, APLOS, IBM-DCS operating systems.

Users' Notes, and bulletins. It provides a useful record of the documentation available at NSSDC, as well as that which exists in the published literature. The references include published and unpublished documents relating to the spacecraft, experiment, or data set which are or will be preserved at the Data Center. No attempt is made to physically maintain references of a general scientific nature or in any way duplicate the services of the NASA library facilities. However, the classification of the documents and the descriptors are usually different because the depth of indexing is tailored specifically to the space science community and to NSSDC retrieval use.

The computer can display pertinent information in a variety of ways. Open (subjective) and controlled keywords are used to cover standard satellite/rocket identification, type and content of publication, and discipline-oriented keywords. The methodology for describing the type and content of publication can be seen in Figure 4. A typical TRF entry is presented in Figure 5.

To overcome the common gap between indexer-selected terms and user-selected terms, the acquisition agents themselves, who are space data scientists and the prime users of this subsystem, review the literature, select the entries, and keyword the inputs.* Thus, each member of the acquisition staff devotes a portion of his time to building up the TRF base and verifying the output. In this manner, up to 120 items per week are entered into the file, which now contains well over 4,000 entries.

As concerns the external uses of the TRF, considerable effort is presently being devoted to the production of a notebook-sized TRF output. Once this is fully implemented, NSSDC will have the capability of producing space science bibliographies ordered by author, discipline, experiment, or spacecraft. To produce special bibliographies upon request, a logical routine has been integrated into the TRF program which allows for the usual Boolean logic searches of AND, OR, and NOT among the entries. Present and additional keywords are also being studied to eventually derive a meaningful thesaurus.

Request Accounting Status and History (RASH)

At this point, the data from the space science experiments have been obtained, entered into the system, and prepared for retrieval. The next step is to facilitate the acquisition and request agents' contacts with the users of these data - this need is satisfied through the RASH subsystem.

*Document storage and retrieval is based upon randomly assigned accession numbers.

PUBLICATION CLASS		CONTENT CODE	
A	JOURNAL ARTICLES	0	BIBLIOGRAPHIES
B	BOOKS	1	THEORETICAL PAPERS
C	GOVERNMENT PUBLICATIONS	2	SCIENTIFIC PAPERS EXPERIMENTAL RESULTS
D	UNIVERSITY REPORTS	3	INSTRUMENT DESCRIPTION PAPERS
E	INDUSTRY REPORTS	4	DISCIPLINARY REVIEW PAPERS
F	MAGAZINES, PRESS RELEASES, AND NEWSPAPER ARTICLES	5	SATELLITE & MISSION DESCRIPTIONS
G	PROCEEDINGS, SYMPOSIA & OTHER COLLECTIONS	6	NEWS RELEASES
		7	DATA PROCESSING PAPERS
H	UNPUBLISHED	8	WORKING PAPERS, MINUTES, ETC.
		9	DATA TABULATIONS

Figure 4. Codes to Identify Publications. Code letters and numerals are combined to identify both type and content of publication by the use of two or more characters, e.g., BG23.

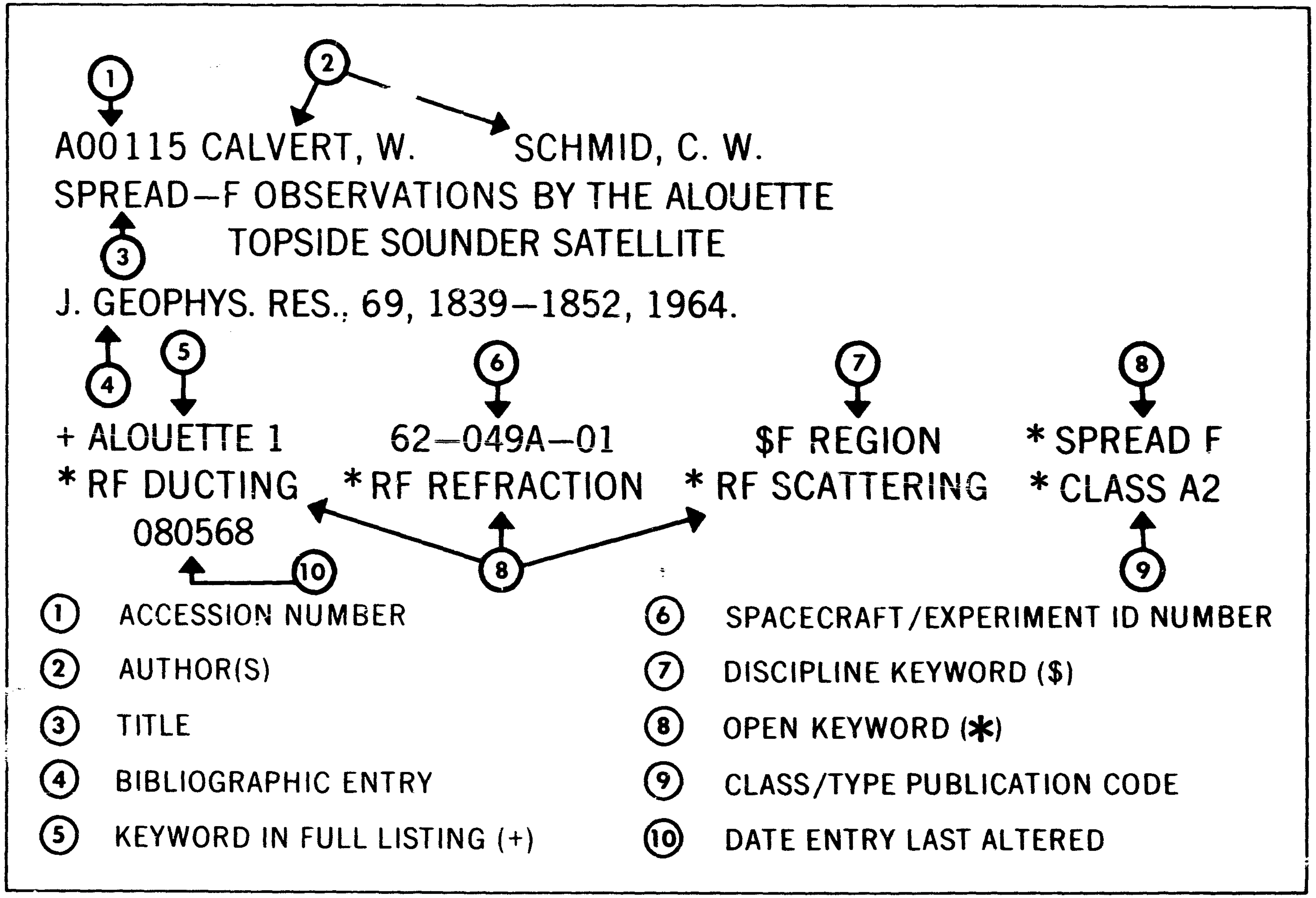

Figure 5. View of TRF Entry

This subsystem provides much valuable information. It is used to aid in keeping track of the progress of requests received by the Data Center and providing management with bookkeeping information. Specifically, RASH is designed to display up-to-date information relating to the number of requests, their status, estimated and actual costs, processing time, and necessary action reminders.* This variety of information can be retrieved by data set, requester, affiliation, date of request, date filled, request agent, status of request, and so forth.

Information on these items is provided in the RASH output, which consists of five weekly reports and five reports generated on an as-needed basis.

Distribution File

As mentioned earlier, there are five special-purpose files set up to help accomplish the Data Center mission: Computer Program Status Report, Magnetic Tape Unit Record, Computer Program File, Rocket File, and Distribution File. The titles of the first four files are self-explanatory, and, consequently, these files will not be discussed. However, a few words should be said about the consolidated distribution list, which, for obvious reasons, is tied to request reporting of the RASH subsystem. It is used for file maintenance and document distribution. In addition to the production of self-adhesive labels, this subsystem will provide listings by recipient, affiliation, publication, etc. It also will have a built-in capability for SDI (Selective Dissemination of Information), should this tool be used in the future.

As concerns the users of NSSDC data and documents, the Distribution File contains the names and addresses of recipients, what they receive, and general organizational classifications in terms of recipient categories. The recipient category codes used by the Data Center are explained in Figure 6. Note that there is a special distribution list for each of the documents that the Data Center produces or for which it has responsibility.

An interesting application of the recipient category code is its use to identify those sections of the potential user community which receive inadequate coverage of the available space science experiment data and documents.

The last point to consider in the Distribution File is the method of updating and purging. This is accomplished through the following mechanisms.

1. Dissemination of copies of the distribution list to all concerned activities at the Data Center

*This subsystem also can aid in the construction of a model of the NSSDC by supplying viable information describing the user community, types of requests and responses, and data sets most likely to be used.

KEY CODE OF ASSIGNED DISTRIBUTION LISTS

1. DATA CATALOG OF SATELLITE AND ROCKET EXPERIMENTS
2. CATALOG OF CORRELATIVE DATA
3. LUNAR ORBITER
4. NIMBUS & TIROS
5. "MODELS OF THE TRAPPED RADIATION ENVIRONMENT," LIST 1
6. OGO PROGRAM
7. GSFC COMPUTER NEWSLETTER
8. "MODELS OF THE TRAPPED RADIATION ENVIRONMENT," (NASA SP-3024 ONLY)
9. SURVEYOR
10. X - DOCUMENT
11. WDC-A SUMMARY OF SOUNDING ROCKET LAUNCHES
12. WDC-A ROCKETS AND SATELLITES CATALOG OF DATA

LIST OF RECIPIENT CATEGORIES

A NASA, GSFC ONLY
B NASA, OTHER THAN GSFC
C DOD
D ESSA (ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION)
E OTHER GOVERNMENT AGENCY
F PRIVATE INDUSTRY
G MUSEUM, PLANETARIUM, OBSERVATORY
H ACADEMIC INSTITUTIONS
I FEDERAL CONTRACT RESEARCH CENTERS
J FOREIGN
K MISCELLANEOUS

Figure 6. Codes Used in Distribution List

2. Continual hand-updates by each section
3. Use of RASH
4. Verfication (annual) by recipients
5. Consolidation, machine update, and redistribution of updated copies

A LOOK TO THE FUTURE

It is apparent that the use of the Data Center will continue to grow at such a rate as to double by 1971. As more staff are added, the rate at which data can enter the system will increase. As the capability of NSSDC grows, the services that it can render will multiply, thereby encouraging further use. However, it is realized that only certain resources will be available for this data activity. Present estimates indicate that if approximately 1% of the budget available for the research areas which produce the original data handled by NSSDC is given to the Data Center, then an adequate facility can be developed to handle the complete data needs.

As the Data Center grows, so must its information system. It must be able to handle large varieties and amounts of data and prepare them for a multitude of different uses. It must rely on new and better software and hardware to effectively perform these tasks, bearing in mind the guiding principle of furthering the effective use of data from space science experiments. The present system software must be upgraded with respect to processing incoming tapes for verification of inputs and quality control - two goals are immediate detection of errors or omissions and standard maintenance and systems quality control programs. Effective purging of the active data base will have to be accomplished. Consideration must be given to a good long-term archival medium as the lifetime of magnetic tape cannot compare with photographic or printed matter, although recent tests are encouraging. Time-phased data compression will be another vital area of concern. Considerations include higher density storage techniques and the actual compression of data. This data compression can occur in various steps. The first step would involve the retirement of alternate forms of the data in which the most useful form would be retained. Then, even this most useful form of data could be subjected to the removal of derived variables, which are computed from basic positional and attitude information. This would still permit recalculation of these variables at a later date, should this prove necessary. At this point, no reduction in the basic information content has occurred. However, if one wishes to use this data, more time and resources will have to be utilized than previously. One is balancing this cost against the maintenance cost of keeping all the derived bits in the active data base. There is a break-even point depending on data usage. As one starts the irreversible process of destroying information content, a sensible approach would be to separate the background information (ambient, quite time) from the event information (disturbed time). This will permit time averaging the background information, say over hours or days, for subsequent use in determining long-term changes. A sizable reduction in the number of bits for a given data set will occur in this process, and yet the information most likely to be used in future studies will still be available. Clearly, data of historical significance would be preserved as long as possible.

In other words, certain points should be considered when planning for the retirement of data.

1. Large volume plus cost of maintenance plus fixed resources dictate the orderly retirement of data.

2. Early in their life, various forms of the data are useful, e.g., time ordered, space ordered, etc.

3. Data can be reduced without losing information content:

   - By eliminating certain forms of data

   - By removing derived variables

   - By keeping only the significant number of bits, not the full computer words

4. Information content of data can be reduced:

   - By breaking out event data from background

   - By averaging the background over suitable time intervals

   - By preserving only special data for historical purposes

   - By preserving only outstanding geophysical event data

   - By compressing data into analyzed forms so that general understanding of phenomena is retained

In short, data can shrink in size and in information content, but knowledge of it never disappears from the scene.

Some thought is already being given to the next generation of the NSSDC information system. One consideration is to provide the Data Center with much greater flexibility and capability by developing varied analysis programs which can be readily applied to the data. Although complete requirements have not yet been defined, it is envisioned that scientists, experimenters, and acquisition agents should be able to interact, on-line, through a computer, to data bases and data sets held at NSSDC. It is also anticipated that in this way the resulting dialogue between two or more scientists can be used to synthesize new information in the process.

These concepts are not too far from reality. With the progression of time, the central processing facilities are performing more work on the raw data before it is sent to the experimenters for analysis. In the beginning, the raw data was sent directly. Now, tapes are digitized and edited, noise flags are inserted, time overlaps are removed, and decommutation is performed. There is interest at present in having the orbit and attitude information merged with the data before it is sent to the experimenter. As high-speed data links become available across the country, there will be no need to send the data to the experimenter. Instead, standard processing will be performed up to the point where detailed analysis can begin. The data in this form could be sent directly into the Data Center where it could be reached via high-speed terminals and manipulated on large computers by the principal investigators using many standardized analysis programs. Special-purpose analysis programs would be constructed on-line by the individual users as the needs arise. At that point, the processing facility and the Data Center will have blended into one operation. There exists today an on-line retrieval system with a 10^{12} bit capacity. (3) This is capable of handling a years worth of space science data at the present rate of generation. It is quite clear that this new type of facility could be a reality in 5-10 years.

REFERENCES

1. "The Operation of the National Space Science Data Center," NASA/NSSDC 67-41, October 1967.

2. Karlow, N., and J. I. Vette, "Flow and Use of Information at the National Space Science Data Center," NASA/NSSDC 69-02, January 1969.

3. Kuehler, J. D., and H. R. Kerby, "A Photo-Digital Mass Storage System," Proceedings-Fall Joint Computer Conference, 1966, pp. 735-742.