DATA MANAGEMENT STUDY

Prepared by: A. Bedford

Approved by: D. Fithian

SUMMARY

This report discusses the problems of handling, controlling and displaying data received from Space Scientific Experiments in the Shuttle Era. A satisfactory system is considered to be one which provides:

1. a suitable response time;

2. easy access to the data; and

3. intelligible human-oriented displays;

for all the various users and all the varied control and display modes. The need for hardware and procedural commonality is stressed as is the requirement for display flexibility. A two tier system of data handling is recommended in which the data is first stored in a data bank and then accessed via CRT/keyboard terminals for display.

This report is primarily concerned with developing criteria for classifying or categorizing the final output of the data management system in terms of response time, quantity of data, type of processing required, purpose of displaying the data and type of display. We believe that analysis of the user requirements in these terms is necessary to provide a sound foundation for the data management system.

TABLE OF CONTENTS

INTRODUCTION

This report discusses the problems of handling, controlling and displaying data received from Space Scientific Experiments in the Shuttle Era. The initial impetus for generating this report was a desire to eliminate some of the problems and restrictions on the data handling and displays which were experienced during the ALSEP program. As the study progressed it soon became apparent that a critical review aimed at up-grading the ALSEP data management system was too limited in scope due to the major advances which have been made in all aspects of data handling during the last ten years. Consequently the scope of the study was widened to embrace all aspects of data management and the applicability of the latest techniques.

The objective of the study is to provide an independent assessment of the requirements and activities for definition and implementation of shuttle payloads data management. A ground rule which was adopted for the study was to perform an initial look at the problem without regard to any other specific studies performed to date.

One major activity of the study was to define the user requirements and objectives so that the data management problem could be evaluated. The "key issues" of any data management system are response time, data access, and data displays.

Data management for space science missions has several different objectives as the mission progresses from pre-launch checkout, through operations, to science analysis and final historical documentation. The first major activity in this study is discussion of the methodology which should be employed to relate the data management objective for any specific data handling activity to the required values of the data management system parameters which control the response time, data access and data display capabilities. The discussion concerns system design principles and relative requirements; no attempt is made to determine specific parameter values for a mission.

The second study activity considers some aspects of the solution to the overall data management problem. The discussion focuses on the ground data handling facilities and the capabilities which they must provide. The key system elements are considered individually and their prime design considerations which are necessary to achieve the desired overall system capabilities are identified.

Finally some comments concerned with follow-up studies and future planning are presented. The recommended study areas are those which are considered to be crucial to achieve a cost effective data handling facility. The recommendations and observations on future planning are included to emphasize the very significant role of data management in a science orientated program, and its wide range of impacts throughout the program.

This report is primarily concerned with developing criteria for classifying or categorizing the final output from the data management system in terms of response time, quantity of data, type of processing required, purpose of displaying the data and type of display. Our previous experience in the data management field indicates that this kind of classification of requirements has frequently been overlooked in the past, indeed many experimenters seen to regard their experiment as a "special case". We believe that analysis of the user requirements in these terms is necessary to provide a sound foundation for the data management system design. We re aware that there is a wide range of significant topics which arise mediately after the foundation has been established and which are not considered in this report. In particular on-board processing and displays, and the identification of all the subsystems necessary to provide a cost effective data management system are not discussed in this report.

DEFINE THE DATA MANAGEMENT PROBLEM

The data management facility must be capable of satisfying several distinctly different objectives during a mission. An effective system is one which will satisfy all the checkout, mission control and science analysis requirements as the mission progresses. It must be clearly recognized that we are concerned with a facility which is inherently required to operate in multiple modes.

Response time is the most important single parameter in system design and selection of system elements. By response time we mean the permissible delay between data acquisition in the experiment and the production of reduced output by the data management system. Reducing the response time raises the system cost. The response time requirements vary widely, and an effective system must satisfy the range of requirements.

The data processing may be performed either on-board or on the ground. he choice of on-board versus ground processing for each data management mode has wider system and mission manning implications which must be recognized.

The data access problem is crucial, it is a significant problem because of the multiple objectives and multiple users of the data. Data storage elements and the data access interfaces are major cost items. Access techniques in previous mission programs minimized the quantity of data supplied to each user. Modern data transmission facilities have improved to the point that minimizing the quantity of transmission is no longer a prime system goal.

Display requirements vary widely, the problem is to select cost effective display elements which provide the necessary flexibility. In this report we classify the displays by their function and emphasize the display system design parameters which are applicable to each function. Many special purpose displays have been provided by users in the past which have restricted application. Smart terminals incorporating a local computer and CRT display have been developed for commercial data processing systems and it appears that these possess the flexibility to satisfy the differing design parameters for a wide range of applications.

THE OBJECTIVES OF DATA MANAGEMENT

The data management facility provides an essential service during several different mission phases. This service is to store, retrieve, reduce and display all the relevant data to permit the performance and science output of each experiment to be analyzed and understood. The phases are:

1. Checkout control and evaluation;
2. Operations control;
3. Science analysis and experiment evaluation;
4. Historical mission documentation.

The performance of the data management facility can be properly assessed only in terms of its value to the experimenters and other users. The objectives of data management are to satisfy the user requirements in all four mission phases.

The response time, data access, data reduction and display requirements are different for each phase. Hence the data management facility must be designed to satisfy several distinct objectives and must be capable of performing adequately over a wide field of parameter values.

RESPONSE TIME

By response time we mean the permissible delay between data acquisition in the experiment and the production of the reduced output by the data management system. The response time requirement is primarily determined by the specific objective which is to be achieved. Due to the inherent differences between experiments the required response time for any mission phase will be experiment dependent. In Table 1, the response time range is shown for each system objective.

The response time is of prime importance in selection of system hardware. It determines the type of subsystem which is appropriate to each ınction as shown in Table 1. At a more detailed level, response time will ɔe a significant parameter in the choice of every system element.

ON-BOARD VS GROUND PROCESSING TRADE-OFFS

In many instances the data processing could be performed either on-board or on the ground. Ground based operations are supported by teams of scientists and engineering specialists. In order to provide equivalent support to on-board operators, in general more sophisticated subsystems are required on-board than for ground processing. The trade-offs applicable to various mission phases are summarized in Table 2.

DATA ACCESS TECHNIQUES

Obtaining access to the experiment data has been a continuing problem throughout the space program. Much of the difficulty arises from the multiple objectives and multiple users of the data which the data management facility must support.

The quantity of stored data which must be accessed is different for each mission phase. Operations control requires access to a relatively small quantity of stored data. Experiment evaluation, which includes troubleshooting, equires access to a significant quantity of stored data. Science analysis must access the bulk mission data.

TABLE 1

RESPONSE TIME	OBJECTIVE	TYPE OF SUBSYSTEM
1 SEC	PAYLOAD CONTROL, CHECKOUT OR OPERATIONAL	ON-BOARD, AUTOMATED
10 SEC TO 10 MINS	PAYLOAD CONTROL, "REAL-TIME" CHECKOUT OR OPERATIONAL	ON-BOARD OR GROUND MANUAL OR AUTOMATED
10 MINS TO 10 HOURS	EXPERIMENT EVALUATION CHECK-OUT OR OPERATIONAL	DISPLAYS, ON-BOARD OR GROUND
1 HOUR TO 10 DAYS	SCIENCE ANALYSIS	DISPLAY OR REDUCED DATA
END-OF-MISSION	SCIENCE ANALYSIS	OFF-LINE DATA REDUCTION
END-OF-MISSION	ARCHIVAL DATA STORAGE	DATA BANK RETRIEVAL SERVICE FOR OFF-LINE DATA REDUCTION

TAL 2

ON-BOARD VS GROUND PROCESSING TRADE-OFFS

ON-BOARD	GROUND
CAN PROVIDE FAST AUTOMATED CONTROL	NOT SUITABLE
REAL-TIME CONTROL REQUIES TRAINED OPERATOR AND/OR CONVERSATIONAL COMPUTER	SUITABLE FACILITIES MORE EASILY PROVIDED AT A GROUND BASE
DATA RECALL FOR REAL-TIME CONTROL REQUIRES EITHER STATUS CHANGE HISTORY OR WRITE/READ DATA RECORDER	PROVISION OF ON-BOARD STATUS CHANGE ANALYSIS CAN SIGNIFICANTLY REDUCE DOWNLINK DATA RATE.
ON-BOARD EVALUATION INHERENTLY LIMITED IN SCOPE	EXPERIMENT EVALUATION REQUIRES P. I. COOPERATION/INTERVENTION
SCIENCE ANALYSIS WRITE/READ DATA RECORDER MANDATORY	SCIENCE ANALYSIS TO MAXIMUM DEPTH NECESSARILY PERFORMED OFF-LINE

There are two possible approaches to providing a facility to satisfy the three requirements. These lead to contrasting system design philosophies. These are:

1. Attempt to optimize three separate services for the three modes of access described above;

2. Provide a unified service which will satisfy all three modes access requirements and which can service all the different users.

The advantage of providing three separate services is that this permits the quantity of data flowing through the interface to each user to be minimized. The disadvantage is the high cost of designing and implementing numerous specialized interfaces.

Historically, the problems associated with transmitting large quantities of digital data were so severe that the separate data access services approach was adopted. Modern developments in data transmission facilities have removed many of these limitations so that minimizing the quantity of transmission is no longer a prime system goal. The exception is certain special missions, e.g. EOS, which will produce extraordinarily large quantities of data which will have to be handled in special facilities.

The unified service will require both a data storage and a data recall and transmit facility. An adequate service will require a large subsystem which will be a significant portion of the overall data management facility.

We recommend that a design study to assess the technical problems, costs and capabilities of a unified access system, e.g. based on disc and tape storage with data distribution via commercial data links, be performed.

DISPLAY TECHNIQUES

The displays which are used to present reduced data to the mission controllers and experimenters are most appropriately classified by the function by which the data is being reduced.

The three data reduction functions are:

1. Experiment control, which typically involves engineering data conversions, status decoding and data change detection;

2. Physical parameters measurement, i.e. the measurement of steady or slowly varying parameters;

3. Event display, primarily science events, which requires the display of rapid, and relatively rare, changes in the data.

The data processing and the displays which are required are significantly different for each function:

1. Experiment control involves minimal processing, fast response and read-at-a-glance displays;

2. Parameters measurement, whether this is based on a single observation or an observations over a long time period, requires extensive data reduction;

3. Event display is predicted on event detection, by data analysis, and implies display of time functions.

The requirements to access stored data also vary widely:

1. Experiment control requires minimal storage, the prime need is for before-and-after command readouts;

2. Parameter measurement requires a medium quantity of data, although for long term averaging the acquisition period may be considerable;

3. Event display requires massive data storage, so that once an event has been detected the data may be recalled for analysis and display.

Traditional display techniques do not implement the differing display system design parameters, i.e. response time, processing, event detection, data storage and display formatting, independantly. Hence there is a strong tendency for each user to design a special display to provide the display system parameters required for his own experiment. These special displays are based on special hardware which is generally neither available nor suitable for use with any other experiment.

In order to illustrate these points, consider the use of strip chart or drum recorders for displaying either science or engineering data:

1. Mass storage at limited resolution is available;

2. Fast access to recent data is provided;

3. Data reduction capability is limited according to the hardware utilized, e.g. some data correlation can be achieved by using multiple pens, and frequency domain shaping can be accomplished by using analog filters in the amplifiers;

4. Event detection is achieved by eyeballing the trace;

5. There is little flexibility in the data format and no data tabulation facility is available;

6. Control by skilled technician is required;

7. The cost of providing technician support is high, especially on a 24 hour per day basis.

The capability of the analog recorder should be compared to that of a smart terminal incorporating a local computer and CRT display;

1. Mass data storage must be provided separately;

2. Fast access to recent data is available;

3. Powerful data reduction capability can be incorporated;

4. Automatic event detection and reporting capability can be provided in software;

5. Great flexibility is available in the display formats, at full data resolution and with hardcopy option;

6. Control is by the user keyboard;

7. Operating costs are low.

SUMMARY-DEFINING THE DATA MANAGEMENT PROBLEM

The task of providing an adequate definition of the data management problem is more complex than would appear at first sight. We submit that the difficulty is twofold; first, we have to design to satisfy multiple objectives corresponding to different mission phases; second, we must identify specific system parameters which define the hardware functional requirements. These parameters are response time, quantity of data, type of processing required and type of display. Different values of these parameters are applicable to each mission phase.

In searching for a solution to the data management problem we must try and satisfy the wide ranges for the parameters in a cost effective fashion. This implies that standardization, flexibility and growth potential will all be significant criteria in selecting subsystems.

The reader will appreciate that the approach to defining the data management problem taken in this report implicitly rejects attempts to build a system by combining in a "side-by-side" fashion the processing and displays for individual experiments. This is because the fundamental purpose of the data management facility is to provide human oriented, easily intelligible displays. We consider that this singular purpose imposes a degree of commonality on the system design requirements which far outweights the technical differences between experiments.

CONSIDER SOLUTIONS: DEFINE THE SYSTEM CAPABILITIES AND ELEMENTS

A system block diagram is presented in Figure 1. This diagram brings together the various functions of the data management system which have been discussed above; it is not intended to be used to identify subsystems in the data management facility. This block diagram provides an introduction to the discussion which follows of the capabilities of the key system elements.

These elements must be designed to provide the required overall system capability. To achieve an adequate system capability the following design features should be incorporated:

- Provision of an automated data access subsystem

- Standardized user data access interface

OVERALL SYSTEM BLOCK DIAGRAM

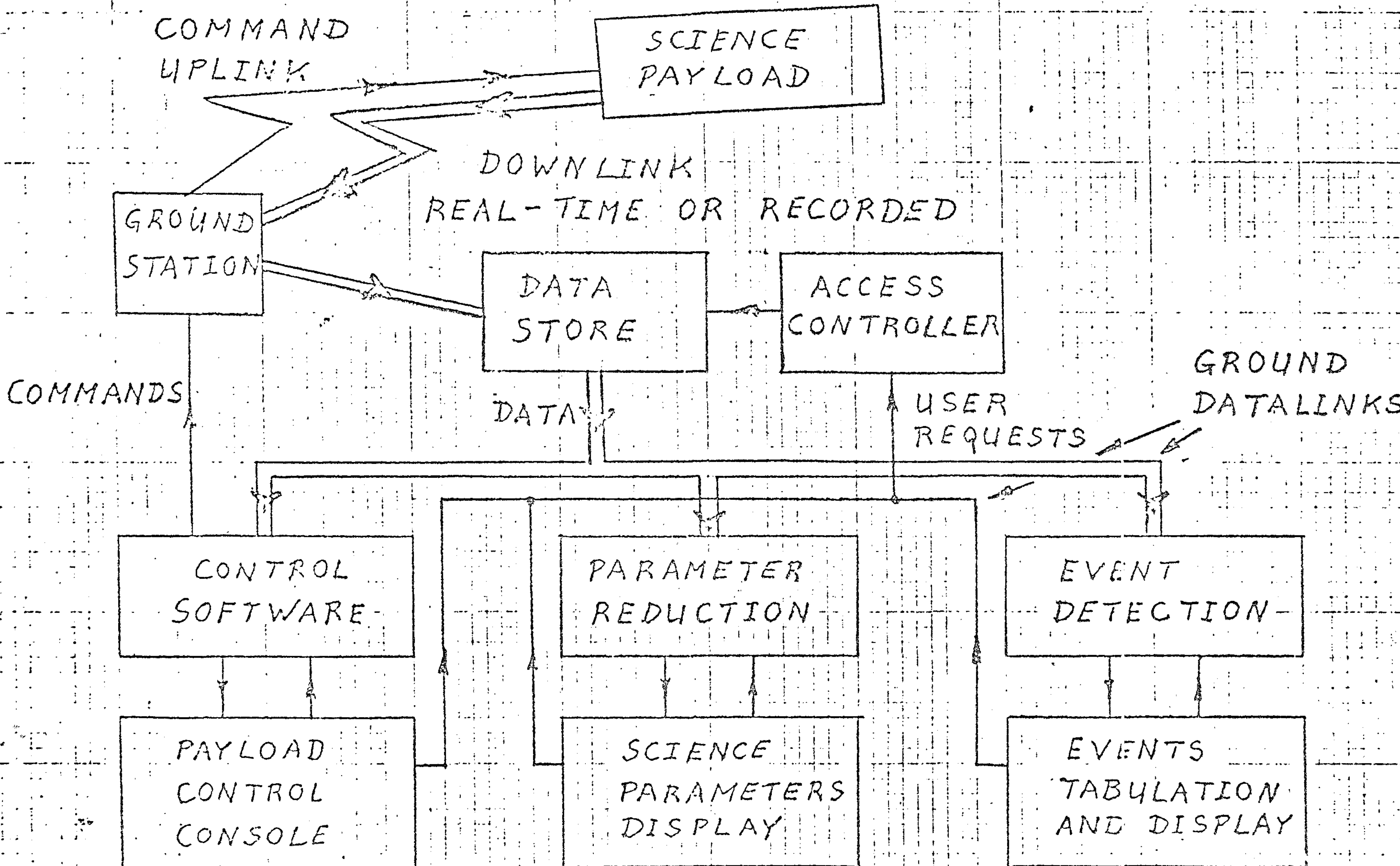

FIGURE 1

- Operation in all three data analysis modes - Control; parameter reduction; event detection and display.
- Ability to tailor display system design parameters to user requirements
- Standardization of display hardware for all modes of operation and for multiple users
- Multi-user flexibility provided by software. User-programmer cooperation required to obtain full benefit of this feature.

THE DATA STORE ACCESS CONTROLLER ELEMENTS

The most practical, economical and readily available storage elements are computer compatible digital mag tape and disc stores. A disc can provide a fast access store for payload control or a relatively small store of recent data. Typical single disc packs store 10^6 to 10^7 bits, and their typical access times are 30 to 100 millisecs. Mag tape provides a mass storage capability with typical access times of 1 to 5 minutes; a single tape can store up to 3×10^8 bits. For reference, an ALSEP system produces approximately 10^8 bits per day.

Control and access to such a store requires a powerful computing system. To provide service to more than one user simultaneously a data interface would be required for each user. Multiple interfaces must be provided to handle the potentially high peak output data rate.

The incoming data from the payload would be put into the store by the computer, and at the same time the computer would prepare a catalogue which would permit automated data retrival. Data access involves:

1. Receive request from the user;
2. Search catalogue to locate data;
3. For data from mass store, mount tape;
4. Search tape or disc for the specified data;
5. Output the required data to the data link.

Fast response to user requests requires sufficient computing power to perform the searches without interrupting data transmission.

The access controller is an on-line computer. Its prime specification parameters are:

1. Number of users which can be serviced simultaneously;

2. The response time to user requests.

Modern computers handle input/output data rates of the order of 10^7 bits/sec. For the majority of payloads, the input data rate is small compared to the computers' capability and is a secondary consideration. Similarly, service to a very large number of users would have to be provided before the output data rate becomes significant.

GROUND DATA LINKS

Data rate is the prime capability of ground data links. Currently, low-speed links typically operate at 2.4 Kbps, and medium speed links are available in the 5 to 10 Kbps range, e.g. 7.2 Kbps is generally available. In the 1980's, high speed links at 50 Kbps should be commonplace. Due to the continuing development of ground data link facilities the cost of transmission of bulk data compared to the cost of special formatting to provide data compression is continually changing in favor of bulk data transmission.

The data rate from the mass data store to the displays is that of the ground data link, and is independent of the payload to ground data rate. Clearly, any user whose display requirements exceed a currently available date rate capability imposes special purpose data reduction/data display requirements. These might be satisfied either by data reduction prior to transmission, or by the provision of a special purpose hardware subsystem. Hence we consider that a plan to upgrade the ground data link data rate capability in the forseeable future is an integral part of a sound system design.

SOFTWARE

Software is required for experiment control, experiment evaluation and science evaluation, regardless of whether these are run in a single computer or performed in discrete facilities.

There is a substantial correspondence between operational control and payload checkout. We recommend that as policy a single software package should be developed for both functions. Correspondingly, the system test facilities should mimic the operational data access procedures, hence the test facilities cannot be implemented before the operational system has been specified.

Note that in these remarks we are implicitly distinguishing between data access as a function and the data access hardware. In the system we are discussing, the access function has been implemented discretely and not as an implied function of a specific ground station or any other piece of hardware. It is this feature which permits commonality of the software modules to be implemented.

Science parameter data reduction normally involves either correlation between data channels and/or over time periods. The processing is experiment specific. The prime requirement to achieve cost effective operation is the integration of the users programs into the NASA facilities.

Science events can only be detected on the basis that data values have changed by an amount which is statistically significant compared to the channel noise level. We believe that general purpose programs can be written to perform the statistical tests, and these tests can also implement other criteria such as the predicted probability of physical events.

Display formats are inherently user oriented. Tabulation, graphical, histogram and mapping displays are applicable to many experiments, and usually more than one form of display is required for an experiment. We consider that the capability for the users to readily format the display to be an important requirement for system design.

DISPLAY HARDWARE

The requirement for flexible multi-user, multi-purpose and multi-format displays cannot be met by any form of hardware alone. Adequate flexibility can only be achieved by a hardware/software solution.

Standardization of the data access interface will permit displays with differing capabilities to be installed by different users.

The prime specification requirements for a display installation are:

1. Number of users to be serviced simultaneously, i.e. the output display capability; and

2. Peak input data rate from the data access subsystem.

These requirements specify the number of CRT displays and the number of input data channels and their data rate. These are the specialized hardware items of the display installation.

The display drive computer can be selected to suit the specific display installation. The conceivable requirements range from a mini-computer primarily performing the display housekeeping chores for a simple single experiment display to a powerful centralized computer at a Mission Control Center containing several display stations. The potential display requirements cover a very wide spectrum; the flexibility and growth capabilities of the displays can be enhanced by selecting computing equipment which:

1. Implements display functions with plug-in modules;

2. Is compatible with back-up high speed memory;

3. Interfaces with disc and/or mag tape mass storage;

4. Is capable of driving additional output devices such as hardcopy printers, X-Y or other graphics plotters.

FOLLOW-UP STUDIES

A wide range of design studies is required to establish all the details of the data management system. We have selected the areas which appear to have the greatest impact on system design in preparing the list below:

- Data Access Techniques; Design Studies;
  Establish the maximum input data rate limitation;
  Establish the simultaneous users criterion;
  Establish the response time criterion;
  Specify the disc data storage requirement;
  Establish the number of tape drives required;
  Computer selection;
  Establish the output data rate limitation.

- Ground Data Links; Economics Studies: Predict costs as a function of data rate and transmission distance in the future.

- Event Detection Software Design: Specify suitable generally applicable algorithms.

- Cost effective display hardware design studies.

- Payload-to-ground data link optimization study.

- Plan to increase the flexibility of the users programmers interaction with NASA.

- Establish the extent of the support of the standard displays to image data presentation.

RECOMMENDATIONS/OBSERVATIONS ON FUTURE PLANNING

Data management concepts are of concern at every stage of a science payload program. This report primarily summarizes the objectives and basic technical requirements for a cost effective data management facility which will provide the capability to optimize payloads operation.

This facility will have a major impact on mission management and planning. In this wider context, some pertinent considerations are:

- Emphasize importance of data management due to impact on shuttle payload capability, science return and economic factors.

- Multi-user multi-usage requirements can only be satisfied by powerful and flexible data management facilities.

- Plan for long term program

- Bracket payload requirements to allow systematic development of data management facilities.

- Data management techniques are universal i. e., not payload specific.

- Design for cost effectiveness

- Engineer payloads and data management system to achieve compatibility.