NASA Conference Publication 2343
Part 2

Planetary Data Workshop

Proceedings of a workshop
held at Goddard Space Flight Center
Greenbelt, Maryland
November 29-December 1, 1983

NASA

NASA Conference Publication 2343
Part 2

Planetary Data Workshop

Hugh H Kieffer, *Chairman*
NASA Office of Space Science and Applications
Washington, D C

Proceedings of a workshop
held at Goddard Space Flight Center
Greenbelt, Maryland
November 29–December 1, 1983

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1984

TABLE OF CONTENTS

Page

1. INTRODUCTION 1
   1.1 Introductory Statement 1
   1.2 The Current State of Planetary Science Analysis 3
   1.3 A Search for Solutions 4
      1.3.1 The Pilot Planetary Data System 5
      1.3.2 The Planetary Data Survey 5
      1.3.3 The Planetary Data Workshop 6
      1.3.4 A Planetary Data System (PDS) 7

2. PLANETARY DATA DEFINITION 9
   2.1 Introduction 9
   2.2 Downlink Data Flow 9
   2.3 Data Content Requirements for the PSD 13
      2.3.1 Levels of Processing 14
      2.3.2 Completeness 14
      2.3.3 Data Compression and Reliability 14
      2.3.4 Longevity and Accessibility 15
      2.3.5 Update and Change 16
      2.3.6 Documentation 16
      2.3.7 External References 17

3. USER REQUIREMENTS 18
   3.1 Introduction 18
   3.2 Uses and Users 18
   3.3 Data Requirements 20
   3.4 Data Access Requirements 20
      3.4.1 Catalog 20
      3.4.2 Examination 21
      3.4.3 Acquire 22
   3.5 Processing Requirements 22
      3.5.1 Data Manipulation 23
      3.5.2 Statistics 23
      3.5.3 Registration/Resampling 23
      3.5.4 Graphic and Image Display 23
      3.5.5 Analysis Tools 24
      3.5.6 Recalibration 24
   3.6 User Environment 24
      3.6.1 Uniform Appearance 24
      3.6.2 On-line Help 25
      3.6.3 Off-line Help 25
      3.6.4 Resource Requirements Forecast 25
      3.6.5 Display and Hardcopy 25
      3.6.6 Large Data Requests 26
      3.6.7 Software Transport 26

4. CONFIGURATION AND IMPLEMENTATION 27
4.1 Introduction 27
4.2 System Structure Considerations 27
4.2.1 Central Management Center 29
4.2.2 Discipline Centers 29
4.2.3 Communication/Distribution Network 31
4.3 Management Considerations 33
4.3.1 Science Advisory Committee 33
4.3.2 Management Center 33
4.3.3 Discipline Centers 33
4.3.4 Periodic Review 34
4.3.5 Resource Allocation 34
4.3.6 Mission Interface 35
4.4 Implementation Plan 35
4.4.1 Prioritization of Activities 35
4.4.2 Scientist Involvement 36
4.4.3 Recovery of Existing Data Sets by Proposal 36
4.4.4 Revised Flight AO Process 37
4.4.5 Competitive Proposals for Participation in PDS 38
4.4.6 Incorporation of Existing Facilities ... 38
4.4.7 Use of New Technology 38
4.4.8 Standards and Common Software 39
4.5 Funding Issues 39
4.6 Implementation Schedule 40

5. IMAGING DATA 42
5.1 Introduction 42
5.2 Elements of a Data Set 42
5.2.1 Digital Imaging Data 42
5.2.2 Supplimentary Experimentors Data Record (SEDRs) 43
5.2.3 Picture Catalogs 43
5.2.4 Special Data Sets and Catalogs 44
5.2.5 Ephemeris Data 44
5.2.6 Photographic Products 45
5.3 Data Description 45
5.3.1 Voyager 46
5.3.2 Viking Orbiter 48
5.3.3 Viking Lander 50
5.3.4 Mariner 10 51
5.3.5 Mariner 9 52
5.3.6 Mariner 4, 6 and 7 54
5.3.7 Lunar Data 54
5.3.8 Psuedo-Imaging Data 55
5.3.9 Imaging Data of the Future 56
5.4 Locations of Imaging Data Sets 57
5.4.1 JPL 57
5.4.2 USGS - Flagstaff 57
5.4.3 Johnson Space Center 58
5.4.4 Foreign Image Center 58

5.5 Archiving and Preservation of Image Data 59
5.5.1 Urgency and Nature of Rehabilitation Activities 59
5.5.2 Archiving Priorities 60
5.5.3 Rehabilitation of Digital Image Data ... 60
5.5.4 Recommendations for Film Products 62
5.6 Imaging Requirements on the Planetary Data System (PDS) 63
5.6.1 Use of Existing Data Centers 63
5.6.2 Planetary Scientist User Requirements .. 65
5.7 Summaries and Recommendations 69
5.7.1 Summary of Requirements 69
5.7.2 Summary of Open Issues 70
5.7.3 Recommendations 70

6. NON-IMAGING REMOTE SENSING DATA 72
6.1 Criteria for Basic Observational Data and Catalogs 72
6.2 Enhancements to Basic Observational Data and Catalogs 74
6.3 Present User Community, Data Sets, and Archival Facilities 75
6.4 Entry into the Data Facility; Nature of a Request for Data 76
6.5 Role of the Data Facility in Data Analysis 77
6.6 Configuration of the Data Facility 78

Appendix 1: Pioneer Venus Orbiter Cloud Photopolarimeter Imaging 81

Appendix 2: Existing Data Sets 82

Appendix 3: Pioneer Venus Orbiter UVS Data Facility (D. Anderson) 84

7. IN SITU DATA 86
7.1 Introduction 86
7.2 The Nature of In-Situ Data 86
7.3 The Need for Processed Versus Raw Data 87
7.4 Ancillary Data 87
7.5 Key Parameter Data Sets 87
7.6 Means of Data Access 88
7.7 Levels of Accessibility of Various Types of Data 88
7.8 Data Manipulation/Analysis 88
7.9 Quality of the Existing NSSDC Data 89
7.10 Rules of the Road for Users 89
7.11 Standards for Future Data Submissions 90

8. RADIO SCIENCE 91
8.1 Nature of the Data and Data Sources 91
8.1.1 Interest Areas Within Planetary Radio Science 91

8.1.2 Basic Measurements and Calibrations 92
8.1.3 General Comments on Levels of Data 92
8.1.4 Data Attributes 93
8.2 Current Situation 94
8.2.1 Active Research Groups 94
8.2.2 Major Data Sets of Interest 95
8.2.3 Radio Astronomy 97
8.3 Description and Requirements of PDS 98
8.3.1 The Preliminary Inquiry 98
8.3.2 Nature of An Analysis Task 100
8.4 Summary, Recommendations, and Concerns 103
8.4.1 Major Recommendations for PDS 103
8.4.2 Rehabilitation of Existing Data Sets ... 104
8.4.3 Requirements of PDS 104
8.4.4 Continuing Concerns 105

9. EARTH-BASED PLANETARY DATA 107
9.1 General Nature of Data Available 107
9.2 Ground-Based Data 108
9.2.1 Imaging 108
9.2.2 Photometry and Polarimetry 110
9.2.3 Photometry and Polarimetry Data Sets Ordered by Technique 111
9.2.4 Spectrophotometry 112
9.2.5 High-Resolution Spectroscopy 113
9.3 Ballon and Rocket Data 114
9.4 Earth-Orbiter Data 114
9.4.1 Orbiting Astronomical Observatories (OAO-A2 and OAO-C) 114
9.4.2 International Ultraviolet Explorer (IUE) 115
9.4.3 Infrared Astronomy Satellite (IRAS) 116
9.4.4 Space Telescope (ST) 116
9.4.5 Shuttle Payloads 117
9.5 Ancillary Data 118
9.6 Requirements 118
9.7 Summary Recommendations 118

10. LABORATORY 120
10.1 Definition of Laboratory Data 120
10.2 Sources of Data and Data Volume 120
10.3 Required Support Documentation for Lab Data ... 121
10.4 Urgency of Building a Computerized Archive 128
10.5 Software Needed for Management of Lab Data 128
10.6 Analysis of Data 129
10.7 Location and Access of the Data Archive 130

11. INTRODUCTION TO THE TECHNOLOGY VOLUME 131
11.1 Overview 131
11.2 User Acceptance, Productivity and Psychological Factors 132
11.3 Executive Software 132
11.3.1 Operating System 133
11.3.2 Applications Executives 133
11.3.3 Integrated Systems 134

11.4 Technical Standards, Software Packages and Portability 134
11.4.1 Programming Languages 135
11.4.2 Device-Independent Graphics Software .. 136
11.4.3 Imaging Device Requirements 137
11.4.4 Software Portability 137
11.4.5 Standard Format Data Units 138
11.5 PDS and the Technical Environment: A Virtual System for PDS 138
11.6 Standard Elements 139
11.6.1 System Entry 140
11.6.2 System Tools 140
11.7 Non-Common Elements 141
11.8 Overall Recommendations 142
11.8.1 The Model 142
11.8.2 Phases of Implementation 143
11.8.3 Specific Recommendations 144
11.8.4 Strawman Schedule 146

12. DATABASE MANAGEMENT 148
12.1 Introduction 148
12.2 Databases and Database Management Define 148
12.2.1 Records, Files and Their Limitations .. 148
12.2.2 Databases and Database Management Systems 149
12.2.3 Database Organization 150
12.2.4 Accessing the Database 151
12.2.5 Protecting the Database 152
12.2.6 Distributed Databases 152
12.2.7 Database Machines 153
12.3 Some NASA Databases 153
12.3.1 Pilot Climate Data System 154
12.3.2 Pilot Ocean Data System 154
12.3.3 NASA End-to-End Data System 154
12.3.4 Solar Mesosphere Explorer Mission Database System 155
12.4 Perspectives on Planetary Database Management . 155
12.4.1 Suppliers 156
12.4.2 Users 156
12.4.3 Discipline Centers 157
12.4.4 Control Authorities 157
12.5 Issues of Planetary Database Management 157
12.5.1 Standards 158
12.5.2 The Data Catalog 160
12.5.3 The Planetary Database 164

13. SOFT-I - IMAGING PROCESSING SOFTWARE 167
13.1 Introduction 167
13.2 Levels of Image Processing Software 168
13.2.1 Level 1: Logging and Formatting 168
13.2.2 Level 2: Radiometric Correction 168
13.2.3 Level 3: Correction for Geometric Camera Distortions 170

13.2.4 Level 4: Geometric/Navigational Corrections 171
13.2.5 Level 5: General Software Tools 171

14. NON-IMAGING SOFTWARE/DATA ANALYSIS REQUIREMENTS 173
14.1 Introduction 173
14.2 Working Assumptions 173
14.3 Non-Imaging Database 174
14.4 Data Definition and Access 174
14.4.1 Browse/Quicklook 174
14.4.2 Search/Sort 175
14.4.3 Data View 175
14.5 Data Manipulation/Analysis 175
14.5.1 Calibration Software 176
14.5.2 Common Manipulation/Analysis Software . 176
14.5.3 Discipline/Specific Software 178
14.5.4 Instrument Type-Specific Sofware 179
14.6 Display Requirements 179
14.6.1 Browse Graphics 179
14.6.2 High Resolution Data Graphics 180
14.6.3 Interactive Needs 180
14.7 Implementation Phases 180

15. GEOMETRY SOFTWARE COMMON TO ALL EXPERIMENTS 185
15.1 Introduction 185
15.2 Geometric State 185
15.3 Implementation 186

16. NETWORKING 189
16.1 Requirements on Networking 189
16.1.1 Use of The Network 189
16.1.2 Required Functional Routines 191
16.2 Assumptions 194
16.3 Selection of Transport Method 195
16.3.1 Transport Media Options 195
16.3.2 Transport Costs 196
16.3.3 Network Terminal Costs 197
16.3.4 Selection of Transport Method 198
16.4 Selection of Transport Protocol 198
16.4.1 Minimum Required Software 198
16.4.2 Choice of Transport Protocol 199
16.4.3 Selection of Transport Protocol 202
16.5 Intermetting and Resource Sharing 204
16.6 Conclusions 206
16.6.1 Primary Functional Needs 206
16.6.2 Meeting Functional Needs 207
Appendix 1 Existing and Planned Networks 210
Appendix 2 Network Control and Services 214

17. PLANETARY DATA SYSTEM: HARDWARE REPORT 215
17.1 Introduction 215
17.2 User Requirements 215
17.2.1 Access to Data 215
17.2.2 Operations on Data 216

17.2.3 Local Data Storage 217
17.2.4 Data Display 217
17.3 Hardware Issues 218
17.3.1 Choosing Hardware 218
17.3.2 Hardware Futures 218
17.3.3 Maintenance 218
17.4 I/O Requirements 219
17.4.1 Magnetic Media 219
17.4.2 Optical Storage 219
17.4.3 Local Area Networks 219
17.4.4 Remote Communications 220
17.4.5 Hard Copy 220
17.4.6 Interactive Input 220
17.4.7 Image Display Characteristics 221
17.5 Work Station Characteristics 221
17.5.1 Hardware Categories 221
17.5.2 Performance 222
17.5.3 Benchmarks 223
17.6 Example Work Stations 223

APPENDIX A: ACRONONYMS AND CONTACTS FOR DATA DESCRIBED IN TABLE 4 224

APPENDICES TO RATIO SCIENCE 228

APPENDIX A: MEMBERS OF PLANETARY RADIO SCIENCE COMMUNITY 228

APPENDIX B: PLANETARY DATA SETS 230

APPENDIX C: JPL PLANETARY RADAR FACILITIES 251

APPENDIX D: INSTITUTIONAL HARDWARE 253

11. INTRODUCTION TO THE TECHNOLOGY VOLUME

11.1 Overview

The following chapters discuss technical aspects of the Planetary Data System (PDS). Several of the major topics addressed are listed below:

o Technologies and techniques which should be incorporated into the PDS.

o The extent to which existing software and hardware can be used and those parts of the system which must be custom-build.

o The level of effort required to develop the PDS.

Previous chapters have stated requirements that should be met by a Planetary Data System. The sum of these requirements can be combined into three points that are of paramount importance in designing such a system:

- o The PDS should facilitate access to all planetary data which is not under proprietary restriction, and that access must be sufficiently simple to allow relatively unsophisticated users to perform basic functions like determining which datasets are available and ordering portions of selected datasets. Access must also be uniform enough to promote interdisciplinary studies which necessitate use of different types of data stored at different centers.

- o The system should provide planetary scientists with enhanced data handling and analysis capabilities. Functions needed by a majority of users should be incorporated into a core system that can be made universally available; however, the system must remain open-ended to allow the addition of discipline-specific and user-specific features as needed.

- o The PDS must not rely too heavily upon specific hardware and software as the evolution of computer technology will render prematurely obsolete any system that is tied to the capabilities of current machines.

This chapter will outline the concept of a "virtual system" that can perform necessary PDS functions without being tied to the particular hardware and software that implements those functions. Also addressed are the software considerations which are relevant to most disciplines in planetary science. **Finally in the last section of this chapter a possible implementation plan for the PDS is outlined. Chapters 12-17 cover the technologies necessary to implement such a system. These include: Database Management, Chapter 12, in which current methods and tools for maintaining and accessing large, complex sets of data are discussed; Chapters 13 and 14 are devoted to the specific software and applications that will be needed for processing imaging and non-imaging science data; Chapter 15 discusses the need for specific software that provides users with imformation on the location and geometry of scientific observations and augments the chapters on science data processing;

Computer networks and allied topics on communication, including the user interface to the PDS and the methods for exchanging data within a networked system are covered in Chapter 16; Finally Chapter 17 discusses appropriate computer hardware available to the PDS, including low-cost workstations, array processors, and display devices.

11.2 User Acceptance, Productivity and Psychological Factors

A large-scale PDS would provide the scientist with capability for locating data, and a means to process and analyze that data. Any combination of software and hardware to accomplish this would have limited capacity and any operation would take finite time. There are trade-offs among system power, system speed, system cost and user satisfaction.

As human beings our judgement of system performance depends on other than strictly objective factors. Studies show that consistent response to input is preferred, rather than rapid but erratic response. Programs written to slow some responses artificially so that all fall within a narrow range of delay have been very popular among users. If long delayed response is necessary, users appear to require some evidence of activity, to assure them the system is still alive. An occassional message showing the progress of the request is sufficient.

The user's satisfaction with a system also depends on prior experience, performance expectations and perceived choice. A user accustomed to punch-card input may be happy with any interactive system, however slow, whereas one accustomed to personal computers will have much higher expectations. The science user may mentally calculate the apparent difficulty of a given operation and be satisfied with an hour's response for a geometrical transform, but be dissatisfied with an image display taking 30 seconds to appear. Finally, users will become accustomed to a certain performance level, given that there seems no practical alternative to the current method.

Improved scientific productivity is our goal, and psychological factors are a key component. We have limited resources, but psychological factors should be used to help allocate resources. In system design, effort may be put into software or hardware development and the trade-off will influence which operations will be favored. More practically, when a prototype system appears, users must be polled carefully and certain operations improved for better response.

A crucial role of the pilot programs will be to assist this effort, by determining thresholds of user satisfaction for various operations. The pilots should serve as a model, so that a prototype planetary data system will be "friendly" to users.

11.3 Executive Software

System software provides the environment in which the user accesses the capabilities of the hardware. It includes the operating system, programming languages and executive software. An important related topic is the utilization of standards in the development of application software.

11.3.1 Operating System

Of particular interest to the PDS community is the user interface to operating system functions such as file management, text editing and communications. Since PDS users will include computer novices as well as experts, we feel that a system in which the user needs to learn as little as possible in order to use this system will have the greatest possibility of success. It is anticipated that the user will work in an environment where different tasks are performed on various levels of workstations tied to a variety of hosts (see Chapter 17 for a description of workstations). To meet our goal of minimizing the amount a user needs to learn, it is desirable that the host interface be consistent from level to level. This is accomplished through the use of operating systems which provide user-definable commands, prompting for command input parameters and substantial on-line help facilities.

11.3.2 Applications Executives

There are two very different approaches to developing software. One is to write special purpose stand-alone programs and the other is to write special functions under a common executive. These two approaches are contrasted and the use of an executive is recommended as an aid to transportability and to provide a common user and programming environment.

a) The historical approach to writing applications software has been to create stand-alone programs to solve particular user requirements. Each program has its own unique user interface, deals with a fixed set of I/O devices, uses unique file formats and is targeted for a particular machine environment (hardware and operating system). These factors contribute to long software development times and high development costs, but may provide execution benefits in terms of more efficient use of machine resources.

Stand-alone software is not only expensive to develop, but the traditional approach has led to software which is not easily extended or transported. In some cases entire separate programs have been written to perform in batch and interactive environments. The software has generally been designed for an experienced computer user; as a result it has not provided help or tutorial capabilities to bring a novice user up to speed. The user has had to contend with both the native operating system in order to run applications software, and the programs' unique way of interfacing with him and the machine environment.

b) A modern approach to writing applications software is to write smaller programs which are encapsulations of algorithms, that run under a common executive. The user is presented with a standard way of running programs, entering parameters, and getting help when necessary. The novice uses menus to locate the function required and uses tutorial screens to enter parameters. As users become experienced they can run the same programs using command sequences. The command language allows the same programs to be run in an interactive or batch environment. The command language shields the user from the native operating system. Programs written under the executive call well defined libraries of routines to provide file access and virtual terminal capabilities. This provides an environment for writing transportable software and reduces software development time and costs.

The Transportable Applications Executive (TAE) developed at NASA-GODDARD is an example of a modern approach at providing an environment for software development. It has been designed to be transportable, but presently has been tested solely on DEC (Digital Equipment Corporation) hardware. It is presently used as the user interface for MIPL (Multimission Image Processing Lab) and the Pilot Climate Project. Among it's strengths is a very good parameter processing and parameter help facility. Among its drawbacks is the fact that it is written in C (no standards, not available for many machines, see below), and that the image size is large (about 15 megabytes of disk space).

UNIX is a Bell Labs product. It is noted for its good program development environment, for its many test manipulation programs, and for its terseness. The image size is somewhat smaller than TAE (TAE will soon run under UNIX), and it has been transported to many types of CPUs, with varying degrees of compatibility and support. In all but a few cases, it has a file system which is incompatible with the standard operating system on the computer.

Other examples of applications executives are IRAF from KPNO (Kitt Peak National Observatory) and APIS from NRAO (National Radio Astronomy Observatory).

11.3.3 Integrated Systems

In the last few years software products that integrate Data Base Management Systems (DBMSs) with other common types of software have been developed for small computers. Typically these products integrate a DBMS, and spread-sheet, word-processing and graphics software. Data can move easily between these components and the user has a single interface to all the functions. Usually each function communicates with the user through a separate "window" on his CRT display. For example, a user might write a report using the word-processing package in one window, leave that window and go to another where he extracts information from the database and converts the data to figures using the graphics software, and then go back to the first window to insert the graphics product into the report.

Examples of commercial integrated software packages for small computers include Context MBS and Lotus 1-2-3. The Apple Lisa computer is based upon a highly integrated environment. Other products, like Visi-On, allow users to integrate their own software.

There is certainly a need for similar integrated systems in science data processing. Integrated software systems called "geographic information systems" have already been developed that marry DBMSs with special software for classifying and analyzing spatial data like land resource imaging. But even most geographic information systems lack the degree of coordination available in the best personal computer packages.

11.4 Technical Standards, Software Packages and Portability

Standards are ubiquitous in data processing, from the width of computer paper to the international committee-designed graphics protocols. Standards are important: equipment from one manufacturer could not interface to that from another, and programs written on one computer could not execute on any other without standards. This importance makes standards difficult to arrive at; standards committees are notorious for taking years to agree, and some have broken up without doing so. Standards tend to codify a certain technical level and inhibit rapid introduction of new methods. We are all familiar with those "extensions" to standard programming languages that every manufacturer adds to their implimentation. Standards are rarely withdrawn, they are simply superceded by new developments embodied (eventually) in new standards. Each user must judge when to follow existing standards and when to deviate for reasons of cost or performance.

Software packages are collections of programs or subroutines designed to be used in many science analysis tasks. Most widely known are the scientific and statistical packages (IMSL, SSP, etc.) which are commercially available. Within a field (such as remote sensing) packages such as the Jet Propulsion Lab's VICAR image processing software are widely used. These packages free the analyst from writing much software. Often the user may simply combine routines from a package to accomplish a specific task.

Portability is the quality that permits software from one system to execute on another with as little change as possible. The use of standard languages is the first step. But scientists often require high performance, which leads to coding of routines in machine-specific languages, and to the use of special purpose peripherals such as array processors or display generators. Both choices lead to very non-portable programs, since computers have different machine languages and special purpose peripherals rarely have been considered candidates for standards. The only solution lies in making changes as simple as possible, by designing programs with many short subroutines each of which does one function only, so that (at worst) non-portable code is clearly segregated.

11.4.1 Programming Languages

A principal concern in anticipating PDS is the transportability of applications software developed by the user community. At the present time, nearly all scientific software is coded in Fortran, despite the popularity of PASCAL and C, especially in university environments. One of the major problems with the development of scientific software in Fortran is that once coded, the execution speed of many routines is inadequate and they are recoded in assembly language, and thereby become dependent on the hardware architecture of the host. Probably the most often cited strength of the C language is its transportability due to the fact that it provides the programmer with access to assembly level functions. The development and utilization of ADA by the Department of Defense will have a major impact on the computer industry and provide valuable insight to NASA on the directions it should take in the future. However, it is unlikely to have any short term effect on software development activities.

There are two areas where significant progress could be made in developing more transportable software. First, a requirement that a pure high level language version (no assembly code) program be maintained as assiduously as any streamlined versions would considerably reduce transportability problems. Second, translators capable of converting Fortran to "C" and vice-versa would allow software to operate on many more systems and be utilized, modified and upgraded by more programmers. This can be emphasized by considering that the cost of a compiler may account for 20fstation hardware. Therefore, there will be many systems with Fortran, C or PASCAL but few with all three.

11.4.2 Device-Independent Graphics Software

The computer graphics field, until a few years ago, had two de facto standards, the Tektronix PLOT - 10 calls, and the Calcomp plotter routines. For the most part, people used one of these two sets of calls (or emulated them); in some special cases, manufacturer specific software was used. These exceptions were acceptable because the devices were both somewhat rare and also costly, thus justifying the additional expenditure.

Current technology offers a vast array of imaging and color graphics products, suitable for applications that range from "quick look" display stations to very high power imaging work stations. The field is such that some unifying principles must be found, to prevent the cost of supporting these devices from becoming overwhelming. Care must be taken that choices made today do not prevent the use of the more powerful and less expensive hardware devices that will become available in the future.

The burgeoning use of graphics and imaging, not only in the purely scientific fields, but in medicine, CAD/CAM, cartography, automated engineering, etc. has created the need for some unifying software standards. There are two standards widely discussed at present: GKS and CORE, which address the issue of device independent software. Both of these graphics standards were designed primarily for vector data, presentation and manipulation operations. The CORE standard derives from work by the ACM SIGGRAPH group, and is under consideration by the ANSI (American National Standards Institute) standards body. At present only the CORE standard is widely implemented. The GKS standard was originally a German DIN design, which is now a draft international standard before both the ISO (International Standards Organization) and ANSI standards committees.

GKS is the more modern of the two standards, has a better defined set of interfaces and calls and appears destined to be the standard of choice. It completely defines the set of subroutine calling sequence for FORTRAN and C, and has a well defined Virtual Device Interface (VDI) and a Metafile definition for image transport and disk storage. The VDI is an important concept: it defines a fixed interface for any program that wishes to talk to any device. The interface defines the generic set of device characteristics, and the back-end (or driver) maps these generic requests onto the specific device when the image is displayed.

Since the specific device characteristics need not be specified in the program, choice of device can be deferred until the image is to be displayed. This separation of program from the device ensures portability and flexibility, and allows new devices to be introduced in a straightforward way. A new device only requires a new driver that translates the VDI commands to the specific hardware in order to be incorporated into existing programs. The metafile also uses this same generic definition, allowing an image to be created, stored on disk, and then displayed at a later time.

Calls are defined that allow device characteristics to be determined as needed. The number of image and overlay planes, can be designed in a flexible way. Escape sequences are defined for access to any device specific routines that are not mapped by the normal calls, and this mechanism can be used to access such items as a video rate processor. These processors, which many of the high-end systems have, are sufficiently different that a the definition is not likely to be possible.

11.4.3 Imaging Device Requirements

Note that neither the CORE nor GKS standards deal particularly well with the problems of images and image data. They deal with vector images, color, line attributes, interactive devices, overlays, and picture segments very well; but do not have facilities for handling multiple image planes, raster rotations, and other functions associated with imaging operations. Run-length encoded data and pixel fill operations are supported, however.

The features covered by the GKS standard are well enough thought out and are useful as a model of device operation, so that several observatories are discussing a set of standard extensions to GKS that support imaging. These extensions are expected to deal with all issues (except perhaps the specialized video processors) in a way that is a compatible extension to the existing GKS standard. The video rate processors and other device specific extensions can be handled via the existing escape sequences. These image extensions should be carried to the ANSI and ISO standards committees once they have been settled on among the Astronomy community.

11.4.4 Software Portability

The issue of software portability must be addressed by any group that sees itself in existence even five years in the future. Major software projects are a large, and necessary, expenditure that must be protected like any other investment. Software must be derived in such a way that it is portable across operating systems. This ensures the ultimate longevity of the software as well as easing the transition across local operating system upgrades. Techniques that enhance portability are well established. They include:

a) Use of well designed, well structured, modular code;
b) Use of a standard commonly available language such as FORTRAN-77 or C;
c) Isolation of machine and operating system dependent code in a small set of interface modules;
d) Exclusion of system or implementation specific features from the body of the code;
e) Use of a table-driven architecture to allow new functions and devices to be easily incorporated.

Other issues will affect its portability in more general terms.

a) Documentation: Program design, installation, modification and use should all be well documented. How to fix it should be covered as well as how to use it.
b) Maintainability: Systems must be designed so that they can be maintained. Structured techniques, clean modular design and good documentation are the best ways to ensure this.
c) Device independence: Device independence should be obtained at both the terminal/user interface and at any display device interface. The GKS package provides image device independence; a terminal definition concept such as the Berkeley TERMCAP package can provide device independence for terminals.
d) Contractual considerations: One factor limiting the use of standardized commercial software is cost. Many of the institutions that will be interested in these systems will be academic/research oriented, for them costs of a few hundred to one or two thousand dollars for a system distribution copy are reasonable. Tens of thousands of dollars are appropriate for commercial customers who can expect to distribute the costs to their paying customers. This consideration limits the use of commercial packages as part of the system unless they support a multi-tiered pricing structure for academic and commercial customers.

11.4.5 Standard Format Data Units

The routine exchange of data can be facilitated by use of standard formats. One scheme designed to achieve this is the Standard Format Data Unit (SFDU) system. The SFDU is a unit of data that has been encapsulated by means of a globally interpretable primary label. The purpose of this label is to provide a means for global identification of the structure of the data unit. The primary label contains both control authority and format ID codes which direct the user to the data format description in a central data dictionary. This dictionary is maintained by the identified control authority, and contains descriptions of the formats in a standard data description language. The remaining structure of the SFDU is provided by the creator of the SFDU, containing additional data description and support labels as needed, and the data itself. Users will be encouraged to create data units in a modular fashion, drawing from a standard set of formatting structures, i.e., standard imaging labels, standard array formats, etc. These standard formatting structures will often be discipline specific, and the various disciplines are encouraged to generate such standards.

11.5 PDS and the Technological Environment: A Virtual System for PDS.

The Planetary Science Data System should not be built in isolation. Most users have existing data systems which frequently involve large investments in hardware and software. The PDS should make use of extisting hardware and software where possible. This will enable the largest possible number of researchers to use the system. This also will protect the large existing investment in hardware and software from immediate obsolescence and reduce the costs of implimenting the PDS. In many cases it is not possible to use standardized hardware and software, since the present system was chosen because of special abilities (e.g., high speed floating point calculations).

Even if it were possible to acquire standardized hardware and software for the PDS, this would not be advisable since it could only be done by acquiring the system from a single manufacturer. This will lead to the common problem with single source procurement and limit our ability to introduce new technologies into the system since advances have not been confined to a single company. To avoid having the PDS stranded by technological advance, the system must be able to accomodate change.

When one considers the potential complexity of the PDS, its need to accomodate evolving hardware and software, demographics, diversity, and data volume, one soon arrives at the conclusion that PDS must be implemented as a 'virtual' system. A virtual system is one in which the user interface is very stable, the software and hardware interfaces remain somewhat stable, leaving the details of implementation as flexible as possible. Thus, a discipline computer could be one computer or a collection of computer sites - but to the user it would appear as a single entity; a new plot package could be purchased for the system - but the calls from other routines would remain the same. Examples of systems which have consistent user interfaces in spite of considerable differences in implementation include FORTRAN, portable operating systems (such as UCSD-P and UNIX), and superset operating systems, such as MVS (which runs other (previous) operating systems within its environment).

The set of user and program interfaces which are given 'virtual' status must be chosen carefully. First, one needs to consider the time and effort for achieving agreement on the properties of the interfaces, guaranteeing easy transportability, and ensuring adequate 'hooks' for adaptability. Second, there are costs incurred in transporting to a variety of systems or hardware and the system must be tailored, hence it is probably not available as a commercial software package. Third, items which are included in the 'virtual' system are, by definition, not easily changed; such items tend to stifle innovation, and tend to be stranded by technological advance. To keep the programming and maintenance effort minimal and to promote adaptability and change, the set of user interfaces and standards which constitute the virtual system should be small. On the other hand, the set should be sufficiently inclusive to provide a satisfactory range of user services, a usable number of programming interfaces, and adequate capabilities and standards to permit design and implementation of the PDS system.

Each of the technology sections which follow describe certain standards which must be met and certain interfaces which must be transportable to various machines and/or software environments. The software sections require a standard set of graphics calls and a method of accomodating various graphics output devices. The DBMS section requires a reasonably standard access language. The network section requires standard addressing, transport protocol (for other protocols to interface to), and file-exchange protocols. The hardware section needs a standard model which diverse hardware can emulate to participate within the PDS.

11.6 Standard Elements

The Planetary Data System would encompass a wide variety of software, hardware, datasets, preferences, and operational styles. It is important that such a system have a uniform method for access to the system. Without a common access basis, the system would be far too complicated for the casual or mildly forgetful user. There would be a significant probability that difficulties in achieving system access would mask and displace efforts to achieve scientific progress.

The attributes of such a system are not difficult to define:

a) The access method must not change frequently (from a user viewpoint)
b) The access method should permit access to all elements of the PDS, and should provide access to several elements simultaneously.
c) The access method should permit running in 'native mode' (i.e., using the standard operating system) on any particular element of PDS.
d) The access method should provide a reasonably uniform method of data interchange via various media.
e) The access method should provide methods of insuring system integrity and of monitoring use of system resources.

11.6.1 System Entry

The requirements for monitoring system resources (retaining maps of resources, tracking resource availability and usage, controlling access) and the need to have a common drop point for the various elements (for mail and other centralized activities) suggests that there is a 'conceptual' central location. It would be undesirable to route all accesses through a central facility for reasons of system reliability and system throughput. It may be feasible to use 'discipline centers' for maintaining a common user access interface, and maintaining the requisite number of element interfaces (conversion to host computer requirements).

The discipline centers could refer to a central control or float that responsibility between them. From the user standpoint, access should be uniform. A single phone number (or data-line, or mailing address) should allow access to all elements. This single access point should provide information on resource availability, resource use, outstanding messages, and a user profile for the given user. The user should be able to change physical locations without changing a significant amount of his access protocol. Once signed onto the system, the user should have available a set of common tools for manipulating data.

11.6.2 System Tools

The system tools will be activated and controlled using the PDS executive. Like the executive, these services should be available on the computers that accept catalog queries and handle orders for data (so that access to the system is available with only a terminal modem and the proper passwords) and on workstations and other computer systems tied into the PDS.

Editor - A simple editor should exist for creating and modifying catalog queries, requests for data, mail messages, etc.

Catalog Access - Software is required to provide the means for querying and browsing a catalog of planetary datasets. Details on how this function might be implemented are covered in the Database Management chapter.

Database Access - Once datasets have been selected through catalog queries, a user may ask that part or all of the desired datasets be transferred. Software must be provided to allow the user to specify, or "order", the data needed. The user should be able to select he disposition of files created as a result of catalog and database searches. Options include network, magnetic tape, video disk or mailed printout.

File Transfer - Workstations and local computer systems will need the capability for transferring files back and forth. This provides users with the ability to share programs, for example.

Mail - An electronic mail system should be available for PDS users. It is possible that a commercially available electronic mail system can be used, otherwise a computer within the PDS will have to be designated to act as the clearinghouse for mail.

Help - The system should provide an on-line user's guide containing information on each major system function and diagnostic messages that explain to a user what he is doing wrong and how to correct the situation.

Break - Users need the ability to interrupt and cancel active functions.

Status - Users should have the means of determining the status of the PDS and the PDS's processing of their particular requests. For example, the system should be able to tell a user the status of any orders for data that he may have outstanding.

Format Conversion - A common problem in a large system is that data formats (for floating point numbers, etc.) are not uniform. Therefore the PDS should provide the means for converting data from one machine's format to another during data transfer. Although this type of conversion has traditionally been difficult, several systems exist or are being developed that provide this capability.

This set of system access and system tools software has been designated SESSION software for the purposes of this report. The user view of this SESSION software remains constant over time. The difficulty of implementing SESSION software is not that of writing (there are many possible existing systems which could be adapted), but that of agreeing upon a standard set among the community-at-large. Achieving this standard may impose one of the largest schedule impacts on PDS implementation. It is also vital for the creation of a smoothly functioning system, one that permits communication between investigators and provides a framework for design of common analysis software.

11.7 Non-Common Elements

The PDS will be a heterogenous environment: users will access the system using a variety of computer hardware and software. Several areas in which the PDS will include diverse elements are listed below:

Computers - There are a few types of computers that are very popular within the planetary domain but the PDS will not have the luxury of compatible hardware.

Operating Systems - Most types of mainframes and minicomputers have their own unique operating system. De facto standards are emerging but the PDS must still deal with a wide range of operating systems. There are, however, two operating systems -- VAX/VMS and Unix -- that are used by many in the planetary science community, and therefore the virtual PDS system should probably be implemented first for these two operating environments and then for other operating systems as the need dictates and money allows.

Data Management Software - Each operating system provides its own software for file and record management so that there are important differences between files and records from different types of machines. Similarly, no particular database management system (DBMS) is available for all the computers that will be found in the PDS. Therefore the system cannot rely upon particular data management packages: it must provide a virtual interface to the data system that is substantially independent of any particular implementation.

Discipline-Specific Software - There is a large body of software that is chiefly useful for data from a particular discipline. Examples include image manipulation software, preprocessing software, calibration programs, etc. It will typically be up to those who work within a discipline to develop (and hopefully share) this type of software.

Application-Specific Software - Includes all the software that the user elects to develop himself. Much of the data processing within the PDS will be dependent upon instrument characteristics, spacecraft data formats, etc., so there will always be some need for this sort of software

11.8 Overall Recommendations

An explicit model must be developed to provide a standard with which one may compare alternative implementations. The model used here is a synthesis of those proposed in various PDS meetings. It is used here to permit development of a sample implementation schedule.

11.8.1 The Model

This model assumes a completed implementation consisting of six discipline centers, a central catalog, and an administrative center. The planetary community (individual investigators) are connected to their appropriate discipline center. The discipline centers provide computer power for catalog searches, data storage, data display, and modest amounts of processing. They also provide review of new data entered into the system and review of the quality of catalog entries. The administrative center provides a central point of contact (for novice users), control of resource usage (password control), a central catalog, a network map, a network phone book, and a mail drop. It also provides the means for maintaining compatibility between discipline centers and for maintaining uniform catalogs.

The user installations are completely free-form with respect to hardware and software: choices are limited only by budget and incentives offered by the appropriate discipline. Discipline centers have uniform access software. No hardware uniformity is required except at the communications link level.

It should be pointed out that the discipline and administrative centers are simply logical constructs. There is no particular hardware reason for them to be physically adjunct or separated. In fact, even user computers could be viewed as either joint or separate entities - from a network standpoint there is very little difference. Most transport methods have costs independent of distance (mail and local telephone calls being notable exceptions), so the discipline and administrative entities are simply for conceptual convenience.

The purpose for these centers is to emphasize the need for a central control and to emphasize the need to maintain a reasonably small set of computers for which common software is actively maintained. It is neither practical nor desirable to attempt to create software which is truly transportable across all computers in the planetary community. Such a software leviathan would be impossible to maintain and would necessarily have a too-small set of convenience features.

The model assumes that the PDS facility(ies) would be developed from present capabilities. Its full implementation would use the PSCN net for most data transport activities. Connections to other nets (MILNET, TELENET, TYMNET, etc.) would be restricted to one (or a few) centers to reduce connection charges (all would have access through the network - merely the number of PADS or IMPS would be limited.

11.8.2 Phases of Implementation

Implementation should occur in 3 phases - a design and test phase (mostly as a PPDS effort), an implementation phase and an operations phase. A 5 year plan with phase I extending from year 1 to year 3, phase II from year 2 to year 4, and phase III extending from year 3 to year 5 has been assumed for planning purposes. This overlapped phasing permits operational testing during system development. The following paragraphs show an example of the level of detail that was used for the schedule. We follow the implementation scheme of Chapter 4.

The development phases for the networking example proceed in this manner:

a) Facilities are upgraded early in this phase to permit rescue of old datasets and to aid catalog creation. Participating computers are linked to create the "net"; and a set of user workstations are purchased early-on to encourage participation by the community. TELENET-like and ARPANET-like gateways and dial-up access are provided to the community to promote data exchange and use of remote data. The pilot data system software, which is fairly mature at this time, is utilized for a startup catalog and browse software set. The "net" is used to test various ideas concerning 'discipline centers'.

b) Missing or deficient software is identified and plans are made to replace or rework those elements required to support analysis activities. Necessary protocols are defined. Development projects are set up in conjunction with NASA, NSF and defense agencies for communication to increase participation. Identified needs at this stage include: standard protocol to link to PDS network, internet protocol, file formats, metafile format, messaging protocols, mail capability, and accounting procedures. The DBMS is developed from the PPDS model for a virtual system. Discipline centers are selected and their facilities upgraded, and more user workstations are supported. Discipline-specific software is developed in conjunction with other discipline centered activities.

This second stage contains the most difficult parts of the process — choosing the standards most desirable from the PDS point of view and gaining agreements from the community to commit to some standards.

c) Complete Implementation includes acquisition of those facilities which have not been implemented during the first two phases. At this stage, it must be assumed that the discipline and administrative hardware must be acquired and maintained, and that discipline centers are fully staffed. More user workstations will be made available to the community. Final development of the DBMS will also be achieved, and the development of discipline-specific software will continue.

This strawman implementation plan permits estimation of schedule and costs. It also provides a baseline for comparing other conjectured implementation schedules.

11.8.3 Specific Recommendations

a) Database Management Implementation Recommendations

The current Pilot Planetary Data System (PPDS) project will deal with many of the database management issues confronting the PDS. In particular, PPDS will develop standards for data within the system, a catalog of datasets, database software for use in archive centers and a virtual system interface to the catalog and database. Major milestones for PPDS data system development include:

- Development of data administration guidebook - preliminary version due Oct 1984, complete version in Sept 1985. Data documentation guidelines available March 1985.

- Catalog of all PPDS holdings available July 1985. Catalog augmented with a taxonomy, cross-reference and bibliography about June 1986.

- Archive center database systems should be available in mid-1986. The NSSDC facility has been used to rescue old data prior to this time.

- Virtual system interface to the catalog and database will be completed in early 1986.

The amount of work that will be required in these areas will be largely determined by the degree of success of PPDS. It is therefore difficult to provide a schedule for PDS database system development. It will be similarly difficult to estimate cost: if components can be inherited from PPDS then costs for development of the database system can be reduced substantially. It must be understood that the initial cost to set up the database system for PDS archive centers is going to be a major expenditure, and will run $200K or more excluding the cost of a general-purpose computer and peripherals. This substantial expense may prove to be an important factor limiting the number of archive centers.

b. Image Processing Software Implementation Recommendations

The Image Processing Software should be developed to meet the following recommendations:

- o A set of universal standard formats should be developed and used for past current and future image data collected by planetary missions. Old data should be reformatted.

- o These formats should be developed and used for the following imaging data types:

  Raw EDR data
  Radiometric calibration coefficients
  Geometric Camera Distortion Data
  Spacecraft and target ephemerides
  Image Data Catalogues
  Image Pointing and Geodesy Data

- o Software for access and processing of these six data types should be developed and made available to the user community. These software modules can be universal from mission to mission and should include: EDR access, radiometric calibration, calculation/removal of geometric distortion, calculation of the target coordinates (lat. and long.) and photometric coordinates (phase, illumination and emission angles etc.).

- o Certain complex software modules including those for geometric transformation (2 and 3 dimensional) of images, calculation of cartographic transformation matricies, perhaps photometric modelling and catalogue searches should be provided as a well documented, transferable software set.

c. Non-imaging Software Implementation Recommendations

Phase 1 - Startup central and discipline center catalogs and browse Software — possibly using Planetary Pilot, Climate Pilot or Ocean Pilot Data Systems software. Years 1-3.

Phase 2 - Clear needs identified in Phase 1 are addressed, including the user interface data access at system level (enhanced catalog inventory, search/sort capabilities) and data processing at discipline center and user workstation levels (enhanced stations, manipulation, graphics and discipline-specific analysis tools). Software standards are defined and enforced. Years 2-5.

Phase 3 - Needs arising from Phase 2 addressed, including full-up encyclopedia at system level, and further discipline-specific analysis tools at discipline center and workstation levels. Calibration software is brought up to process raw data. Years 3-5+.

11.8.4 Strawman Schedule

The development schedule for this phased implementation takes into account available technologies and reflects the schedule in Chapter 4. The activity schedule has been divided into four separable projects: data access (networking); data base management; hardware acquisition; and common analysis software. These projects in turn have been subdivided into tasks which are individually scheduled. This expansion of the schedule in Chapter 4, shown in Figure 11.1, enables one to make assumptions about the extent of the overall task and could support both manpower and cost estimates.

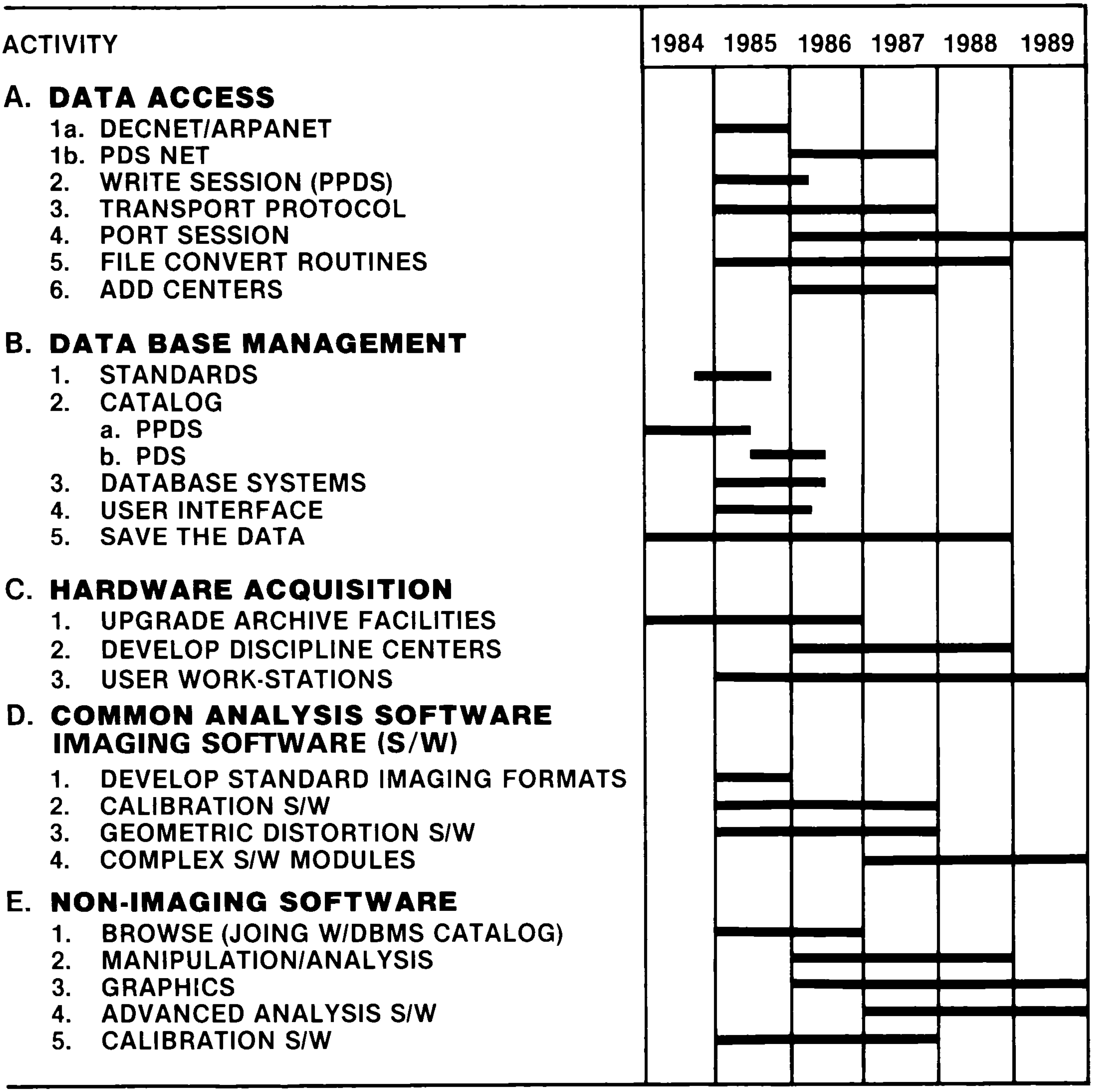

FIGURE 11-1 EXPANDED IMPLEMENTATION SCHEDULE

12. DATABASE MANAGEMENT

12.1 Introduction

This chapter of the PDS report addressed management of the data within a planetary data system (PDS). The chapter opens with a section that describes principles of modern data management and another that briefly examines several large NASA scientific database systems. The penultimate section outlines PDS data management and introduces the major data management issues. The final section discusses these issues at length.

12.2 Databases and Database Management Defined

This is an introduction to an important data management concept -- the database. This document addresses the specialized software for managing databases along with principles of database organization, access and protection. The unique problems of managing a distributed database are also discussed. This section ends with an examination of a new and important data management technology: database machines.

12.2.1 Records, Files and Their Limitations

Computer users are able to store and retrieve data without having to understand the complex way in which those data are actually represented on a medium like magnetic disk. They are freed from dealing with sectors and tracks on a disk because abstract representations of data storage have been devised, which are easier to comprehend and to use. In the most common abstract representation, familiar to almost all computer users, data are arranged into files and records. As we shall see, other representations are possible.

For every representation of data storage there must be data management software to translate between the abstract representation and the physical format. The software for creating and accessing records and files is called a "data management system". Data management systems are often supplied as part of a computer's operating system. While data management systems simplify the task of storing and maintaining data there are still some serious shortcomings:

a) Users must keep track of a myriad of details, including the names and locations of files, the length of records within a file, the order of data items within a records, etc.

b) It can be difficult to determine which data are stored in which files. In the absence of adequate documentation, it is often impossible for an outsider to determine the contents of a particular file without examining the programs that read and write the file (even that may not provide the answer if the programs are not well written).

c) It is difficult to determine the relationships between data in separate files. If two files each contain the same data item (attitude, for example) then the information within the two files can potentially be compared or combined. But it would be hard for our hypothetical outsiders to determine that such a relationship between files existed without examining documentation or program listings.

d) Programs are dependent upon the format of records and files they use. Altering the structure of records or files usually necessitates modifications to all programs that read and write them.

12.2.2 Databases and Database Management Systems

The problems listed above are particularly acute when dealing with large sets of data, when many people need access to the same data or when there are complex interrelationships between the data. There has been much work, both theoretical and practical, aimed at easing the task of maintaining large and complex sets of data. Most of this work is based upon the concept of the "database". A database is simply a collection of related data that can be managed and accessed as a whole. It can be nothing more than a coordinated set of files, but the full power of the database concept is best exploited when other abstract representations of data storage are substituted for the record/file model.

The software for creating and maintaining a database and providing high-level access to the database is called a "database management system", or "DBMS" for short. Aside from managing large amounts of data, DBMS' alleviate many of the problems outlined above:

a) Users are isolated from the low-level details of data storage. DBMS users typically to not need to know where or how their data are stored, how records are structured, or exactly how the data finds it way from the database into their programs.

b) DBMS' often include "data dictionaries" that define each data item and describe the layout of the database.

c) The relationships between data within the database can be easily expressed and exploited.

d) The organization of a database can be changed substantially without affecting programs that access it via a DBMS. For example, if the structure of an existing type of database record is altered by appending several new data items, only programs that require the new data items need to be modified; other programs, even ones that access modified records but do not use the newer items, will still run as they did before.

Interestingly, one of the first DBMS' -- IBM's -- was developed to track inventory and processing in NASA's Apollo program. But it is only within the last few years that NASA has begun to apply DBMSs to scientific data.

DBMS' are not supplied as part of a computer's operating system but are sold as separate software packages. DBMS', as with many computer innovations were once available only on large, expensive mainframes; now DBMS' are widely available for minicomputers and business and home microcomputers. The capabilities and prices of DBMS' vary significantly. The most popular personal computer DBMS — DBase II -- has a list price of $700. DBMSs for large minicomputers typically cost $10K - $50K. For mainframe computers the cost is about $100K - $200K.

Let us recapitulate some of the ideas presented above: A collection of related data is called a "database"; the software that maintains and provides access to a database is called a "database management system" (DBMS); and the entire process of creating and maintaining a database ? which consists of people, procedures, software and hardware ? is called "database management". Note that DBMS' are only a part, albeit and important one, of database management.

12.2.3 Database Organization

A database exists at many levels simultaneously. At the lowest level, called the "physical" or "internal" level, the database is a set of sectors and tracks on a disk. Often those sectors and tracks are arranged into records and files (in fact, many DBMS' rely upon a computer's standard data management system to maintain the physical database level).

DBMS users are not required to know much about the way the database is organized at the physical level. A DBMS isolates users from the details of the physical database by creating a high abstract level called the "conceptual" database. The most important aspect of the conceptual database is that it replaces files and records with a "conceptual model" that attempts to capture the essence of the interrelationships between data within the database. Theoretically there are any number of candidate conceptual models for a database, but in practice nearly every database managed by a commercially available DBMS is patterned after one of the following conceptual models:

a) NETWORK - The network model arranges the database into a simple directed graph. Each type of record is represented as a node of the graph and the links between nodes represent the relationships between records.

b) HIERARCHICAL - In the Heirarchical model a database is represented as an inverted tree, where each node represents a type of record and the children of a node are associated with the parent in some relationship. Relationships between records in both the network and heirarchical approaches are usually predefined by the database designer and explicitly stored within the database as pointers.

c) RELATIONAL - This model arranges data into "relations", which are essentially tables of information to which certain set operations can be applied. Each column of a relational table is reserved for a particular data item. Each row of the table, usually called a "tuple" (rhymes with "couple"), consists of a value for each column. The range of values that any particular column can have is called the column's "domain" (the domain of a latitude column, for example, would be the real numbers between -90 and +90). The relational model allows tuples from different tables to be associated whenever they each contain a column whose value is drawn from the same domain e.g., if two tables each contain a latitude column then an association can be created). Further, these associations can be created "ad hoc" by the users themselves.

Some DBMS' can present the database to a user as if it were designed soley with him in mind: the user sees only those data items that he needs and they are arranged and formatted just the way he wants them. This is accomplished by creating an additional level called the "external level". The external level consists of a set of "views" of the database. Each view designates a subset of the database that will be accessable to whoever uses the view; everything else in the database is off-limits. A user specifies the view to be used. The view in turn guides the DBMS to the proper part of the database and determines how data are formatted when they are returned to the user.

The overall plan for a database is called the "schema". Developing a good schema for a large database is a difficult task and the job is usually done by a specialist called the "database administrator" (DBA). The DBA builds and maintains the schema using a "Data Definition Language" (DDL) that is included as part of a DBMS. A DDL is solely for creating and modifying the plan of the database; it does not provide for manipulation of the data. Schema definitions written in a DDL are processed and all information on the form and content of the database is usually stored in a "data dictionary".

12.2.4 Accessing the Database

Each DBMS possesses a "Data Manipulation Language" (DML) which provides the capabilities for reading, writing, modifying and deleting portions of the database. For an applications program to access the database requires that the DML be embedded within a "host" data processing language like FORTRAN or PL/I. Usually this is done in one of two ways:

a) The DML can have a syntax that is a subset of the host language syntax. Usually such a DML consists of calls to a set of subroutines that perform all data manipulation. For example, Call DBWRITE (...) might be the DML command for entering data into the database.

b) The DML can have a syntax that is entirely different from that of the host language. A program containing this type of DML command must be run through a precompiler that translates the DML commands into code in the host language (precompiling essentially reduces this method to method 1, above)

Most relational systems provide a DML of the second kind and even allow DML commands to be processed outside the context of a host language. Users able to perform many tasks as a result, (tasks that do not require significant computation or data manipulation) without having to write a program in a host language. We often call such DMLs "query languages". Most query languages have an English-like syntax; for example, in the language SQL (usually pronounced "sequel") queries are expressed in the following form:

```
SELECT one or more data items
FROM one relational table or multiple associated tables
WHERE specified conditions hold true
```

In response to such a query, the DBMS finds all tuples that meet the specified conditions, extracts the values of the selected data items and prints them or stores them in a file.

Some query languages are clearly designed for interactive use rather than use within a host language. Such languages typically rely upon "user-friendly" techniques like menus and graphics. An example is IBM's QBE query language which asks the user to name the relational tables to be accessed and then draws a template of those tables on a CRT screen. The user fills in the template, selecting the data items desired from each table and specify conditions that control which tuples are returned. There are also a few DML processors that can handle queries in a "natural" language like English.

12.2.5 Protecting the Database

As much as 25% of the software code in a DBMS is dedicated to preventing accidental or malicious damage to the database. There are four major types of protection afforded by a DBMS:

a) SECURITY - Preventing unauthorized access to the database by restricting any given user to selected portions of the database and to selected operations on the available portion. Typically access is granted on a user-by-user basis by the database administrator.

b) INTEGRITY - Checking the accuracy and validity of data inserted into the database. For example, a DBMS may check to be sure that a value for latitude is between -90 and +90.

c) SYNCHRONIZATION - Preventing two or more users, accessing the database at the same time, from interfering with each other. For example, a DBMS precludes two users from simultaneously attempting to update the same record.

d) RECOVERY - Restoring the database to a known state after a failure. All recovery methods require redundancy, such as periodically backing up the database onto tape.

12.2.6 Distributed Databases

Most databases are centralized, meaning the entire database resides with a single computer system. Distributed databases are separated into distinct pieces with the pieces resident on geographically dispersed computers and connected by a network. Distributed databases can be implemented with each computer using a different DBMS to manage its piece of the database or a single DBMS controlling all pieces. A system consisting of separate DBMS' is called "heterogeneous" and a system built around a single DBMS is termed "homogeneous". There are currently very few DBMS' designed specifically for the distributed environment.

Most distributed systems attempt to make the distributed nature of the database transparent to users. There are two important types of transparency that distributed DBMSs try to achieve:

a) LOCATION TRANSPARENCY - The user should not have to know at which site any particular piece of the database actually resides. Requests for data ought be made in a form that is independent of the location of the data.

b) REPLICATION TRANSPARENCY - If certain data are available in more than one place, then the user shouldn't have to specify from which site to get the data: the DBMS should automatically access those data via the least costly path. When a user transfers data from another site to his local database, for example, future references to the same data ought to be satisfied from the local database rather than the original site (as long, at least, as the data at the original site have not changed).

Querying distributed databases presents special problems. In a distributed relational system a query that requires associating relational tables located at different sites will necessitate transporting the tables to a single site, so that, the query processor can work with all the tables at once.

12.2.7 Database Machines

There has been significant research in recent years into specialized hardware that can perform database management functions. Several such "database machines" are now commercially available. a popular example is Britton-Lee's IDM series available for DEC VAX and other computers. The IDM machines provide a general-purpose data management system and a relational DBMS. A database machine connects to its host computer and to one or more disk drives containing the database. When a user issues a query or requests data, the request is passed to the database machine which then performs the necessary operations required to locate and return the selected data.

A database machine may improve throughput in a computer system freeing the host computer for other things, but it is not true that a database machine will always perform better than a traditional DBMS. Careful analysis and modelling of each potential application is required before deciding whether to use a database machine or a DBMS.

Database machines are currently available for mainframes and large minicomputers at a cost of about $100K - $300K.

12.3 Some NASA Databases

It is instructive to see how previous NASA projects have designed database systems. The examples below include two of the "pilot" data systems developed with support from the OSSA's Information Systems Office, one system that can capture and store data at rates approaching 50 Megabits/second and a system that supports an active satellite. Beside these examples, many NASA flight projects are currently assessing or building database systems, including AMPTE, UARS and Space Telescope. Development of two new pilot data systems, the Pilot Planetary Data System and Pilot Land Data System, are also underway. Each of the systems discussed below uses a wide range of data management techniques, from simple file/record management to DBMS. It should be noted that each of these systems is a centralized system. NASA is just now beginning to deal with the unique problems of distributed databases. Probably the first project to tackle these problems in earnest will be the Pilot Planetary Data System.

12.3.1 Pilot Climate Data System

The Pilot Climate Data System (PCDS) archives information on the earth's climate gathered from NASA and non-NASA sources. The system is implemented on a VAX-11/780 computer at the Goddard Space Flight Centre. Data are provided by experimenters on magnetic tape. A detailed description of each dataset is entered in a standard format into a online catalog. Information on how to find each dataset on tape is stored in an online inventory. Both the catalog and inventory are managed by a commercially available relational DBMS called ORACLE. With ORACLE's SQL query language, users can search the catalog to determine which datasets are available and then search within a dataset for data with specific characteristics. Custom-built data access programs are provided to extract data from the climate database and put them into a standard format called a Climate Data File (CDF). Extensive software is provided for manipulating and displaying data stored in the CDF format. the Transportable Applications Executive (TAE), developed by GSFC, provides a friendly interface for PCDS users.

12.3.2 Pilot Ocean Data System

The Pilot Ocean data System (PODS), like the PCDS, is funded by OSSA's Information Systems Office to study the desirability and feasibility of storing data for an entire branch of space science. The PODS database consists of satellite observations of the Earth's oceans from a number of missions including SEASAT. The system resides on a VAX-11/780 computer at Jet Propulsion Laboratory and is available to qualified users via dial-up lines. As with PCDS, the majority of data reside on tape and PODS maintains an online catalog with information about each dataset that is managed using the RIM relational DBMS. After a user searches the catalog and locates the data desired, those data can be extracted from the database and transferred via phone line or magnetic tape.

12.3.3 NASA End-to-End Data System

The NASA End-to-End Data Systems (NEEDS) was designed to be a testbed for hardware and software that can acquire and archive massive amounts of data at rates approaching 50 Mbits/second. Data comes into the NEEDS system in packets, with a standard header on each packet. The system's Packet Management System software accepts packets as they arrive, strips off packet headers, stores the headers in a packet directory and writes the packet to disk. Eventually all data will be archived on an RCA optical disk juke-box storage system with a capacity of nearly 10 Terabits. The packet directory is managed using the ORACLE DBMS and the directory can be queried using the SQL language. The NEEDS has become an important data system for scientists in the fields and particles domain. The system is implemented at the Marshall Space Flight Center and comprises three VAX-11/780 computers and several minicomputers connected by optical fiber links. The NEEDS is accessible to a large number of facilities via dial-up and dedicated lines.

12.3.4 Solar Mesosphere Explorer Mission Database System

The Solar Mesosphere Explorer (SME) is an earth-orbiting ozone monitoring satellite operated for NASA by the University of Colorado's Laboratory for Atmospheric and Space Physics (LASP). LASP processes and analyzes all data from the satellite. Level 1 data, and some of the level 2 and 3 products, are managed by a DBMS built specifically for the mission. The system does not use a catalog as such, but users can query the database to determine which data are available. All data are archived on tape and users can promote? older data back to disk whenever access is required. The SME DBMS permits, even encourages, users to generate their own views of the database. The views not only specify which data are to be returned to the user and his programs but also determine how the data are formatted. Even though a particular data item may be stored on disk as a 2-byte integer, for example, a user can specify through his view that the item is to be returned to him in floating point format. Similar format conversions take place automatically when data are written into the database. The SME DBMS also provides special Processing Summary software that automatically tracks and documents the processing of all data.

12.4 Perspectives on Planetary Database Management

A planetary data system would bring together the following kinds of people and organizations:

a) SUPPLIERS - Organizations and individuals that analyze scientific information and who make these data available to others via the PDS, either directly or by transferring the data into an archive center.

b) USERS - Scientists, teachers and students searching for and accessing data via the planetary data system.

c) DISCIPLINE CENTERS - Organizations responsible for inserting data into the planetary database, maintaining those data and disseminating them to consumers upon request. This is based on the "center of excellence" concept: institutions with the scientists, technical staff, and computer resources necessary to obtain data from and provide it to a large segment of the space science community.

These are functional divisions only and any one individual or institution might perform more than one role. For example, scientists at a discipline center would probably also be suppliers and consumers. There is a need for an additional organization to perform functions similar to that of a database administrator:

d) CONTROL AUTHORITY - Centralization of certain aspects of the data management process will be needed. Chief responsibility of the control authority would be to ensure that data from different missions, disciplines and archive centers can be retrieved by consumers in as standard a way as possible.

12.4.1 Suppliers

A supplier acquires sensor data from his experiment, analyzes those data and stores them in his own database. Suppliers might make their data available to others via the PDS in two ways:

a) Directly, by allowing other users to access the data through the supplier's computers.

b) Through a discipline center. The supplier would contact the appropriate center for his discipline when he is prepared to submit data to the planetary database, then work with the center staff to determine how to modify the schema of the planetary database to accommodate the new data. When both parties are ready, the supplier ships his experiment data, along with ancillary data and documentation, to the center and the data are inserted into the planetary archive.

The first method offers speed and flexibility; it would be a fine way for co-investigators to share data during a mission. The second method would make it easier to provide standard access to the data and facilitate the development of a catalog of all datasets.

12.4.2 Users

In many respects, accessing a planetary archive will be similar to buying a car. Consider the activities of someone in the market for a new automobile:

a) The buyer searches through brochures and magazines to determine what makes and models are available. After a first look at the market the buyer begins to narrow his search by getting more information on the cars that possess the features he desires.

b) Once the buyer knows which models he likes best, he goes to a dealer and examines individual automobiles until he finds the one that has the right color, options, etc.

c) The buyer orders the car from the dealer and the dealer fulfills the order by delivering the car to him.

Now imagine a scientist accessing data via the PDS:

a) The first activity is to determine which data are accessible that might be pertinent to a particular scientific study. To do this the user needs the equivalent of the car buyer's brochures. Descriptions of each dataset might be provided through queries to a catalog similar to those available with the PCDS and PODS databases.

b) After our scientist identifies useful datasets, he may wish to examine a sample of those data to make sure they are really what he wants. This selective examination of he data is often called "browsing."

c) Finally the user orders some portion of the data and the order is fulfilled by delivering (electronically or otherwise) those data to the scientist. As with the car buyer, the organization that fulfills the order may not be the original supplier, but an agent acting on the supplier's behalf (i.e., a discipline center).

d) The scientist reviews and analyzes the data that was sent. If the data were acquired from a discipline center, he might help the center evaluate its performance by returning comments on data quality and the service she has received.

12.4.3 Discipline Centers

The process of transferring data from suppliers to discipline centers will necessarily be more formal than most scientist would like. Before accepting data from a supplier, the center staff will have to review the data being submitted, its ancillary data, processing history and documentation. If costs are to be kept low, the centers will have to be somewhat hard-nosed about not accepting data that doesn't conform to standards or that lacks documentation. To avoid problems center staff must work with future suppliers, keeping them aware of all requirements for submitted data. If the relationship between centers and suppliers is not a close and cooperative one, with mutual understanding of the problems faced by the other, then it is likely to become adversarial and ultimately untenable. Similarly, if centers are not responsive to their users, the users will find ways to circumvent them. This too will drive up costs as duplication of effort increases.

12.4.4 Control Authority

The control authority would set both scientific and technical standards; only the latter are discussed here. The control authority would initiate or ratify most data management policies within the PDS and set guidelines for discipline center operations and data system usage. There may be inherent resistance to this level of centralized control over PDS operations, so we reiterate an earlier point: it is almost universal that each database, no matter how distributed, is under control of a single database administrator or administrative organization. Large databases -- and a PDS will be one of the largest — require full time database administrators. It is therefore likely that the PDS control authority will need a technical staff of several individuals dedicated full time to the task.

12.5 Issues of Planetary Database Management

A number of the issues that arose in the previous section require additional discussion. In particular, the following will be of major importance when designing and developing the PDS:

a) STANDARDS - What standards are necessary to guide (and even bind) those who submit data to the planetary database and what standards are required so that users can easily access the database?

b) DATA CATALOG - The database catalog is a user's window into the planetary archive. What information should be in a planetary database catalog? How should it be organized? How will it be accessed?

c) PLANETARY DATABASE - What is the appropriate organization for a planetary database? What types of protection are required? What role can database machines play in the PDS?

12.5.1 Standards

Anyone who has attempted to process information that originated on a computer different in type from their own can attest to the obstacles created by the lack of standards for representing and transferring data. This lack of standards is not only bothersome but costly. For example, many years of work were required to create the software for handling all the datasets residing in the Pilot Climate database. The NSSDC has loaded over 150 datasets into an on-line database and their experience indicates that several weeks of effort are required before a data set can be loaded: this programming effort is the major cost, by far, in constructing the online database. Software development effort - and the attendant costs - can be considerably reduced by appropriate standards for data supplied to and residing in a planetary database.

It may be difficult to impose standards initially because of the wide variety of existing data that might be "grandfathered" into the system. as time goes on, however, standards ought to become increasingly prevalent. Standards need to be developed that cover the entire life cycle, from data acquisition to data archiving. Effective standards would probably exist at three levels:

a) Mission - Data from an experiment should meet the needs of the particular mission. Mission standards might be established by mission science steering groups in consultation with discipline centers that would receive data from the mission.

b) Discipline - Standards should be imposed on data within each discipline to make it easy for co-workers to share data.

c) System - System-wide standards that cut across disciplines should be established when possible. The definition and enforcement of system-wide standards would be facilitated by the control authority.

Areas needing standardization are outlined below.

a) Standard Identification

The simplest form of standard would be to attach a label to all data within the PDS that unambiguously identifies the data. The labels would include information like the target, mission, instrument, data format, originating computer, and perhaps a processing summary. Standard labels provide important information to the users, but they also permit data management and data processing software to automatically determine how the data are to be handled.

The space science community already has rich experience with standard labels: planetary images are commonly prefaced with a VICAR label, Landsat CCT tapes have a header record identifying each scene and the radio astronomy community has developed a standard set of labels as part of the FITS format for data interchange. An international effort is currently underway to develop a new labelling and data registration convention that would allow data to be packaged into "Standard Format Data Units" (see Chapter 10 for a discussion of the SFDU concept).

It is not required that each record or even each dataset in the planetary database have a label attached to it. The labels are principally useful when data are moved from one location to another or even from one process to the next and they can be constructed and appended to the data by the transport software. How data are transported (i.e., tape, local network, packet network, etc.) is not of particular concern from a data management viewpoint. But if each data transmission is accompanied by appropriate labels then the data become self-identifying: the receiver can determine from the labels alone what he is receiving and process it accordingly.

b) Standard Organization

The term "organization" here refers to the way in which data are arranged on mass storage devices. Some space data are organized in a straightforward manner: images, for example, are typically two-dimensional arrays with one byte per pixel. But often data organization is highly idiosyncratic, making it difficult to use the data without specialized software (and much patience). There are some standard data formats currently in use like VICAR and FITS, but some of the so-called standard formats have permitted new and incompatible versions to proliferate.

Future decisions about how to organize experimental data must be predicated upon the eventual needs of the users who will access those data via the PDS. Additionally, the database systems used by the PDS will impose some requirements and restrictions on data organization. Suppliers who expect to eventually place their data in a discipline center will have to arrange their data to be compatible with the center's database. The PDS, through its control authority and discipline centers, must provide suppliers with information on how data should be structured. Although this may appear to be a burden upon the suppliers, it will work to everyone's benefit by reducing the need for custom software and diminishing the potential for redundancy, error and inconsistency.

c) Standard Descriptions

When data are submitted to the PDS they will have to be accompanied by a significant amount of descriptive material. Standards governing the content and form of this documentation should be established to ensure completeness. This is particularly important since the documentation becomes the raw material from which data catalog entries would be created. Required documentation should include:

- o Descriptions of science instruments and their characteristics

- o Descriptions of each dataset, including not only what data are available but why they were acquired

- o Descriptions of available analysis software

- o Data formats

- o History of processing applied to the data

- o Descriptions of ancillary data and their relationship to the experimental data

- o Bibliographic and reference material

Data formats pose an interesting problem. As we noted earlier, each DBMS has a Data Definition Language for defining all data items, records, etc. But if the PDS is a heterogenous database system or, as is likely, the chances are high that different database software and hardware will be used in the future, then data formats should be defined in a way that is independent of any DBMS.

One area listed above, standardized histories of the processing applied to data, is only now received appropriate attention. Discipline-wide, and if possible system-wide, formats for processing history information should be defined in a way that is independent of any DBMS.

One area listed above, standardized histories of the processing applied to data, is only now receiving appropriate attention. Discipline-wide, and if possible system-wide, formats for processing history information should be defined, and standard software developed for creating, maintaining and utilizing processing histories.

d) Standard Administration

Any large distributed data system needs standards governing activities at all nodes. In the PDS these standards would principally apply to discipline centers, but they also affect suppliers and users. Areas requiring standardization include data modelling strategies to be followed when developing the planetary database, naming conventions, standard terminology and configuration control guidelines. The control authority would develop and maintain these standards and enforce them as necessary.

12.5.2 The Data Catalog

One of the most successful aspects of the Pilot Climate and Pilot Ocean systems has been their online catalogs describing the datasets available. There is strong support for a PDS catalog that would contain information about missions, experiments, datasets, ancillary data, data processing, data formats, and more. This subsection explores some of the data management aspects of such a catalog.

a) Implementing the Catalog

The experience of Pilot Climate, Pilot Ocean and NEEDS provides a convincing demonstration of the suitability of commercially available relation DBMS' for maintaining catalogs of a data system's holdings. Almost certainly a centralized catalog can be built for the PDS and cared for using a relational DBMS or a database machine; however alternatives to a centralized catalog should be considered. Catalog contents could be totally distributed; that is each discipline center and supplier can maintain and provide access to its own catalog. A distributed system keeps the catalogs close to their source, thereby reducing the delay before the availability of new data is reflected in the catalog. However, total distribution poses significant problems. How do users know where to look for particular data? How many different nodes might have to be contacted to fulfill a single complex query? How does the control authority guarantee commonality and standardization in such an environment?

One way to avoid some of the above problems is to have a hierarchy of catalogs. There would be one central catalog with some information about each dataset. A dataset's description in the central catalog would indicate where the dataset resides and how the user goes about accessing it. A user would start out by querying the central catalog but he might then be directed to another catalog at a specific node for more detailed information — down to the level of an individual image. If major PDS nodes are connected by a network any switching between the central catalog and node catalogs may be transparent to the user; otherwise, he might have to dial the number of the node catalog computer himself. The hierarchical scheme has some of the best characteristics of centralized and distributed systems: users always know where to start looking for data and only relatively small amounts of information have to be transferred from suppliers and discipline centers into the central catalog. It is not without problems, however. A significant effort on the part of the control authority would still be required to produce commonality between all the catalogs.

Another possibility, compatible in greater or lesser degree with all three types of catalogs outlined above, is to copy some or all of the catalog contents to magnetic tape or optical disk and disseminate them to the scientific and educational communities. The user can then perform queries directly at their site. This edition of the catalog might be "illustrated", containing samples of the data. Special software and possibly hardware would be required to read the catalog.

A detailed analysis of costs and benefits — well beyond the scope of what this document can provide -- is required before committing to a particular approach, but the Pilot Planetary Data System is examining the issue of catalog implementation and will hopefully determine how the PDS catalog should be organized.

b) Querying the Catalog

To be of greatest utility, the catalog system should be accessible to anyone who has a proper terminal, modem, account and password. Software for processing queries should reside with the catalog and not within a user's computer.

To access the catalog a traditional query language like SQL, an interactive language like QBE or a language designed specifically for the needs of the planetary science community might be used. Planetary scientists should decide what they want from a query language and the overall format of the query language. Implementing a new query language is not overly difficult and it is preferable to suffering with a query language that does not quite do the job.

What would be the catalog system's reply to a typical query? Answers to a query would often consist of a table of information -- for example, the names of datasets containing the desired data and the times for which data are available -- printed at the user's terminal. Users with proper software and hardware might have this table transferred to them as a file, permitting further manipulation on the user's computer. In a hierarchical catalog, a reply might consist only of pointers into another catalog and instructions to the user about what to do next.

An an example of catalog usage, imagine a researcher looking for information that might pertain to a study of volcanoes. The first query he makes can be paraphrased as "what information does the planetary database contain on volcanoes?" Hopefully the answer comes back: "besides the earth and its moon, there are volcanoes on Mars and Io and very likely Venus as well, and the database contains many datasets that might be of use to you." The answer should include the names and descriptions of each dataset that might contain information on volcanoes and sufficient information to allow the researcher to continue the search. The answer to our volcano query might say, in part: "instruments onboard the Mariner Mars and Viking spacecraft obtained data on several Martian volcanoes. If you are interested in pictures of volcanoes you can search the imaging catalogs from those missions. If you are interested in a particular target — the volcano Olympus Mons — then search for images centered with five degrees of 134 degrees west Martian longitude and 18 degrees north latitude." Of course these answers would be presented in a terse tabular form and not in English sentences. The system should contain the necessary aids to help the user frame his query properly and even to permit the system to determine the meeting of fuzzy queries. A taxonomy of space science included as part of the catalog and available to the user can guide him to the proper categories to query. An on-line thesaurus could help the user find acceptable synonyms for terms that he uses but that the catalog system does not understand. The catalog system might even search the thesaurus for the user and then verify with him whether it found the proper acceptable term before processing a query. These aids are not typically part of a DBMS but they can probably be added to the catalog system with only modest effort.

Obviously the catalog system could be overtaxed by large queries posed wittingly or unwittingly, therefore, the catalog system should restrict the size of the search performed for any query. Some relational DBMSs estimate the number of tuples that will have to be searched before responding to a query. Provided with this information, the catalog system could estimate the time required to fulfill a request and compare it to a quota associated with the user to determine whether or not the search will be made. In some instances the system might refuse to perform the search; in other instances processing may simply be delayed until non-prime time. The catalog system should also explain clearly to the user why it is delaying or denying any query.

c) Browsing Through The Catalog

Catalog browsing might be implemented in several ways:

1. Digital media may be supplied to consumers containing the data that can be browsed. In many respects this would be an extension of the illustrated catalog discussed above. Special hardware and software would be necessary to support this method. Digital video disks might be a good medium for supporting this type of browsing.

2. Data can be supplied to the consumer in an analog form like video tape. This method is already being used successfully for imaging data. It too requires special hardware and software.

3. Data at a discipline center could be browsed via a telecommunications link. Bandwidth limitations would impose serious constraints upon the type and volume of data that could be browsed in this way.

As an example of some of the concepts discussed above consider the following query-and-browse system available to users of earth imaging data:

The INORAC System

INORAC (INquiry, ORder and ACcounting) is a database management system for processing the inquiries and orders of earth image data from the USGS EROS Data Center in Sioux Falls, South Dakota. Data cataloged in the system includes imagery from NASA U-2 flights, Genimi and Apollo missions, Skylab and Landsat. Data types include synthetic aperture radar, passive microwave, thermal, panchromatic and color infrared photography and multi-spectral scanner data (both digital and photographic).

INORAC provides access to the EROS database from most types of terminals and modems. After logging into INORAC the user initiates a program (RESORD) which provides the database inquiry capabilities. The user specifies the longitude, Landsat path and row, etc. Identifying information about each qualifying image is put into a temporary table. the user can continue to specify other restrictions -- such as data source, maximum cloud cover, recording technique, satellite, instrument, etc. -- and the number of entries in the table is reduced to include only those images that meet all qualifications. The remaining entries in the table can be printed on the user's terminal or on a printer at the EROS center and mailed to the user.

The user can browse the images located for him. Images are available on microfilm cassettes accessible by image number. They can be examined at National Cartographic Information Center facilities. There is at least one such facility in each state and most are located near major population centers. The desired images can then be ordered through the EROS data center.

12.5.3 The Planetary Database

The system that has been discussed so far is a complex one, composed of large database nodes (discipline centers), smaller database nodes (suppliers providing access to their own data) and a diverse group of users. It is difficult to find an existing database system which is similar. The PCDS, PODS, NEEDS and SME database systems have been discussed but they are all on a significantly smaller scale. Other large distributed systems are in the works but in many respects the PDS will have to break new ground, not only for NASA but for data systems in general. The following discussion of planetary database implementation follows the outline of Section 12.2 covering organization, access and protection.

a) Planetary Database Organization

We noted previously that commercial relational DBMS' do a fine job of managing science data catalogs. Unfortunately the situation is much less clear about how to manage the database itself. Remember that PCDS and PODS use DBMSs for their catalogs only; the databases are accessed through special software developed by the pilot projects. This does not mean that DBMSs cannot manage a science database: the SME database system indicates that DBMSs do the job very well, but there are definite obstacles. Some have already been discussed, particularly the need for standardization of the data to be installed in the database. Others obstacles have to do with inherent limitations of current DBMS'. Commercially available DBMS' typically do not support data types (e.e., single and double precision floating point) and data formats (i.e., vectors and arrays) that are required for a science database. Much careful analysis will need to be done to determine the proper blend of DBMS (and perhaps database machines) and non-DBMS software for managing the planetary database. DBMs can be quite cost-effective but they are not the entire solution to managing a planetary database. Some non-DBMS software will always be required (to process catalog entries, for example). If commercial database systems are used then even more non-DBMS software may be needed to handle the types of data that do not fit within the DBMS framework (i.e., images). That is why standard data organization is important: if there are limits on the number of data formats used in the PDS, then less non-DBMS software will have to be developed.

What would the PDS database look like? The options seem to be:

o Don't place any restrictions on how a node implements its portion of the planetary database. This means a proliferation of formats for data and the development of a great deal of custom software. Since software is the principal cost and schedule driver in a system like the PDS, this option may be too costly, too risky, too difficult to create, maintain and use. In many respects this in no improvement over the current state of affairs.

o At the other extreme, create a homogeneous database by having every node on the PDS install a specific DBMS and supporting software. This would promote uniformity and make it easier to tie the nods together. But it will probably not be possible to find one DBMS that can handle all types of planetary data or that is available for all the types of computers likely to be found in the PDS. There is no reason to expect that scientists will agree on a database management package any more than they agree about which computer, operating system and programming language to use.

o Accept a "controlled" heterogeneous environment where discipline centers and suppliers can choose from a few different sanctioned database management packages and a common user interface to all systems is provided. This means, for example, that special software would be developed to translate requests for data from a standard system-wide format into the format required by a particular node's database management package. This approach presents many problems but it is quite feasible. Although this option might be somewhat costly to develop, it could significantly reduce overall lifecycle costs.

Different parts of the PDS will require different mixes of DBMS and non-DBMS software. A discipline center might be able to afford the best DBMSs, even database machines, but many suppliers and most users could not. The user interface would perhaps be the key element in such a system. Not only would it provide uniform access to the PDS but it could provide a highly integrated environment, tying together the data management and data analysis software. Therefore much thought must be given to the overall design of the user interface. The interface should incorporate some of the techniques commonly found in small computer interfaces like menus, tokens and windows.

b) Accessing the Planetary Database

Data might be loaded into the planetary database on an orbit-by-orbit or day-by-day basis or, if data are transferred to a discipline center, the entire dataset may be entered at one time. None of this would present any problems for available database systems. There are important issues, however, about how to read data from the planetary database. Should users be able to get into a discipline center's computers, browse through the data and extract data themselves? This is more of a policy decision than a technical one. It requires that sufficient computer resources be made available to the discipline center to support.

One easy way to provide user access to the database would be to have an online system for ordering planetary data. The ordering system would be much like a central catalog in that the user could call one number to order any data, regardless of where the data actually resided. The order would then be forwarded to the proper node and the order filled. The user would only need a terminal and modem to place an order. Proper safeguards should be implemented to prevent losing orders if the system failed. After an order is placed, a user could call into the order system and determine the current status of his request.

A mechanism should exist for informing users about additions to the database. Each discipline center and supplier might maintain a list of users interested in particular topics and issue a notice to those users (electronically or otherwise) when new data on those topics are available. The same mechanism would make it easier to "recall" data — notifying users that data they have may be suspect or invalid.

c) Protection

Fortunately a planetary database will not require the safeguards that are necessary for banks and other commercial enterprises. The level of protection that is provided by most DBMS' should be adequate. Standard database security mechanisms can ensure that users only access those parts of the database for which they have permission. Scientists typically check and screen their data carefully during processing and so DBMS integrity checks performed at the time data are entered into the database are useful but not of paramount importance. Synchronization of database access is not a major issue since users will not write or modify the planetary database directly. A small fraction of the planetary database will be changing at any one time, since only standard techniques for recovering data will be sufficient.

13. SOFT-I - IMAGING PROCESSING SOFTWARE

13.1 Introduction

The planetary research community has now matured to a point that it requires direct access to the enormous volume of planetary digital imaging data. Hard copy generated in the course of the missions or by data consortium activities is now insufficient. It is one thing to archive the digital database for the community; it is another thing to retain/establish the means to use it. The imaging data exist in an enormous range of states in terms of processing maturity. Many investigators now wish access to the original raw instrument data. The capability for handling the raw instrument data from past missions even in terms of having the necessary software to read the tapes is rapidly being lost. Even the lead facilities can no longer read and calibrate raw Mariner 9 images. Necessary software is obsolete or has been lost due to the rapidly changing computer processing environment (multiple changes of hardware and software), lack of concern for software portability in the original design, and no delegated responsibility or funding to maintain the software. Viking and Voyager image data may reach a similar circumstance in the next few years.

Processed data are in better shape in terms of portability and ease of use by a range of investigators, but such data are scarce. The Mars and Lunar Consortia generated reduced image files with common format and resolution thus sacrificing much of the information from any particular investigation. Calibrated and geometrically transformed Viking Orbiter or Voyager images are rare and generally were not funded.

The problem the Planetary Data System (PDS) must address is to provide the necessary software, calibration data, and geometric knowledge to the scientific community so that the imaging data base, ranging from raw to the most highly processed data can be accessed and analyzed by any user, with any experience, on any computer. The specific concerns in this task are as follow.

a) To what level of maturity should the imaging data be processed before it is distributed?
b) Should the raw imaging data be processed and distributed or should it and the software to process it be distributed, at what level?
c) How should radiometric calibration data be standardized; who should be the curators; can it be standardized?
d) How can the information for image geometry be standardized so that pre-mission, and post-mission data have a systematic relationship to one another and can be mixed or compared to other missions?
e) What are the general tools that users need for image manipulation? Are some items (high-order geometric transformations) more valuable for general distribution than others (e.g. filters, stretches)?
f) What should be the standards to enable portability of functional software modules?
g) What are the opportunities/advantages for common interactive executives?

13.2 Levels of Image Processing Software

In order to structure the discussion of image processing software requirements we have grouped the software into a series of levels. In general, the higher the level of processing, the less specific (styled for a particular experiment) the processing becomes so that at the highest levels the software becomes conventional. At the lowest levels the software operates on raw data. Development and maintenance of software for these lowest levels requires great familiarity with the spacecraft, the detailed operating characteristics of the instrument, the geometry of the observation, and the conditions (radiation background) under which the data were collected.

13.2.1 Level 1: Logging and Formatting

The planetary spacecraft framing camera images are traditionally archived by each mission on a series of computer compatible magnetic tapes referred to as EDRs or Experiment Data Records. These represent the record of the spacecraft images in their most primitive form. The EDR files are not yet in image raster format. Each line is treated as a separate, minor frame each with a series of ancillary information related to the spacecraft, instrument and data link conditions. The latter provide information on the quality of the data and the error rates encountered during transmission. Minor frames not received are absent from the EDR. An image can therefore be partial and can occur in several segments in the EDR files.

Programs developed to read and format the EDRs into image rasters are referred to as logging programs. Such programs are complex; they require a major amount of decoding and error checking and contain options for bit error restoration. Additionally the EDR formats are mission specific; separate logging programs are required for each mission. One option would be to distribute the archival EDR files and along with this data to also distribute the software needed to format them (in a portable form). This option is attractive in that the user has the option of inspecting and treating the raw image, and its original ancillary information, so that software can be customized to the particular scientific research application.

Variations on this option include logging the EDRs into a raster format with the ancillary data attached to the end of each image line. Lines of missing data blank and bit errors could be left uncorrected or these could be corrected and the changes marked. In these options the raw data is still available but the images can be placed in a single raster format to be read by a single program.

13.2.2 Level 2: Radiometric Correction

The raw raster camera images in image data numbers (DNs) are converted to conventional radiometric units by this operation. In the most simple form this involves modeling the camera response and subtracting the dark current. Non-linear response functions are removed and the image is scaled to absolute radiometric units.

Planetary cameras like those used on the Viking Orbiter and Voyager spacecraft are complex devices having many operation modes. These include selectable read-out rates, gain states, light flood options, offset options to prevent black clipping, optical filters, and simultaneous exposure options for wide and narrow angle cameras. All of these modes affect the radiometric signature of the cameras. The radiometric performance for a given set of operating conditions can change as functions of the operating temperature of the cameras, aging of camera components, or radiation conditions that cause the dark current to vary. Hence the radiometric calibration files must be functions of a wide range of operating conditions, spectral band passes (filters), and time for each camera. Finally, most cameras have a variety of artifacts which optionally can be removed during radiometric correction. These include coherent noise patterns (such as microphonics on Mariner 9 and Viking Orbiter II), residual images (most prevalent on Mariner 9), and the random noise and bit errors common to all planetary imaging data.

One of the greatest problems with the radiometric calibration files is their volume. The many operating modes, filters, and time-variable properties have a factorial multiplying effect implying hundreds of files for each camera. When one considers that the traditional method of handling non-linearity is to break each light transfer curve up into linear segments rather than storing them in a functional representation (e.e. linear, second-order, exponential, etc.), as many as 12 segments or "planes" can be used which add seriously to the volume.

Operational noise levels in the cameras had been reduced to such a low point by the era of the Viking Orbiter (cameras), that radiometric correction had to be performed at higher precision than the eight bits of the original data encoding. Corrections at a lower precision resulted in low frequency contouring of the image. As a result the low frequency content of Viking images is known to a much higher precision than eight bits. The additional precision, however, has increased the complexity of the calibration files. An additional consideration in the complexity of calibration files is the pixel density of the file. Correction values can be provided for each pixel or for a block or group of pixels. The optimum size of these groups is yet to be determined. Recent work with Voyager dark current has shown that granular patterns at the scale of a pixel is recurrent; if dark current valves are stored pixel-for-pixel, the noise can be removed. For each type of calibration file cited the greater the number of valves stored in the file, the higher the complexity of the calibration procedure.

The extent to which the imaging science community wishes to recalibrate raw imaging data, will determine the products which must be made available for their use. If raw data* is desired then the radiometric calibration data, application software, and documentation procedure must be provided. Alternatively if the requirement is for more mature data, calibrated image files (level * or above) could be distributed.

The disadvantage of the second option is that radiometric calibration files and techniques are in a constant state of refinement. Even today the calibration for Viking Orbiter 2 is in the process being refined. If the raw EDRs are distributed radiometric processing and calibration data modules can then be updated.

Certain tradeoffs must be made if calibration files are distributed. For instance it can be argued that the natural errors and variance in the calibration is larger than the difference in storing a simple functional fit and retaining the full segmented calibration files. Additionally, most of the time-variability is now known to be in the dark current rather than in the shading or responsivity. The dark current appears to be a strong function of the read-out rates; the shading files seem invariant to many such mode changes. The dark current could be retained at higher resolution and precision compared to the shading files and additionally the dark current is independent of the filter bandpass. All of these considerations suggest that a practical compressed form for the radiometric calibration files can be developed and distributed with the raw data.

Handling the additional mission-unique complications such as residual image, coherent noise removal, and bit error restoration is another consideration. These perhaps should be considered as optional additional modules to be added to the distributable radiometric software/calibration data. Finally, it would be desirable if a single standard could be developed for the radiometric processing algorithms and calibration data formats so that software maintenance could be simplified. Ideally an individual institution would be charged with the responsibility for developing, maintaining and distributing the radiometric software and data.

13.2.3 Level 3: Correction for Geometric Camera Distortions

The planetary framing cameras that have flown on planetary spacecraft have vidicon sensors and have inherent internal geometric distortions. These arise from irregularities in the pattern of the electron beam which scans the image stored on the photoconductor in the vidicon. Two common types of distortions are 1) "barrel" distortions producing severe distortions in the frame corners, nominally fixed from frame to frame and 2) beam-bending distortions caused by deflection of the beam by the charge distribution of the image itself, formed on the photoconductor. The second of these varies from image to image. Other distortions can be introduced by variations in the ambient magnetic field. The solution is the use of control points called reseaux that are burned into the photoconductor surface. These produce black holes in the image whose geometric positions on the photoconductor are known with great precision. Software is used to automatically locate the reseaux in the image. From these, data correction matrices are derived which provide the map between the digital image and undistorted "object space"

Alteration of the geometry of the image at this stage is optional. It may be sufficient for a user to know the corrected geometric position of each pixel in a camera coordinate system or reference frame. Other users may require a geometrically transformed image for registration of successive images. The geometric transformation is usually performed by mapping the image into a new roster utilizing the matrix of values relating the distorted geometry to object space geometry. Following this operation the camera-introduced corrections are, in theory, completely removed; that is each pixel is reduced to a standard radiometric energy unit in a known geometry, centered in the camera frame-of-reference. "Perfect" camera pointing relative to the target body and the sun is addressed in Chapter 14, "Geometry Software Common to All Experiments."

13.2.4 Level 4: Geometric/Navigational Corrections

Distinction is made here between the process of deriving information about the acquisition geometry from processes which utilize that geometric information to alter the image. The latter processes include geometric transformations and the removal of model photometric functions. The geometric data base includes: 1) the relative positions of and orientations of the target body (a planet, a satellite, planetary rings, etc.), of the spacecraft, the sun and the earth and 2) the camera pointing information in terms of target body coordinates. From these data the user can derive for every pixel the position on the target body in latitude and longitude, the emission angle and azimuth to the spacecraft, the solar incidence angle and solar azimuth, the phase angle and the distance to the spacecraft from the target. If the user is interested in doing photometry he is not necessarily interested in altering the image beyond radiometric calibration. The radiometric brightness of a pixel and the geometric conditions may be his sole requirements. Another user may be interested in making measurements of planetary shape, topography, feature dimensions with the raw image data. Again he may need only the geometric information for the image and measurements taken from the new data to complete his task. These needs can be addressed by Common Geometry Software, discussed in Chapter 15.

13.2.5 Level 5: General Software Tools

There are several advantages to producing a general software library of some commonly used image-related functions. These include:

a) avoiding unnecessary duplication of effort in software development.

b) providing more image manipulation flexibility than is currently available at many sites.

c) allowing compatibility for performing high-level operations simultaneously on several data sets from different missions and different disciplines.

d) rapidly providing research-level study tools for new data sets.

We recommend that the software included be a suppliment to user developed routines and commercial packages such as input-output routines. The following is a list of specific image handling functions which might be included in such a library. Centers which currently have software that could be used as prototypes for these functions are given in parentheses.

a) Cartographic functions, projections and map drivers (USGS-Flagstaff)

b) 2-D and 3-D geometric transformations, stereo manipulation (IPL)

c) Image registration

d) Photometric function modeling and removal

e) Catalog functions - sorting by picture label parameters (Wash. U.)

f) High-level plotting functions - section plots, mosaicing, picture differencing.

14. NON-IMAGING SOFTWARE/DATA ANALYSIS REQUIREMENTS

14.1 Introduction

The goal of the Non-imaging Software Splinter Group of the Planetary Data Workshop is to identify the analysis software needs of the non-imaging planetary data user and to begin to establish a framework for analysis software within the Planetary Data System (PDS). There are several objectives supporting this goal:

- Establish working assumptions as to the nature of the Planetary Science Data Center (PSDC) or Centers where the data physically reside, the nature of the user workstation, the existence of a computer network linking users and PSDC(s), and the quality and nature of the planetary data itself.

- Identify data or experiment types within the purview of non-imaging data, so that clear analysis needs may be assessed.

- Identify facilities that users are likely to need to define and access data.

- Define data manipulation and analysis needs: What are facilities common to all non-imaging data users.

- Establish display software attributes.

The eventual design of a non-imaging analysis software system must address the functional requirements derived from the above considerations.

14.2 Working Assumptions

The development of non-imaging data analysis software for planetary data is predicated on the existence of online datasets residing at either a central planetary data center or dispersed discipline centers and linked to scientific user workstations via a national network. The online data is presumed to be high quality, verified data approved by the Principal Investigator for distribution; this data would have been gathered as the first phase of the Planetary Data System (PDS) effort. It shall be assumed for the moment that the data is in the "final" calibrated form of physical units (level 7 or 8); the issue of raw or semiprocessed data will be addressed later. The scientific user workstations are assumed to consist of a graphics-supporting terminal, hard copy capability and some level of processor and storage capability. The network interface command language is to be simple and user-friendly, and network line rates must support at least 2400 baud dial up for typical scientific needs.

All levels of the system will be "Help" supported so that a user may learn what options are available and receive some instruction on how to use them. A menu-driven system accomplishes this easily, with each menu item providing access to further sub-options and further documentation through "Help" (e.g., a VAX VMS-like technique). Experienced users, however, will wish to shortcut intervening menus and proceed immediately to the task level desired and such a shortcut provision must be included in the system (as in TAE). All operating system and utility-level facilities should be help-supported; some analysis and display-level facilities may not be help-supported but should at least allow the user to get back to a level or facility that he/she understands.

Finally, a User Steering Group must actively oversee the development of all software to assure user-friendliness and usefulness. This group would presumably be at least partly comprised of PI and non-PI space scientists, people whose input is crucial from the user-impact point of view.

14.3 Non-imaging Database

The data types and experiments to be accessed by the non-imaging data analysis software fall into two broad categories: (1) In situ data and (2) Remote sensing data. The former includes particles and fields, direct atmospheric measurements by probes and spacecraft, and planetary surface sampling, geology and meteorology, while the latter includes radio, radar, microwave, infrared, visual, ultraviolet and x-ray and gamma ray measurements (spectrometry, radiometry, photometry, polarimetry, interferometry) from either Earth-based or spacecraft-borne instrumentation (see Table 1) or from the laboratory. The diversity of these measurement types also reflects diversity of analysis strategies; it is this diversity that will govern what software is common to all non-imaging planetary data and what software is discipline- or experiment type-specific.

14.4 Data Definition and Access

In order for a planetary science data user to know what data is available for study, to search and sort that data for desired parameters, and to review and access that data, several Data Base Management System (DBMS) interfacing facilities must be available. One is a catalog facility (discussed in the chapters on User Requirements and on Database management) which could provide the user with an overview of the planetary datasets by mission, by planet and by measurement type. A "browse" facility would allow the user to review key parameter datasets containing limited (low resolution) information for rapid display and review. A "status" facility would remind users of their current dataset directory, search/sort configuration etc. Finally, if the system can support it, some set of display and analysis routines for user-selected data can facilitate a reasonable scientific return to the user. Another facility would allow the user to search or sort the datasets for specific parameters. These facilities are common to both imaging and non-imaging disciplines.

14.4.1 Browse/Quicklook

The "browse" or "Quicklook" facility provides the user with an overview of selected datasets. At a minimum this provides a text summary of user-selected parameters in a user-selected dataset (or several datasets) for selected times, targets, or other parameter. More useful is a display capability, plotting various user-selected data together for given parameters. One example of this would be to plot specific Pioneer Venus orbiter data on a common time base, say ionospheric electron number density and temperature, magnetic field strength, and plasma wave intensities as time series for 30 minutes about periapsis. Such a display (a version of which exists at UCLA) can be browsed on an orbit-to-orbit basis, allowing the user to search for key events. Rather than accessing the detailed high resolution data in the primary database, the browse facility handles data from special "Quicklook" parameter files made up of summary (low resolution) data from selected instrument channels. This arrangement facilitates the rapid display and I/O required by the browse philosophy without bogging down the machine. The "Quicklook" parameter files are to be installed when the primary database is installed.

14.4.2 Search/Sort

The search/sort facility is another DBMS-related function. This facility could be keyed to select user-specified parameters for specified times and/or locations for various selected instrument datasets. The result of this operation would presumably be an online dataset available for some limited display and analysis, or for transmission to the user by tape or network for analysis on the user's home system. Searching and sorting of data in the "In situ" category is usually keyed off time or some spatial parameter such as spacecraft altitude or latitude and longitude. Target or pointing location might be the most frequently used search parameter for remote sensing data. This facility must have a multiple dataset search capability.

14.4.3 Data View

The actual convoluted structure of datasets must be transparent to the planetary science data user. The translation from actual data storage format to organized usable physical parameters is handled by the DBMS. This entity not only tells system software how data within the dataset are stored, but it also defines the "appearance" of retrieved data to the user (integer, floating point). The user can (via system prompting) define a "data view" or "data map" for selecting desired quantities from one or more datasets, and can assign names to the different kinds of quantities. This "data view" capability actually represents a module that can both write and read the user-specified data subset. It must be saveable so that a user does not have to recreate it. (An example of "Data view" is given in Section 4.2).

14.5 Data Manipulation/Analysis

The level of display and analysis software available to the planetary science data user is determined in part by the load this places on the PDS computers. The amount of display and analysis being done by twenty or thirty users, each manipulating several tens of kilobytes of data, could radically slow down the discipline center system. To reduce this load, it is highly desirable to support some analysis software at the workstation. One might begin by incorporating the less compute-intensive analysis needs in the discipline center system first, and include more and more complex software as the PDS and workstation computing power is upgraded. There are several conceivable levels of data manipulation and analysis software; each level is more diverse and probably more CPU intensive than the last. The lowest level is the software common to all calibrated non-imaging data, no matter what the experiment or data type. Next is the software common to all data within either the In-Situ or Remote Sensing subdivisions. A third level may be found within each of these subdivisions; for example, under the In-Situ subdivision there may be generic analysis software packages specific to (1) Fields and particles, (2) Atmospheric measurements and (3) Landed or surface measurements. A possible fourth level (not far removed from the third) is instrument type-specific analysis software.

14.5.1 Calibration software

Calibration software falls in the instrument-specific manipulation category. This PI-contributed software would be instrument-specific code that converts raw/semi-processed data (level 6 or 7) into calibrated physical parameters (level 8) in the same way that Level 1, 2 and 3 Image processing software calibrate imagery. One virtue to this data-producing technique is that calibrations may be updated -- the data is always the best possible. Another virtue is that the entire dataset does not have to be reprocessed as newer and better instrument calibration becomes available, only the subset of interest to the user must be upgraded. One liability is that such a system has a CPU overhead; computing cycles must be devoted to calibrating data that might otherwise be devoted to analysis of previously-calibrated data. Another liability is that the calibration code, presumably developed by the instrument P.I., must be converted to operate in the PDS environment.

Since it is likely that the calibration software would require a long development time, it may be desirable to first establish a preliminary version of the calibrated dataset (with the understanding that calibrations will be updated), allow users to access and analyze this preliminary dataset while the calibration software and raw database are brought up to speed. Eventually the preliminary dataset is superseded by the calibration processing module. To reduce the associated central CPU burden, the calibration processing module may need to be transportable to some level of user workstation. As discussed in the User Requirements chapter, complete documentation of calibration codes is essential.

14.5.2 Common Manipulation/Analysis Software

The user may gain access to the calibrated data by defining "filters" using a program module which select data according to desired criteria (for example, certain longitude and latitude intervals). In addition, the user may need to cull out "bad" or "noisy" data; such a capability should also be in the filter facility.

The user can select any data within "data view" meeting defined filter criteria from the specified datasets and produce output data for study. The following scenario illustrates this system (as a conceptual model only):

a) Jane Doe logs onto the system.

b) Using "status" she obtains a list of data files which she created yesterday.

c) In today's work she wants to compare Viking Lander surface pressure data with Viking Orbiter water vapor measurements; first she sets up some filters. The system software writes the filter program modules; all that is required from Jane is to specify the filter criteria.

d) Filter definition:

```
TERM1=('VIKING ORBITER'.AND.'MAWD COLUMN ABUND'_
     .AND.TIME(>1976:200:0,<1976:365:0)_
     .AND.LAT(=48,/DELTA=10).AND.LON(=226,/DELTA=10))_
TERM2=('VIKING LANDER TWO'.AND.'PRESSURE')
TERM3=COINCIDENCE(TIME,TERM1,TERM2,/DELTA=0:0:3600)
SEARCH(TERM3)
```

e) The system performs a search and creates a temporary storage file of indices of events which satisfy the search criteria of TERM3. It reports back that 78 data events are retrievable.

f) Jane decides to retrieve those events and write the results into a data file called H2OPRES.DAT.

```
SELECT /OUTPUT=H2OPRES.DAT
```

g) Since no data view yet exists, the system software responds with a list of questions about what data Jane wants to include in her output records. Some of the questions and responses are:

```
Name of data view?  H2OPRES.VU

Select parameters for
VIKING ORBITER
MAWD COLUMN ABUND

DAY        INTEGER(2)
REM=JULIAN DAY-OF-YEAR OF DATUM
Include?        Y
Name?           VODAY

LVP7       FLOATING(4)
REM+7 VOLT POWER SUPPLY MONITOR (VOLTS)
Include?        N

H2O        FLOATING(4)
REM=WATER VAPOR COLUMN ABUNDANCE (PRECIPITABLE MICRONS)
Include?        Y
Name?
    .
    .
    .
  etc.
```

h) The selected variables are printed out.

Variables selected for H2OPRES.VU:

```
VODAY      INTEGER(2)
REM=JULIAN DAY-OF-YEAR OF DATUM
H2O        FLOATING(4)
REM=WATER VAPOR COLUMN ABUNDANCE (PRECIPITABLE MICRONS)
  .
  .
  .
 etc.
```

i) After the program runs, Jane has a file called H2OPRES.DAT, which contains 78 correlated measurements of surface pressure and water column abundance. After plotting both quantities as a function of time, using the system plotting software, she decides that she wants to have the data sent to her on a tape.

```
GENTAPE /VAX /ADDRESS=DOE /FILE=H2OPRES.DAT,H2OPRES.VU /DELETE.
```

The specified files will be written on tape in VAX compatible format, and sent to the address listed in the file DOE.ADD. The operator request to mount the tape includes a mailing label and printout of the tape contents, so all the operator has to do is mount the tape and put everything in a box afterwards. Since Jane doesn't want her grant to be charged for storage space for the file once it is copied, she specifies /DELETE, so the file will be automatically deleted after the GENTAPE operation is complete.

Jane would like to use the data immediately, but she has to wait for the tape to arrive. She is looking forward to the installation of the new data line in her lab, because she can then request that her data files be electronically transmitted to her lab minicomputer.

Other common (cross-discipline) analysis software will be limited to such facilities as:

- Simple statistics, including averaging, auto and cross-correlation, and simple regressions on data selected with "Data view".
- Simple transformations (may be part of Geometry software - Chapter 14).
- Fast Fourier Transform
- Subtraction of or normalization by a Standard Model.

14.5.3 Discipline-specific software

The second level of manipulation/analysis software is related to the principal differences in the nature of (gradiometrically corrected) Remote Sensing and In-Situ data. For example, it attacks the problem of filtering and stretching

Remote Sensing data, or providing transformations such as radiance versus wavelength to radiance versus wave number. It also provides resolution and format adjustment, as well as passband integration (spectral resolution modification) so that data from two different instruments and/or times can be inter-compared properly. In-situ data is rarely manipulated in this fashion (one exception being when in situ data is used in mapping). It is also important to recognize that non-imaging data is often used in conjunction with imaged data. One example of this would be using imaging to improve instrument pointing information. Thus there is overlap in user requirements between the two (imaging and non-imaging) software regimes, and care should be taken in PDS development not to completely strand one from the other.

14.5.4 Instrument Type-specific Software

Instrument type-specific software facilitates analysis of data from a given experiment class, for example, infrared radiometers or magnetometers. An example of this level of software might be to convert IR radiance measurements to a brightness temperature. Another example would be the integration of moments of a particle distribution function measured by a plasma instrument to yield total plasma densities, temperatures, bulk flow and heat flux. Analysis at this level might have to be supported at the user's workstation to avoid bogging down the PDS computing. One currently operating entity of this type resides at UCLA; it is a comprehensive analysis program for vector time series such as magnetometer data. Software in this class must be developed with the approval and guidance of the User Steering Group.

14.6 Display Requirements

Table 2 shows common types of graphics displays for various non-imaging instrument areas. This is roughly graded from simple x-y plots on the left to more complex three dimensional plots on the right. The hierarchy of implementation should also be from simplest to more complex. The browse facility will drive displays of only the simplest sort such as the first two columns of the table, while the analysis software package could drive more complex displays. The levels and types of displays follow the same sort of hierarchy as the manipulation/analysis software. Examples of these graphics can be found in the chapter on hardware, section 2.4.1.

The user's data display needs depend heavily on analysis requirements, instrument type and the form of reduced data. One might require anything from a simple plot of x vs. y to a three dimensional view of a particle distribution function to color contour plots of dynamic power spectra. This hierarchy of display requirements ranges from the simplest graphics shared by all (or most) non-imaging investigations, to non-shared instrument-specific (and perhaps even work station-specific) graphics. These needs can be covered by some reasonably capable and complete graphics package supported either on the system or at the workstation.

14.6.1 Browse Graphics

The non-imaging graphics software associated with the browse facility might be fairly inflexible. It would utilize standard display formats of quantity y vs. quantity x, where x and y are user-selected quantities in the Quicklook data base. Multiple quantity plotting, y1, y2, y3, ... yn vs. x, should also be available with preselected labels and plot scales. These quantities would presumably have been selected by the user in the "data view" definition module. The browse facility, by its nature, should include some limited data definition module, allowing a user to browse only data from a particular time or location. For example, a user may be interested in browsing Voyager I & II plasma data at Jupiter only for times when the spacecraft were near the Io L-shell. Much of this sort of information could be found in Catalog. One enhancement to the bottom level browse display is user-selectable plot scale.

14.6.2 High Resolution Data Graphics

Display of complete, high resolution data rapidly carries us into instrument-specific graphics software. However, all the display elements discussed above are applicable to this database. The user should be able to specify all plot attributes, or let the software create a "default" plot with scales and format determined by the range and type of data to be plotted. Once the user has determined suitable display parameters, the system should be able (upon request) to save these parameters in a file for future graphics use. Display parameters include scales of ordinate and abscissa, number and labeling of tick marks, labels used to identify ordinate and abscissa, whether the plot scale should be log or linear, etc. Many remote sensing display requirements approach the level of imaging: the display of maps, instrument footprints on existing images, and spin-scan generated measurement arrays are in this category.

14.6.3 Interactive Needs

In many cases, a user will wish to display the results of an analysis or some data manipulation he/she has just completed and stored in a workspace dataset. If something has gone amiss in the analysis, the user may not know it until the analysis is done, the dataset written, and the software package invoked to display the result. It would be preferable, in many cases, to provide interactive analysis/manipulation and display. Using the example discussed earlier, the user may wish to take the Voyager Io encounter plasma flow data, remove the Jovian corotation field and transform the resulting vectors into some new coordinate system. First, the Io data from the primary database might be displayed as a time series; after removal of the corotation flow, the new vectors are now plotted for the same time. Finally, after the transformation to the user's new coordinate system, a third time series is displayed, and the user either has the desired result or is learning where he/she went wrong. Users must be able to invoke plotting software at any point in the analysis/manipulation phase.

14.7 Implementation Phases

- o Startup central and discipline center catalogs, and browse software -- possibly using the Pilot Planetary, Pilot Climate or Pilot Ocean Data System.

- Clear needs identified in Phase 1 are addressed, including data access at system level (enhanced catalog, inventory, search/sort capabilities) and data processing at discipline center and user workstation levels (enhanced statistics, manipulation, graphics and discipline-specific analysis tools). Software standards are enforced.

- Needs arising from Phase 2 addressed, including full-up encyclopedia at system level, and further discipline-specific analysis tools at discipline center and workstation levels. At this point calibration software is brought up to process raw (EDR) data.

Table 1. Non-imaging Planetary Data

IN SITU

Fields & Particles	Atmosphere	Surface
B, E fields	Structure	Meteorology
Plasma waves	Winds	Geology
Cool plasma	Neutral Mass. Spec.	Seismometry
Hot plasma	Clouds	
Cosmic ray	Gas chromat.	
Solar wind		

REMOTE SENSING

Radio/radar	Microwave/IR/Vis./UV/X and Gamma Ray
Occultation	Radiometry
Gravity	Photometry
Atmosphere	Polarimetry
Altimetry	Spectrometry
Surface reflec.	Thermal structure
Interferometry	& mapping (IR)
Planetary Radio Astronomy	Lab Spectroscopy

Table 2. Display Types for Non-imaging Data

Data type	Display type	Measured Quantity1 vs. time	Measured Quantity vs. Spatial Variable2	FFT Power Spectrum (power vs. frequency)
Fields and Particles		X	X	X
In-Situ Atmosphere		?	X	X
Surface		X	X	X
Radio		X	X	X
Radar		X	X	?
IR		X	X	?
Visible		X	X	X
UV/X-ray		X	X	?

1 Measured quantities include data processed or semi-processed to some physical level.
2 Spatial variables include altitude, radial distance from planet, L-value, longitude, latitude, solar zenith angle.

Table 2. Display Types for Non-imaging Data (Continued)

Gray Scale or Color Maps	3-d Plots
X	X
X	NA
X	NA
X	X
X	X
X	X
X	X
X	X

15. GEOMETRY SOFTWARE COMMON TO ALL EXPERIMENTS

15.1 Introduction

All imaging remote sensing and in-situ experiments require information about the geometry and location of observations. Valid comparisons of the results of experiments require that the geometric information be internally consistent among them. The basic geometric information, the desired parameters, and the associated software have a great deal in common among all spacecraft experiments. Geometric information commonly changes with time due to improvement in information about the spacecraft position in attitude; if different versions of geometric information are used for different experiments, comparisons may not be valid. This can be very important for some remote sensing observations made near periapsis, where timing uncertainties may be equivalent to many fields-of-view. Hence, comparisons made between experiments may be invalid if different geometric versions are used.

Geometry data has traditionally been delivered via a "supplementary experiment data record" (SEDR) wherein the geometric variables desired by each investigation are calculated by the project on a time basis specific to that investigation and formally delivered along with the experiment data record. (See Figure 2.1 (General Downlink Data Flow), levels 5 and 6G). This process has always been a source of difficulty; in few, if any, cases has a complete and accurate SEDR been delivered at the time agreed upon. There are several inherent problems with this system: 1) information both on pointing direction and spacecraft position tends to improve with time, making earlier geometric calculations; 2) obsolete investigators have had to request all geometric items of any foreseen application, making the volume of geometric data large; 3) the size of the software and management systems and the volume of calculations involved make it impractical to regenerate SEDRs.

15.2 Geometric State

A solution is to identify the fundamental information (the geometric state, or GS), upon which geometry calculations are based and to maintain or deliver these in separate packages which are easily replaced when improved information is available. Along with this geometric state a standard (across most missions) software tools package would be available for calculation of specific geometric parameters used in science analysis. The geometric data and software should be treated in the same manner as an investigation in terms of access to the data, software, and accompanying documentation, data delivery, etc. This new method should gently reduce the cost of "SEDR" generation, in that the same data are supplied to all investigations, and the volume of data delivered should be far smaller on the average.

The geometric information associated with radio and radar experiments represent a special case in as much as these experiments can themselves generate fundamental information about the location of the spacecraft and the ephemeris of the solar system. These experiments have special requirements for geometric information, particularly in terms of the gravitational parameters of the solar system and spacecraft non-gravitational accelerations.

15.3 Implementation

The software system should allow for the incorporation of pointing information updates based on the information from science instruments themselves, such as the location in images of features tied to a geodetic net, or the time of detection of a limb crossing. The software system should be able to interpolate through times when telemetry data related to spacecraft position and attitude may be incomplete. Where possible, the geometric data for past missions should be incorporated into this system.

Some geometry parameters are frequently used in searching thru data, such as latitude and longitude of the center of an instrument's field of view. For practical reasons, it may be desirable to include these parameters in the data files accessed by data base management systems. The decision whether to recalculate observation parameters from the GS every time they are desired or to precalculate and store them as data set parameters becomes a practical decision to be made individually in each instance. The particular parameters desirable for each discipline should be determined by an appropriate science group. In either event, computer variables should always be traceable to the version (date) of the following: the software package, the navigation data, the spacecraft and scan platform attitude data, the planetary ephemeris, the calibration files relating instrument pointing directions to spacecraft or scan platform pointing, the calibration file of physical time-constants used for smoothing, the file of information relating spacecraft clock or instrument counts to Universal Time.

A possible practical implementation would be for the project to provide the basic navigation (spacecraft position), ephemeris (target body positions, shape and orientation) and spacecraft attitude information in whatever coordinates and time resolution appropriate for this information. The software package would be used to construct an intermediate file which contained the spacecraft position and pointing information in an object-body coordinate system appropriate for science analysis (such as: origin at the planet center, Z axis parallel to the planetary spin axis and the X-Z plane oriented to include the sun) and on a time-base appropriate to the science investigations. From these vector and matrix quantitites, all other geometric parameters can be rapidly computed as needed.

Figure 15.1

This scheme can be shown diagrammatically as follows (compare with Figure 2.1)

```
5B BODY EPHEMERIS <U   5N NAVIGATION <U   5O ORIENTATION <U

         |                   |                   |

         |                   |                   |

6B BODY EPHEMERIS      6N NAVIGATION      6O ORIENTATION   6T SOFTWARE TOOLS

      |                  |                  |                  |

      |__________________|                  |                  |

                         |__________________|                  |

                         |                                     |

      Instrument         |<------------------------------------|

        times---------->|                                     |

                         |                                     |

     6.1 interpolate to instrument times                       |

     and transform to science coordinate system                |

                         |                                     |

        6C GEOMETRY COEFFICENTS                                |

                         |                                     |

   Time range and        |                                     |

   parameter request---->|<------------------------------------|

                         |

              get geometry

                         |

              7G GEOMETRY
```

Updates may be input independently for: pointing, at level 6 or 7; navigation, level 5 (only the navigation team usually has the knowledge required for this); body ephemerides, level 5 (rare).

The level 5 and level 6 files are identical, the level 6 files have simply been distributed to the investigation teams. The level 5 files are in that coordinate system native to their calculation (e.g., EME 50).

For imaging experiments where the full geometry is a much smaller data set than the experiment data, computation and storage of all geometry items is advisable.

16. NETWORKING

16.1 Requirements on Networking

In discussion of the PDS design, the term "networking" refers to the process of exchanging data and supporting communications between users, between nodes, and between users and nodes. The term is not restricted to electronic exchange of information but includes all forms of exchange (e.g., mail) required to support the PDS activity. Networking includes the protocols necessary to permit communication and the protocols needed to provide a uniform environment for accessing data (or at least the catalogs). Protocols for using data are discussed in the software sections.

There are various ways in which a network may be used, and a minimum number of "human interface" routines which are required. The expected uses and required functional routines are detailed below.

16.1.1 Use of the Network

The functionality provided by network capabilities is required for successful implementation and utilization of PDS. Frequent interaction between working scientists and various datasets generates the cohesiveness required to maintain enthusiasm for, and participation in, PDS activities. Frequent use also creates an environment conducive to achieving defacto standards. A network enhances the capability to interact with, and utilize resources of originators of datasets - a primary objective of the pilot planetary project.

A slight extension of traditional network services would include the transport of datasets by mail. This is desirable since PDS requires a uniform method for all final disposition of data requests and since, during startup, surface transport will be the principal mode of transmission available.

A discussion of the expected utilization of the PDS network, follows:

a) Resource Sharing

It is seldom desirable to duplicate capabilities at each investigator location. It can be too expensive, too time-consuming, some of the duplicates may be underutilized, required staff may be impossible to duplicate, or the ascent user may simply wish to develop more modern, but incompatible, facilities.

Examples include very high-speed computing (expensive and time-consuming to program), image processing (good staff scarce and expensive), and manipulation of old datasets (some of the computing systems are now simply unavailable and no one would accept them as a gift for their institution).

Another aspect of resource exchange involves utilizing processing capabilities on a remote machine for data generated on a user machine. An example might be using generic PDS routines, resident on a "discipline" computer, to catalog and graph data generated on a smaller or software-incompatible computer.

b) New Data Distribution

The timely distribution of new data is important principally for mission environments. The PDS activity should support this distribution, in concert with missions. It should be necessary only to utilize a set of PDS protocols which are consistent with mission data distribution. The operating costs should be carried by NASA telecommunications services.

c) Data Conferencing

The SCAN network has started testing data conferencing - a concerted research effort by several investigators at, or having facilities at, dispersed locations. Results seem to indicate that this is a productive method of collaboration. It is probable that this will become an integral feature of the network.

d) Software Exchange

ARPANET has been used frequently for exchanging programs between investigators. As previously mentioned, such exchange helps to establish de facto standards by propagating the more functional subroutines.

e) Communication

Good communication is necessary for carrying on the business of science. It is needed for locating data sets, resolving problems, cross-fertilizing ideas and for resolving a wide variety of operational issues. Typically, a form of communication is required which is faster than mail and more reliable than trying to find someone by telephone. Operational experience indicates that network mail provides this popular, heavily utilized service.

f) Queries

Conventional wisdom dictates that on-line catalog queries will be an important part of the PDS function. These queries will require rapid response but transfer relatively small amounts of data. The advantages of maintaining a catalog on-line consists mainly in the ease of update and the ability to search on given parameters. There is also a possibility for outside users to gain quick access to catalogs they might not ordinarily have (similar to long given distance information service by the telephone company).

g) Transfer of Historical Data

It is anticipated that most historical data will be obtained by mail. Transfer of small datasets, or samples of large datasets, may occur frequently. The network upon implementation should facilitate such electronically.

The loss of the above services by failing to implement some amount of electronic networking would increase the difficulties of maintaining cooperation between users and of managing the archive. It is quite possible that the data curation facilities would eventually fail from disuse if paper, magnetic tape, and mail remain the principal modes of communication. Most of the functions listed above could be achieved by dial-up techniques. It will be demonstrated that, except under conditions of trivial use, dialup phone charges exceed the cost of more suitable alternatives.

16.1.2 Required Functional Routines

The communications subsystems should provide mechanism for two basic kinds of information transfer, interactive communication and bulk data delivery. The former provides nearly instantaneous response with a substantially greater emphasis on speed. The latter provides bulk transfer, with lesser emphasis on response time, and must incorporate a variety of media and communication environments.

It is not necessary to define which functions occur on which of the two transfer modes. Rather, the mechanisms for each mode should have the potential for upgrade, expansion, and refinement as usage demands. The query language may initially provide capability that is highly interactive but allows no direct access to the data, an activity that can occur satisfactorily at 1200 bps, but must have good response. Alternatively, all direct access to the data might occur in a delayed response (batch) mode with the results delivered via some suitable medium (paper, microfiche, magnetic tape, video disk, laser disk, or high speed link as appropriate).

The boundaries between the two modes are expected to change (as interactive speeds become higher, and ultra high speed links become available, and as query software becomes more sophisticated), but the two fundamental needs will remain unchanged. Regardless of available communication speeds, some queries will always generate delayed responses (due to processing requirements) for which interactive communication is unreasonable. Even as high speed links become widespread, there will still be a need for the unattended message form of delivery.

Alternatively, even as bulk data delivery mechanisms evolve to higher throughput rates, the need for highly responsive interactive links will remain. This need extends beyond the departmental workstations into the homes of the scientists; its requirements for coverage and response far outweigh its requirements for speed.

There are specific functions and characteristics which must be provided or permitted by the network. These have been summarized in the technology introduction. Their relationship to networking is described here and in Section 16.4 (Selection of transport protocols).

a) File transfer - File transfer is the primary requirement for the PDS network. Virtually all computing CAN be accomplished by this means, though the actual operation can become tedious (and slow) from the lack of an interactive capability.

File transfer requires programs running in both sending and receiving computers. The programs must verify data integrity, permit renaming the file to the receiver's convention, and support format conversion. It is desirable (for the simplest connection) that these programs can transport binary files through terminal handlers (implies conversion to 6 bit). Many of the existing file transfer programs require attended operation on at least one end. PDS transfers will require a method of confirming file source, content, and format (file label conventions).

b) Browse - Browse capability is a stated requirement. The intent is to permit search of catalogs akin to the paradigm extant in libraries. This implies the ability to 'flip' through pages (with books one can achieve a rate of 5-10 per second), to rapidly understand the structure of the catalog, and to support sparse (decimating) searches. Enhanced abilities should include the ability to search on prescribed conditions (as in CA-online or NASA library searches).

There exist distinct technical problems in emulating a library browse. The data rate required to support a 5 page/sec rate is about 6 Kbaud. A typical terminal page contains much less information than a written page. It is difficult to present the spatial organization of, say, 10 pages on a terminal. Delays introduced by satellites (or computer load) are significant compared to the dedicated use of a document. Page replacement and menu selection are very terminal-hardware dependent operations. At a minimum, a prefetch algorithm must be developed to enhance browse operations.

c) Remote session (edit, run interactive) - There is recognized need to support remote sessions. Unusual I/O devices, specialized software, and very high speed computing will require remote access for the forseeable future. Remote access and original data capture (including mail) will be the primary tasks of the PDS network when large local data bases become practical. It is difficult to support remote interactive computing over transport media having significant delays (satellites, and, increasingly, multiplexed telephones) since full duplex character echos slow the effective transfer rate significantly. The problem is worse when a variety of equipment (better stated, a variety of protocols) is used. This is a significant programming problem which will have to be solved, since the alternative (dedicated lines) is expensive and increasingly difficult to acquire. Solutions include 1. Uniform hardware (software) on net with local interface routines to foreign equipment or 2. development of a local echo routine and the necessary protocols (uniform editor, uniform response to control characters and escape sequences).

d) Mail - Mail (and voice communications, if achievable) is necessary for smooth functioning of the data system. Operational matters such as error reporting, help requests, and event notification are required. In addition, good communication between working scientists is extraordinarily important. If a proprietary network, or a transportable O/S is chosen for transport protocol, then providing mail service is trivial. If a development path is chosen, then a mail system which works on a variety of computers is required. In such a case, a telemail-like implementation (message center) would probably be the most cost-effective choice.

e) Accomodate substantial delay - For a heavily used network, satellite connections are already by far the most cost-effective transport method. There exists a substantial delay (about 1/2 second) for signal travel time and switching. Telephone connections are being multiplexed (or transmitted by satellite) more frequently. The turn-around time introduces substantial delay (which may be circumvented by using two circuits). Finally, connection through several computers (a common DECNET implementation) can introduce significant delay. This implies that the networking protocols chosen should accomodate substantial delay. This impacts interactive modes severely.

f) Accomodate 'foreign' terminals - It is desirable to accomodate 'foreign' terminals on the net to protect existing investments, to avoid sole source headaches, and to retain the ability to introduce new technologies. Most operating systems are notorious for being dependent on terminal characteristics, usually peculiar to the manufacturer's hardware.

g) Accomodate change - The first handheld calculator was introduced about 10 years ago. Since that time, 'cheap' computing power has decreased at least an order of magnitude in cost, size, and power consumption. Advances have not been confined to a single company - in fact, some of the most significant advances (e.g., the 68000) have been made by companies which were not known for computer technologies a decade ago. It is a reasonably safe assumption that these trends will continue. To avoid having the network stranded by technological advance, the system should be able to accomodate change.

There are several precautions which can be taken. 1) The cost of equipment and specialized software should be minimized to minimize the agony of abandoning out moded systems. 2) Dependance on a single manufacturer should be minimized. 3) Complexity (and therefore, presumably, services) of the system should be minimized. Example: the transportable executive TAE provides many convenient features. Unfortunately, it has proven to be very expensive to develop is not yet transportable (depends on VMS), is too large to fit on many machines, and is complex enough so that it is difficult to modify. The network software should avoid these pitfalls.

h) Permit use of existing software packages - If a transportable operating system is chosen, it may reduce the number of software packages (e.g., IMSL, BMD, special routines) which may be used.

i) Security - The network must protect connected systems from unauthorized use. Concerns include malicious mischief, uncontrolled use of computer resources, and avoidance of improper commercial use.

j) Data rate (1200, 4x real-time to 56K) - The network should support 1200 baud as a minimum rate (300 should be supported, but not encouraged). The 1200 rate should be supported in the spirit of the "free" NSSDC distribution, that is authorization for use should be easily obtained, connection should be trivial, and most services should be available (catalog, some computing - with prior agreement).

Higher data rates should be available as required for mission support or complex tasks. A goal should be that mission data can be delivered at least 4x the data rate. This could be relaxed for extremely high data rate experiments (such as imaging). Current costs dictate that the maximum distribution rate outside of NASA centers should be limited to 50-100 kilobits/sec.

k) Charging algorithm - The network should support a charging algorithm. This permits monitoring system performance, allocating resources, avoiding saturation of the system due to misuse, and establishing priorities for network use (for mission critical deliveries). A recharge algorithm also permits sharing the network between several agencies (which is perhaps desirable for linking universities together).

l) Transporting other protocols - Transportation of other protocols is an important service of the net. It is probable that dispersed data analysis groups having similar equipment will wish to network those computers in a native environment. This permit utilizing the full range of native-mode services for the particular project.

16.2 Assumptions

Several assumptions must be made for the purposes of network design.

16.2.1 The distribution of NASA investigators is assumed to be that given in the SYSTEM90 report (Attachment 1). Ninety percent of investigators reside at 30 institutions. This fact simplifies system design, since a relatively costly facility can be shared by several investigators via a local area net (LAN).

16.2.2 The distribution of computer types in the planetary community is assumed to be that given in the PPDS report (Attachment 2). The computer manufacturer is predominantly DEC. Note that the report reflects acquired, and in most cases, aging computer systems. Conventional wisdom has it that most of the community plans to buy a VAX and can afford at most a 68000-based system.

16.2.3 The network should encompass as much of the planetary community as possible as quickly as possible. This is necessary so that the PDS system can be validated by a large number of users as concepts and implementations are developed. This may imply that the startup choice should be a proprietary net. This startup system should evolve to a system having;

a) A network-specific packetizing scheme having 1) universal network addresses (which allow transporting "foreign" packets on the network) 2) universal network data identification (source and type - e.g. SFDUs);

b) Protocols which support direct and delayed connections (see requirement e.) above;

c) Protocols for data conversion (i.e. a P-code-like set of data protocols) and;

d) A set of transportable standard software.

The required set of software is described in the software introduction. The set which should be directly tied to the network includes executive services and a simple editor. An optimistic network estimate of software development time is on the order of 15 man/years.

16.3 Selection of Transport Method

16.3.1 Transport Media Options

There are several choices for transport media.

1. DIALUP has the favorable properties of low fixed monthly costs, ubiquitous availability, and low-cost hardware. It has the problems of high noise, low transfer rate (this is improving), relatively high connect costs, and, increasingly, difficulty in supporting a full duplex mode (switched voice circuits or satellite delays form unsatisfactory dialup connections).

2. TELENET-type links are somewhat less expensive than dialup. Connection to foreign countries is relatively simple. High rate connections are as difficult to establish as leased lines, and, in many cases, are more expensive. Costs must be considered at the user end (very low for 1200 baud dialup, high for a 9.6 Kbaud PAD), and at the computer end (high).

3. TAPE is an effective, and well parameterized, medium for data exchange. Transfer costs are non-trivial by the time material, copy, and shipping charges are totaled. In addition, total system throughput is typically unsatisfactorily slow.

4. READ-ONLY SATELLITE connections show promise. Costs are not significantly different from leased lines but bandwidth is higher and such connections assume an assymetric data load: more in than out, or vice-versa. A development effort is required before such connections are commercially available.

5. LEASED LINES provide transfer rates up to 9.6K without becoming prohibitively expensive. Difficulties include a long lead time for installation (and a substantial installation charge), moderately expensive modem equipment, and fixed, point-point routing.

6. VIDEO DISK appears to be a promising low cost medium for LARGE datasets. Delay times are probably similar to those for tape.

7. SATELLITE connections are available for about $2500/month. Advantages include high bandwidth, high connectivity (i.e. can address many stations), and efficiency due to packet organization. Disadvantages include high costs at low utilization and inherent delays due to travel times. Locating the (3 meter) dish can also present problems.

16.3.2 Transport Costs

There are alternative methods of designing the PDS network. The most straightforward is that of designing a network built for the exclusive use of PDS. This net design should not be implemented (because of wasted bandwidth) but serves to illuminate potential costs to outside users. A second alternative is to design a network excluding consideration of the physical datalinks. The physical links would be provided by a CODE T project (Program Support Communications Net) to NASA investigators. This is a more appropriate approach, since the PSCN bandwidth, as well as most development costs, would be shared by a variety of applications. A third design involves the utilization of a variety of existing networks to carry a PDS "virtual" network. This design has would make PDS available to a wider community, and would increase the potential for a cooperative development effort with DOD (which is currently very active in internetting activities).

The strawman design given in Section 5.4(?) uses a combination of these options - a virtual PDS net is carried over the PSCN network to active planetary scientists. External or new users gain access via commercial networks (such as Telenet) or long-established networks (like ARPANET), which have gateways into the PSCN. We consider here the appropriate transport methods for various conditions of use of the PDS net. The following paragraphs consider the component costs of networking.

Transport costs are the major cost consideration for a network. Monthly line rentals quickly exceed the cost of hardware at the termination points. Evaluation of transport costs is complicated by the fact that the amount of information to be transferred by the PDS network is indeterminate at this time. The best approach to evaluating transport costs is to develop a relatively simple algorithm for determining cost-effectiveness of various transport mechanisms for various loading conditions. This algorithm is summarized in Attachment 3,(?) a graph of cost/month versus the amount of data transferred. Costs are a combination of initial hardware costs plus monthly lease costs, plus any charge per data transferred. The costs for zero bytes transferred represent hardware + lease costs (purchased equipment is depreciated over 5 years). The slope on this log-log plot is determined by the cost for transfer. In the case of dialup, this is the average connect cost (about $20/hour), for telenet, this is the packet charge ($12/megabyte), and for mailed tapes it is the media charge + mailing charge + copy charge (about $2/megabyte). Leased 9.6K lines have been assumed to cost $1000/month. This price varies. The cost is generally about $1 to $3/mile-month.

Note that four of the transport media (read-only satellite, mail video disk, leased line, and leased satellite) are very insensitive to the amount of data transferred. It is clear that these transport methods are preferred when the data volume is high (i.e. greater than about 100 megabytes/month). This is the case for distribution of most project data, transmitting pictures, and large file transfers (greater than 10 tapes). The charge-per-hour (or packet) services are most attractive when the transfers are small. A megabyte represents about 1000 screen refresh operations. For a 1200 baud connection, a refresh takes about 10 seconds; a megabyte represents about 2.3 man-hours of browse. From the graph we see that it would be cost-effective to use dialup as long as the browse (or edit) hours remain below about 50/month for a long distance connection.

Data rate is a serious concern for some operations. A 10 second refresh operation is not particularly satisfactory when one is searching catalogs or large lists or when one is performing a large edit. Increasing the bandwidth of the link for these intermittent operations increases the fixed cost of volatile connections over those indicated in the graph. Packet networks become much more attractive (in theory) under these conditions (though present high rate packet-net charges generally prohibit realizing the theoretical saving). Leased lines also become more attractive when effective throughput of the dial-up connection is low, as indicated by the histogram in attachment 5-connect time costs.

The conclusions reached at the Workshop are:

1. If the connection is intermittent or involves small data transfers, then dial-up or, if the costs at the computer end are paid elsewhere, dial-up telenet is the connection of choice.

2. If connect time exceeds 50 hours/month or the data transferred exceeds 12 megabytes/month and the connection is used each month, then leased lines are more cost-effective. Telenet provides a savings if the amount of data transferred lies between 12 and 80 megabytes/month.

3. Tape is only marginally cheaper than leased lines, and becomes more expensive than leased lines for data transfer when data quantities approach 120 megabytes/month. If the connection is intermittent, then tape is a better choice.

4. Satellite connections are cheaper than leased line when data quantities exceed a gigabyte per month or when the cost of the leased line exceeds about 2300/month or if several connections are required.

16.3.3 Network Terminal Costs

Access to the net requires at least a terminal (modem costs are included in transport). The simplest access requires an ANSI terminal (assuming full screen capability). Cost is about $1100. Graphics access has a minimum price of about $3k. If significant delays exist (as for satellite access), then local intelligence is required. Simple buffering can be done with a personal computer (about $2k). If protocol conversion is required, then a faster machine is required ($5k-10K). Machines in this price range can support proprietary network protocols, and should be regarded as an entry-level communications station.

Entry-level communications stations require a reasonably fast CPU. Available hardware includes 68000, 8086, and 11/23 CPUs. These CPUs are also adequate for workstations, so terminal hardware requirements span interests of the group at this point.

Massive protocol conversion can be accomplished in hardware. Prices start at about $40k, so this type of high-speed conversion should be contemplated only for very heavily used nodes.

Satellite earth stations include local intelligence for the purposes of dividing bandwidth and entering data onto a local area net (LAN). It should be possible to use this equipment for protocol conversion and traffic accounting. It may not be cost-effective to do so, however.

16.3.4 Selection of Transport Method

Consideration of transport and terminal costs, and speculations on the evolution of transport media, lead one to conclude that a mix of transport methods must be used. Occasional connections or low- volume connections should use dialup or telenet-type connections. Intermediate volume frequently used routes should use leased lines. High volume traffic should be conducted by satellite. Multipoint connections should generally use telenet-like or satellite communications.

The picture is complicated somewhat by the need for centralized network control. The centralized control is needed for maintaining routing addresses, controlling access (password control), and network mail services. This star configuration is useful only for the above services. Once a route is established for a session, it is undesirable to route all traffic through a central node - both because of bandwidth limitations, and because of potential increases in transport costs (routing data from Pasadena to Los Angeles by way of Huntsville is an example of this). The star-like access, since it is of low volume, could be maintained by dialup.

When the methods are available, existing networks should be utilized to transport the selected protocol. ARPANET connects to most universities and may provide reasonable cost routing. The planned PSCN will provide many economies in routing. The cost of connecting to the PSCN nodes may not be cost-effective for all users.

Startup routing should utilize dialup, telenet, and leased lines where they presently exist. A satellite net should be implemented between the 30 largest users (see SYSTEM90 final report) on the times scale of 1-2 years. The satellite net sets an upper limit on the cost of providing service to a node at about $2500/month (excluding network development and control, which is a cost common to all methods). It presently appears that this high speed 'satellite' net will be provided by the PSCN on an appropriate timescale.

16.4 Selection of Transport Protocol

16.4.1 Minimum Required Software

The PDS system must have a uniform access method to be usable. Any large diversity of response when accessing various nodes would make the system unwieldy and difficult to learn. Such a system would be underutilized.

The services which must have uniform characteristics (at least from a network point of view) are those provided by most computer operating systems. A detailed list of services required by PDS occurs in the software introduction. This set of software lies at about level 6 in the ISO network model (see Section 16.4.3).

16.4.2 Choice of Transport Protocol

Protocol can be selected by choosing a proprietary (hardware unique) system (e.g., DECNET, SNA), by choosing a transportable operating system (e.g., UNIX with Usenet or Berknet), or by choosing a transportable applications package (e.g., ARPANET, "NASA development"). A summary of the advantages and disadvantages of each method follows:

PROTOCOL METHODS:
SUMMARY OF CHARACTERISTICS

Advantages

Proprietary

*Fast to implement
*Lowest startup cost
*One brand already predominates
*Many services provided
*Transportable Applications may be implemented within system
*Provides defacto standards

Transportable O/S

*Has advantages of both Proprietary and Transportable Applications
*Available on micros to mainframes

Transportable Applications

*Greatest generality
*May be tailored to specific needs
*Perhaps more flexible to change
*Fewer sole source problems

Disadvantages

Proprietary

*Sole source problems
*May lockout new technology
*Ties national system to single company
*Present systems don't accomodate delay
*Probably impossible to connect over some existing networks

Transportable O/S

*May be expensive to adapt some centers to UNIX
*UNIX has different implementations on different hardware
*UNIX may interfer with existing local software
*Questionable support for many UNIX machines

Transportable Applications

*Implementation time may exceed useful lifetime
*Probably costly
*Historically very complex to modify
*NASA must bear design costs
*NASA must bear maintenance costs
*Protocols must be chosen

There are performance requirements against which these three options may be tested. These have been detailed in Section B and are listed in Table 4-1.

TABLE 16-1

PERFORMANCE REQUIREMENTS FOR TRANSPORT PROTOCOL

PERFORMANCE REQUIREMENTS FOR TRANSPORT PROTOCOL	PROPRIETARY	TRANSPORTABLE OPERATING SYSTEMS	TRANSPORTABLE APPLICATIONS
a. File transfer	provided	provided	existing or trivial software
b. Browse	provided	provided (editor)	existing or trivial software
c. Remote session (edit, run interative)	provided	provided	difficult to accomodate many computers
d. Mail	provided	provided	difficult to generalize
e. Accomodate substantial delay	not yet provided	not yet provided	difficult s/w, can be included in design
f. Accomodate 'foreign' terminals	choice restricted	many choices	inclusion implied, trivial
g. Accomodate 'foreign' computers	inclusion difficult	inherently provided	inclusion implied, but formidable task
h. Accomodate change	at mercy of company	reprogramming possible	at mercy of complexity
i. Permit use of existing packages	available in some cases	very few supported	hard to port
j. Provide security to connected systems	generally breakable	generally weaker than than for proprietary systems	rarely good during development
k. Provide adequate data rate	yes	yes	probably
l. Provide an accounting mechanism	supported	supported	programming required
m. Transport 'foreign' protocols	not supported	not supported	programming required

16.4.3 Selection Of Transport Protocol

The transport protocols must provide the services detailed in a fashion that is reasonably transparent to distributed implementation and to users. A brief discussion of the ISO model will help show what protocols must be adopted.

Table 16-2 shows the ISO model and the DNA model (from Low and Perry).

TABLE 16-2

	ISO Model	DNA model	PDS Problems
7	Application	USER Network Applications	Imaging, non-image software
6	Presentation		see 'minimum required software' section
5	Session	Session	
4	Transport	End Communication	provide error-free data
3	Network	Routing	addressing, flow-control
2	Data Link	Data link	form and read packets reject errors
1	Physical	Physical	buy hardware that is compatible

The minimum required software for the net (detailed in section yclept SESSION)(?) should be a package of transportable software that is developed for PDS net. This SESSION package provides a common user interface for the search and browse functions. This SESSION package must be developed or adapted, and can be developed independent of the choice of transport protocol. Transportability, and the space limitations of many existing or inexpensive machines, dictate that the SESSION software be small and truly a minimum set. Extensions to that set could be allowed for a small number of larger machines.

Excellent discussions of the model in Table 4.2 exist. Two articles well worth reading for this context are 'NETWORKING DEC AND IBM COMPUTERS', W. H. Mish, GSFC, and 'TRANSPORT AND INTERNET PROTOCOL EVALUATION FOR TACCN', D. L. Gallop, JPL. The Gallop article examines the relative merit of public net protocols (TCP/IP, UNET, X.25, NBS). Proprietary protocols (SNA, DECNET) are treated less completely, because of the need (for that survey) of linking a multitude of types of computers together. The article concludes that the DOD TCP/IP protocols (essentially ARPANET) provide the most robust and complete services at this time. It also projects that TCP/IP is unlikely to become an international or national standard, when committees finish their work (ca. 5 years from present).

The network splinter group agrees with the conclusion that a modified TCP/IP protocol is the preferred way to connect diverse computers. For startup implementation, DECNET is probably the least expensive choice (from both hardware and software considerations). Some strong drawbacks of the DECNET approach are the potential for technological stranding, such as, dependence on a single company, and the resource-intensive nature of DECNET (a single active session consumes currently 27% of a VAX780 cpu, additional sessions SPAN costs consume about 1% per session).

A large proportion of the user community owns some DEC equipment so that (refer to PPDS data survey), entry-level equipment is relatively inexpensive (about $10K), and a large number of required software services are provided, the Network splinter group recommends that the PDS net first be implemented as DECNET. During this implementation phase, the PDS SESSION software should be developed and beta-tested. The DECNET transport layers (4 and below) should be replaced by the TCP/IP standards as quickly as possible (1-3 years). These protocols, and the SESSION software, should be implemented for most or all types of computers in the PDS net on a similar time scale.

16.5 Internetting and Resource Sharing

Sharing resources with respect to networking can encompass a number of areas, including transport media, software modules or systems, planning and development efforts, and perhaps hardware items either by actual sharing or by combining purchase orders, etc. Efforts to achieve such sharing must be ongoing, to include continuing contact with organizations and agencies having related and compatible communications needs. These presently include several organizations within NASA, communities of university scientists (UCAR and NCAR, e.g), DOD (ARPANET and MILNET), and possibly NSF.

The earliest efforts should be directed toward joint planning and development efforts, potentially yielding great benefits in design costs. Secondly, there are various opportunities to share communications media, ranging from the simple expedient of subscribing to a packet switching network (thereby sharing the backbone network with other subscribers), to utilizing existing proprietary networks (such as ARPANET), to convincing some agency (NASA, NSF) that an appropriate satellite transponder should be set aside for use by the scientific community at large, and thus be used for communications as envisioned here. Even if all network related software cannot be obtained from academic organizations or commercial sources, any required developments should of course be targeted toward compatibility with as much existing hardware and software as possible.

There are many perceptions among NASA researchers as to what a network is and what it can do. The most prevalent view is that "a network" should be a multimegabit broadband data and image transfer vehicle. Another view is that "a network" would tie together high-powered computational tools. A third view is that "a network" was a tool to provide data cataloging, storage, and remote-access transfer and retrieval. Others say, "There is no need for a network" and then indicate they are avid users of Telemail for electronic communication and heavy users of direct-dial lines to access computers remotely-clearly networking activities. There are also prevailing views that "a network" is a single entity, and that its cost would be very great, and the research programs could not bear the burden. This section of the report addresses these perceptions and concerns by investigating strategies for interfacing existing equipment, and research centers, and for providing a variety of resources at reduced cost through internetting.

The PSCN will provide a network that will support many of the requirements voiced. It will be a broadband satellite network suitable for data collection, transfer, storage, retrieval, and analysis. In addition, it should be able to support communication functions such as electronic mail, remote host connections, and terminal and graphics workstation access. The PSCN is not yet in place and probably will not be fully operational for three to five years. Therefore interim strategies are needed. We have seen that the views of different researchers about networks follows the blind-man-and the-elephant parable. If NASA headquarters is building the "elephant", it is imperative that "the elephant" be a hard working pacyderm and not preferably not white. This will only be true if the potential subscribers to the PSCN network provide NASA Headquarters with recommendations for protocols, services, configuration, and management above the backbone level <u>now</u>. NASA Headquarters will soon be reviewing proposals from bidders to build the PSCN. The Planetary Data System (PDS) working group (as well as other NASA research groups) could benefit by participating in this review process if at all possible, and should petition for a mechanism by which this is possible. If review of the initial implementation of the PSCN is not possible, then the PDS should be planning and formulating its recommendations for management, configuration architecture, and equipment above the backbone level, and make their needs known to NASA Headquarters as a group.

The first order of business would be to survey the PDS constituency to find out just what networking services are currently being used and for what. Such a survey would also be extremely useful to researchers who need to make networking decisions now. They might well find an interim solution already exists, and thus save money and effort by combining forces; or they could choose strategies that are compatible with existing tools and upwardly compatible to the PSCN.

Important questions to ask are:

- How many users/centers are covered by existing network services?
- Do these existing services get the job done now?
- Are there existing services now that are better or cheaper?
- Can these existing services be shared with others by internetting or other means?
- Does everyone know they exist?
- Is cost data available?
- Are these services open-ended or closed systems?

A second order of business might be to survey existing hardware/software, and research programs that are now on, or are candidates for immediate connection to a network.

Questions to be asked here are:

- Are the existing networks suitable to support the hardware/software or program.
- Are there resources or researchers on other networks that are accessible from existing networks through internetting.
- Are the plans for connection upward compatible to the PSCN with a minimum of disruption or expense. If not, would it be a reasonable decision to delay network connection until the PSCN is available.

These two surveys, taken a matrix of who is on which net doing what, can be generated. This can be used for several applications:

- To identify which network(s) are providing useful services now
- To indicate where internetting or gateway access would be useful
- To identify who can communicate with whom, and what computer resources are available
- To plan strategies for porting the PSCN.

16.6 Conclusions

16.6.1 Primary Functional Needs

The networking splinter group concludes that there are significant needs for the transfer of data within the project, and that these needs must be addressed in a comprehensive and structured fashion. Thus, there is a PDS "network" design whether or not it includes a collection of high speed links or other components that sometimes constitute what is termed a network. That is, the network concept includes data movements of all kinds, including non-electronic transfers such as the mailing of magnetic tapes.

The data flow which this system addresses does not include the primary ingestion of raw data, but does include all transfers of data among major centers (computing and archiving) and users. It also includes a variety of transactions that occur between people and computers, sometimes over great distances. The network design addresses the following primary functional needs:

1. Datasets must be moved between various major computer centers (ingestion sites, PI processing locations, archives, etc.) and to end user systems. For most applications the time scale for delivery is on the order of days, the quantity of data is on the order of several to tens of megabytes, and the need for integrity is very high.

2. Users need to have interactive access for learning about available datasets and requesting their delivery. The interactive nature requires nearly instantaneous response, the quantity of data is on the order of tens to hundreds of bytes per transaction, and the need for integrity is moderately high. Note that this item refers to information about datasets, not to the contents of those datasets, hence the low volume.

3. Investigators must have access to computing systems on which datasets can be manipulated, processed, and examined in various ways. Depending upon the distribution of people, functions, and processing power, this may be accomplished or it may require significant remote computing. In the latter case, the response time must be on the order of seconds to hours, the quantities of data are on the order of hundreds to thousands of bytes, and the need for integrity is moderately high. Users will need to examine many datasets graphically. In case such access is remote, the data transfer problems can be considerable -- response time must be on the order of seconds or at most a few minutes, data quantities are on the order of ones to hundreds of kilobytes, and the need for integrity is moderately high.

4. Communications between people form an integral part of any scientific endeavor, especially one requiring collaboration and sharing of resources. This report excludes consideration of voice communications, but electronic messages are important and require data flow comparable to that of item 2.

16.6.2 Meeting Functional Needs

These functional needs must be met within a realistic framework of cost restrictions and existing or available systems and components. The conclusions of this group were reached under the assumption that costs per user institution should be on the order of a few hundreds of dollars per month. However, the group feels strongly that this figure is marginal (unless extensive collaboration with other network organizations is utilized) and that reluctance to fund communications may result in considerable hidden costs such as for tape/disk drives, tape/disk media, tape storage, computer operators, tape/disk handling software, error recovery efforts, losses in the mail or due to physical damage, and human frustration and loss of productivity due to the inherent latency of all non-electronic delivery methods.

The group recommends that electronic media be applied to all four of the above functional needs, at least for the time frame beyond 1986. It is recognized however, that funding constraints may prevent short term realization of this goal for transfers that require transmission rates higher than 1 or 2 kilobits per second. This largely affects only functional needs 1 and 4, delivery of datasets and use of remote graphics.

The overall recommendation of the splinter group is one of sharing and contibuting to the network resources of one or more cooperating organizations. Primary candidates (in order of desirability as presently perceived) are NASA's PSC network, DOD's ARPANET, (and NASA's SCAN network), and the potential for and the ramifications of participation must be pursued vigorously. Ideally, a suitable collection of internetwork arrangements (gateways) would facilitate mutual sharing of resources among all three of the aforementioned networks.

Factors that must be considered in selecting a network (or several networks) in which to participate include: coverage of pertinent institutions, especially the "main" computation and archive centers; widespread availability of interactive access (e.g., via telephone even from a scientist's home); generality and standardization of protocols and interface requirements; the provision of high speed services as needed for dataset and graphics transmission; the overall integrity of the network, including its ability to deliver error free data and the experience and performance record of its governing organization; and of course the initial and ongoing costs of access and participation.

The networking splinter group believes that that nearly all of the previously stated functional needs can be met for most users via reasonably priced electronic means. In fact, the costs may fall within the desirable range (a few hundreds of dollars per month per user institution) provided participation can be realized in the previously mentioned NASA or DOD networks. However, the following cautions should be observed:

a) Widespread interactive access (comparable to Telenet) is not presently planned for the PSC or SCAN networks. This form of service is essential and should be obtained by separate contract if necessary (with, e.g., Telenet, Tymnet, or Uninet).

b) The administrative and physical details of network access must be explored thoroughly to ensure against unanticipated snags and delays. Of special concern are hardware and protocol compatibilities at all levels, and costs for links, modems, and interface units.

c) The major NASA and DOD networks may leave some communities without high speed service such as is needed for dataset and graphics transmission. Alternatives to be considered for such service should include receive-only schemes such as being considered by NCAR or mobile equipment for requirements of short duration.

d) If used for transmitting images or other high volume datasets frequently, the network may experience considerable unexpected congestion. Careful traffic estimates should be prepared and updated regularly as information to the management organization for the network services.

e) It is unclear whether the networks being considered for participation provide adequate reporting of activity by user, especially for purposes of accounting and charge-back.

f) Even if no physical network management system is required in the planned system, there is a need for administrative management, to address such issues as accounting, charge-back, access permissions, network addresses, user satisfaction, load requirements, user assistance, and network information. This may pose significant difficulties, especially in a multi-network environment.

As mentioned above the ideal networking solution arises from an internetwork arrangement that draws upon the strengths of several existing and planned systems. Among the significant advantages of this approach are:

a) Major computer/archive centers, various research centers, and individual users can each determine the most appropriate network connections based on their own needs.

b) The distinct strengths of all three networks can be applied as appropriate. Examples include the SCAN (DECNET) system's comprehensive capabilities for resource sharing and interprocess communications and its ease of connection to Digital equipment that is in widespread use; the ARPANET system's adherence to widely adopted and powerful protocol standards and its versatility with respect to internetwork connections; and the PSC system's planned capability for high speed data transfers and its sources of funding.

c) Dial-up access is clearly facilitated and reduced in cost by an internetwork arrangement, provided the gateways permit establishment of virtual circuits. In particular, the ARPANET's plans for Terminal Access Controllers (TACs) may obviate the need for subscribing to commercial packet switching networks (such as Telenet) for obtaining universality of access (both national and international).

In conclusion, significant benefits at reasonable costs an accrue from PDS networking efforts. Appropriate funding should be commited and suitable people should be identified to pursue the concept as sketched above and detailed in the remainder of this document.

Appendix 1. Existing and Planned Networks

There exist several networks which presently exist, or are in advanced planning stages, which could carry PDS data will varying degrees of support. Most of these networks encompass a substantial portion of the nodes envisioned for PDS, so add-on costs should be small. These nets could be used for test purposes, startup, or expanded to support PDS activities for the forseeable future.

A. The Space Plasma Computer Analysis Network (SCAN) is a network which links together computers used for space plasma research. SCAN currently features a modified star topology using DECnet and dedicated 9600 baud lines. The central node (NEEDS) is located at Marshall Space Flight Center. This network is nation wide with a Telenet gateway to France to be opened soon. Some networks available to SCAN-users through gateways are ARPANET, SU-NET, the Los Alamos local area net, Telenet, etc. Current uses of SCAN include corrolative analysis of spacecraft data from DE, ISEE, IMP, ground based radar measurements, Shuttle PDP, Voyager, with further use expected in the fields of Planetary and Spacelab data analysis. Mostly funded by burden and hope.

B. Telenet, Tymnet - Commercial packet switching networks. Connections available by dialup, leased line, or local PAD (packet assembly-disassembly) + leased line. Uses X.25 protocol (more or less). International connections one available. Average distance for connections is 441 miles. Costs: dial-in ($3.00/mo access fee), 9.6 connection or computer connection (around $1500/mo), data transfer cost ($12/megabyte.

C. NCAR - The National Center for Atmospheric Research is beginning to design a network to link it's substantial computing capacity with a number of universities. Many of the requirements and destinations appear to be similar to those of PDS. NCAR has a fair amount of experience of linking dissimilar computers together (e.g., IBM - defacto standard, DEC - numbered backwards, CDC - large word size). The connections are for automated file transfer. The necessary data conversions are carried out in hardware. NCAR presently uses a commercial packet-switched network and dial-up for remote 'publc' access.

D. PPDS - The Planetary Pilot is developing a network sufficient to provide proof-of-concept. Present plan is to implement DECNET in 9600 baud dialup.

E. PODS - The Pilot Ocean project has implemented a small network with dedicated lines. Protocol is DECNET.

F. PSCN - The Program Support Communications Network is substantially funded by NASA Telecommunications CODE T. It is planned to provide a backbone network for transferring data, voice, and video between (about 14) NASA nodes. The RFP is just about in press at this time. This RFP includes a development phase for the successful bidder. The PSCN may mature at about the time the PDS net is ready for full implementation.

G. ARPANET/MILNET - the first, and currently the largest, heterogenous packet-switched store-and forward host-to-host network. It was designed and built in 1970 under the direction of the Defense Advanced Research Projects Agency (DARPA) as a research experiment to test the feasibility of the packet-switched architecture and design. A working group of scientists from universities, military and government agencies (including NASA), non-profit research establishments, and industry were involved in this design and implementation experiment which proved to be wildly successful, since ARPANET/MILNET was the forerunner of most of today's packet-switched and virtual circuit networks. The designers and builders of the network were also its users, so they incorporated many features designed to assist working scientists and engineers.

In 1974 the management of the ARPANET was turned over to the Defense Communications Agency (DCA), and at that time it became an operational military network. Many hosts were added and its use expanded rapidly. In 1979 the transport protocols (TCP/IP) were adopted as DOD standards for all military (and many government) communication networks. This prompted many vendors to incorporate the protocols into their vendor products so that they would be DOD-compatible. In 1982 the Defense Data Network (DDN) was established as an "umbrella" network incorporating all of the military networks (such as ARANET, MILNET, COINS, DODIIS, WIN, MINET EDN, etc.) intermitted together by means of TCP/IP protocols.

In August 1983 the ARPANET split into two networks, the ARPANET R&D network and the MILNET operational network. Both are managed by DCA's DDN Program Management Office (DDN-PMO); however DARPA sets policies and conduct networking and related research on ARPANET, and collaborates with DCA and other military agencies in transferring useful technology into operational systems.

DARPA has large research efforts in interneting, wide band satellite communications, packet radio communications, artificial intelligence, network protocols, gateway design, electronic messaging, ULSI, graphics, robotics, network standardization, and very large data base handling. Since this research is government-funded a wealth of resources is available in the public domain for use by other government agencies such as NASA.

Neither ARPANET nor MILNET are classified networks; however their use is restricted to the conduct of government business so they are government, not public networks. Military and government agencies provide the sponsorship (funds) to run ARPANET/MILNET. NASA is one of these sponsors. DOE, NBS, and NSF are others. Many large universities have been connected to the network from the beginning, and its users include scientists and engineers and students from many disciplines other than computer science.

ARPANET/MILNET was designed to be a resource-sharing network. It was also designed for operability and survivability, with an extremely robust architecture. It is comprised of more than 100 node computers, called Interface Message processors (IMPs) and TACs are BBN-C/30 computers, manufactured by Bolt, Beranek and Newman, Inc. (BBN), with bacbone 50kb Telco lines. Eight (or more) computers can be attached to each IMP. The same nodes are available commercially for use in private nets or LANs. (IMPs and TACs were originally developed with government funding, but are now commercial products. Technology transfer from ARPANET/MILNET to the public sector has been very similar to that of NASA for its space research).

As stated before the network is managed by the DDN-PMO. Network Services are provided by BBN, who provides the Network Control Center (operations, maintenance and analysis) and SRI International (SRI) who provides the Network Information Center (host name service, online directory service, protocol depository, network newsletter, information services.) Both organizations are under contract to DDN-PMO to provide these services.

There are many features of this network of interest to NASA scientists and administrators:

1) It is operational and NASA already sponsors and is on this net.
2) It supports wideband communications
3) It is reliable
4) It is one of the largest, most geographically accessible nets in existence (CONUS, Hawaii, Europe, and Korea with other access imminent).
5) It allows connection of virtually every kind of computer and operating system to every other kind. (DEC, IBM, CDC, Amdahl, HP, Xerox, Data General, etc.
6) It's protocols support internetting (tying one network to another via gateways).
7) It permits users on a local mesh to connect to a remote host, do work, and transfer the results back to the local computer interactively.
8) It has a wealth of public-domain software that is easily downloaded across the net.
9) It supports file transfer (FTP) which lets a user on one machine push or pull over files. to/from other machines regardless of machine word size or format.
10) It supports electronic mail (which started on ARPANET/MILNET) including some multimedia mail.
11) It is internetted to many of the world's major long-haul computer networks.
12) It supports one of the world's leading research programs in wideband Satellite communications.
13) Many of its users are scientists engaged in research of interest to the NASA scientific community.
14) Many major universities (e.g., MIT, Stanford, CMU, Columbia, USC, UCLA) are connected as well as most high-energy nuclear research labs (e.g. Los Alamos, Livermore, LBL, Argonne, Brookhaven, Natl. Physics Laboratory, NYU.)
15) Many commercial and not-for-profit firms (e.g., Bell Labs, SRI, RAND, MITRE, DEC, IBM, Lockheed, TRW, Aerospace, BBN) have access as government contractors.
16) It's charging algorithm (at this time) does not pass down to individual users, so there is not "meter running" in the commercial sense while it is in use.

17) It's cost is shared by all the sponsors, so the total burden would not be on NASA alone.

18) There are many collaborative computer science research efforts under way on this net in which NASA scientists could participate.

19) Its use does not rule out the use of other networks.

Appendix 2. Network Control and Services

What network control consists of is highly dependant on the size, implementation, topology, uses, etc. of the network. There are probably two major ideas as to network control. One is the idea of an "operations center" where the network is constently monitored for performance, routing control, etc. This would probably fit in with the idea of a star topology where a single central node would control the entire network. A very different type of net control would be that of a distributed network with no hard-and-fast "central" or "controlling" node. The topology of such a network would have nodes connected in a way such that if a single node dropped out of the network for some reason, an alternate route would still exist for the rest of the network to function through. The impact of such a loss on the rest of the network would thus tend to be minimal. Network control in such an environment would be directed more towards network planning and coordination with a lesser amoumt of involvement with minute-to-minute operations.

Many network control functions would be the same regardless of the type of network. Some of these functions would be: 1) the coordination of network node addresses, 2) coordination of communications services between nodes, 3) installation and maintenance of network software in association with remote node managers, 4) definition of network parameters such as those relating to logical line cost, time out values, etc., 5) handling of network trouble reports such as bad lines, nodes that have gone down, speed problems, etc., 6) overall performance monitoring of the network to determine if there are any communications bottlenecks or resource overloads. The goal of any management effort should be to keep the network up and running, thus maximizing network resource availability.

17. PLANETARY SCIENCE ANALYSIS SUPPORT SYSTEM: HARDWARE REPORT

17.1 Introduction

This section describes the computer hardware requirements for a planetary science data analysis support system. Present practice, state-of-the-art, and predictable developments are considered. Specific recommendations are presented for those who plan to acquire new computational tools, those who must install and use them, and those who pay for them.

17.2 User Requirements

The computer systems required to process and analyze data for planetary science users must perform several functions:

1. Give Access to Data

2. Operate on and Transform Data

3. Store Data

4. Display Data

Planetary science data comes in many types from scalar or vector time series spectral plots to multi-dimensional remote sensing data. Each discipline analyzes data in different ways, thus the four functions above change greatly depending on who uses them.

17.2.1 Access to Data

The nature of data studies has changed in the past few years. Previously investigators used only observations from a single instrument on a single spacecraft study. Now more sophisticated studies require data from several sources. This makes additional demands on data access.

Data is generally available from a distribution center in discrete units: a frame for imagery, the entire data set, or a substantial fraction of it for non-imaging data. Orbiting spacecraft typically return far more data than fly-by missions. The orbiting mission data unit may be a single orbit. Whatever the unit, we assume a specific discipline scientist makes requests for data in a regular pattern. That is, the researcher requests a unit of data at discrete intervals.

(Data per Request) X (Request/Unit Time) = rate of Data Delivery

Our experience is that time for analysis is typically the limiting factor in research: we assume that the time to deliver is short compared with the time between requests. Both parameters affect the choice of transportation media or communications speed.

Examples of Data Access Requirements

1) Fields and particles: projected use of Galileo magnetometer data

   10**6 words X 32 bits/word X 12 requests/year = 5 X 10**8 bits/year

2) Imagery: Viking Orbiter research

   8 bits/pixel X 10**3 pixels X 10**3 lines X 3 frames/request X 5 requests/year = 2 X 10**8 bits/year

Access may be accomplished through the use of tape drives, optical disk readers, floppy magnetic disks, remote communications facilities or a combination of these.

17.2.2 Operations on Data

The system must transform raw data into meaningful physical parameters and perform any special computational procedures required by a given analysis, as quickly as practical.

Operations on data may be divided into two types: preprocessing to a form suitable for analysis and analysis procedures to aid in the physical interpretation of data.

1. Preprocessing of Data

The processing of raw data from planetary missions into physical parameters requires access to considerable computational power. For some instruments the data archive contains raw unprocessed data and calibration coefficients or calibration code to turn this data into physical parameters. For example, because of the high rate of data delivery, broadband plasma wave data is fully processed by experimenters only for intervals of special interest.

2. General Analysis Processing

Typical space science analysis functions include: fast (near instantaneous) display of graphs and images, contrast enhancements, algebraic transforms, windowing or browsing in the data, geometric warping transformations, coordinate rotations, noise analysis, fast Fourier transforms, spatial filtering, statistical and numerical analysis and mosaicing of images.

Analysis operations frequently require comparable computer power to that required to process the data. To see this, let us do a simple order-of-magnitude calculation. Given an image 100 pixels square, execute a convolution that requires 10 operations per pixel, each operation taking 10-5 sec. We have 10**6 pixels X 10 operations/pixel X 10**-5 sec., or 100 seconds per convolution. This time increases quickly for larger images (Landsat Thematic Mapper frames have over 4 X 10**3 pixels) multi-band images, repeated operations or very complex (geometric) transforms. Such speed problems are common to all image processing systems, since most computers, however fast, execute only one operation at a time. Only array processors or powerful vector processors (Goddard Massively Parallel Processor (MPP) or Gray) escape this bottleneck.

17.2.3 Local Data Storage

Local data storage divides into short term active storage and longer term archival storage. A research group may need both a large disk storage system for frequently used data and tape or writeable optical disk for longer term archival storage.

1. Working Storage

During analysis the researcher needs to maintain small subsets of the data and several transformed versions of that data for immediate access.

Working storage required = (Subset size) X (Versions) X (Number of different subsets)

2. Archival Storage

Archival storage is a function of the role played by the user's laboratory. If the lab is a curatorial facility for a data set, it will have a complete set of all that data set, plus processed versions of all or part.

Archive storage required = (Data set size) + (Data set size) X (versions) X (version % of data set)

A smaller laboratory may store only a subset of any particular data set, with accumulated versions as required by the level of analysis activity.

Archive storage required = (Subset size) + (Versions) X (Data unit size)

3. Examples of Local Storage Requirements

Working Storage

Event storage of magnetic field data can vary from ^10**5 bits for a single event study to ^10**8 bits for a long statistical study.

Archival Storage (curatorial facility)

The magnetic field data archive from Galileo will total ^10**10 bits. The entire low rate science archive from Galileo will total 5x10**11 - 1x10**12 bits

Archival Storage (ordinary laboratory)

The Galileo magnetometer will produce data at about 5x10**8 bits/month.

17.2.4 Data Display

Planetary science data is displayed in many ways. These range from simple one-dimensional plots to three-dimensional displays with color and shading.

Many planetary science non-imaging data sets are time series. Graphs of various parameters versus time are perhaps the most common type of plot in planetary science. An example of this is seen in Figure 1 where Pioneer 10 magnetic field data from Jupiter have been plotted versus time. Most graphics requirements can be met by using similar formats although more complex formats are used also and will be used more in the future. Frequently it is useful to look at 3D plots from more than one perspective (Figure 2, Sentman et.al., JGR,86,7487,1981). Contour plots of the same data are given at the left. In Figure 3 we have an example of the use of color to display data in the third dimension. These are data from the plasma wave experiment and the retarding potential ion mass spectrometer (RIMS) on Dynamics Explorer (DE) (courtesy of S. Shawhan).

17.3 Hardware Issues

Rapid hardware change is now predictable, if not controllable. Given the expense of writing software, tying an applications package to a specific hardware configuration may in time leave the user stranded with obsolete, expensive to maintain hardware. Thus software and application interfaces that move easily from system to system are most desirable.

17.3.1 Choosing Hardware

In spite of this need to preserve software compatibility, scientific users may be compelled to purchase imcompatible hardware for several good reasons. Planetary science researchers have severely limited budgets, and equipment costs may be a larger percentage of grant funds. The researcher is under great pressure to purchase the most cost effective hardware available at a given time. And at any given time one manufacturer or another may be ahead in this contest. There are yet more mundane factors at work. A particular manufacturer may have general discounts available at the moment, or will extend discounts to researchers. Manufacturers may be convinced to donate equipment. When an entire system is purchased, the way a manufacture or distributor bundles its components will affect price.

17.3.2 Hardware Futures

The use of mainframes and shared access mini-computers for research is well known in space science. Yet the expense of setting up and maintaining such systems has discouraged many from acquiring local computing power. The personal computer revolution has encouraged manufacturers to design ever more powerful microprocessors, and these can now form the heart of an inexpensive yet powerful single user scientific workstation.

17.3.3 Maintenance

Hardware maintenance is typically 10%-20% of the initial hardware purchase price per year. As equipment ages it requires more service, and service providers may either refuse to continue to maintain older hardware or raise prices to high levels. In the last two decades, a pattern has emerged. Economic equipment lifetime is typically less than five years, shorter for mechanical peripherals such as disk or tape drives. New technology is usually more reliable and thus less expensive to maintain than older equipment of the same performance level. The cost both to lease and maintain improved hardware may be less than that of simply maintaining the older hardware.

17.4 I/O Requirements

A prime constraint on hardware systems for image analysis is the data volume that must be accommodated. Digital images from remote sensing satellites such as Landsat can have 10 ** 10 bits of information per scene.

17.4.1 Magnetic Media

Current practice for mini-computers is to use magnetic tape media for data storage and distribution. Although many different media such as floppy disks, data cartridges, etc, are in use, the media cost of 1/2" 9-track magnetic tape remains least expensive per data volume. It is the common denominator, readily available, low cost, and moderately long lived. The major expense in using magnetic tape is the cost for the tape drive.

Floppy disks and data cartridges provide read/write capacity of 0.1 - 2 megabytes (floppy) and up to 50.0 megabites (cartridges) of data per individual media. These media are convenient and inexpensive for low volume storage programs and data, but unsuitable for the larger image data sets. Although the hardware costs are low, they have limited portability and compatibility since there are so many varieties of each.

17.4.2 Optical Storage

The analog videodisk player is a valuable peripheral to support workstations analysis with its high capacity and low cost (<$2k). A complete archive of planetary images is available on two double-sided laser-disks (200,000 images). Images on the disk can be readily displayed following a data base search. While the resolution of images on the disk is limited, it does provide an excellent browse tool for selection of images or image subsets for digital processing.

Digital encoded data on analog videodisk, digital audio disks and digital write/read optical storage systems are all looming on the horizon. These optical disks are not yet available, but will be in the near future. The potential for storing gigabytes of image data on a single disk makes this optical storage of great interest for archival storage and distribution.

17.4.3 Local Area Networks

Although systems with local tape or disk storage may stand alone, several workstations may be connected to a local file server via a network. The file server manages expensive peripherals such as tape drives, magnetic disks, optical disks, printers and other hardcopy units, and provides these resources to individual workstations. Networks with file server support permit diskless workstations to be used. These diskless workstations may be suitable for graphics applications and low resolution imaging. Because workstations can be upgraded with disks as needed, a natural growth path is provided.

The Ethernet protocol which specifies a coaxial connection of 10 megabit bandwidth between devices is an industry standard. In spite of the current diversity of software protocols, Ethernet remains the best way to interconnect diverse equipment with a high bandwidth link, though software limits present individual user throughput to 40-50 kilobytes/second.

17.4.4 Remote Communications

Workstations will still need access to the non-local data. This could be direct line access to local mini or mainframe equipment, access via communication nets to other groups, and/or dial up access to and from remote laboratories.

Telephone system based links at T1 bandwidth (1.544 megaband) or dedicated data lines (56 kbaud) are means of high bandwidth interconnections. Low speed links use low speed asynchronous modems, direct connections or packet witch links. These technologies are standardized and well understood but suffer from bandwidth limitations. Transfer time for even a 512 x 512 x 8 bit image at the 9600 baud data rate typical of asynchronous links takes approximately ten minutes.

Some modes of research that require extensive computations or modeling will need access to substantial computing facilities. These may be local facilities or national centers that have existing large computers now on-site.

17.4.5 Hard Copy

Aside from the digital data forms, there is often a need for film output and other hard copy output. High quality laser film writers are so expensive (>$50k) that they would be difficult to justify for a single workstation. Black and white and grey scale "off the video screen" copiers are available at reduced costs, but suffer from fading media, low resolution and low contrast (washed-out). Similarly, a variety of low cost printers are available that can generate black and white or crude gray scale images.

Medium-resolution (512x512 to 1024x1280) color film writers can be had with a variety of film backs for 35mm, 4x5 polaroid, 8x10, etc. Even though expensive (>$8k), one unit can be shared among a few display units by use of a simple switch. Lower cost (<$3k) 35 mm film writers are also available with lower resolution (480 line). Another alternative for color hardcopy image storage is video tape. Professional quality video recorders can be used for storage, editing and even film production with 512x480 broadcast TV standard images. For a modest system, satisfactory images can be photographed directly off the display screen face if a suitable hood is used. For graphs, medium cost (<$3k) pen potters produce publication quality output (.001" resolution).

17.4.6 Interactive Input

Interactive devices such as light pen, digital tablet, mouse or trackball should be used wherever available to enhance user access. These devices work naturally in the "point at what I want" mode which eliminates much typing and chance for error.

Alphanumeric screen display, whether using a separate monitor with key board or an overlay plane is basic. A bit-mapped graphics screen is useful both because of its inherent rapid graphics display capabilities and because of the flexibility it allows in the presentation of simultaneous text and graphics.

17.4.7 Image Display Characteristics

Images may be displayed with a variety of hardware from dedicated image devices to simple add-on boards. For the low to medium end workstation the add on boards and simple frame buffers are the most likely options.

Costs for image display hardware are directly related to resolution and number of available colors or shades of grey (bits per pixel). Costs are typically linear with number of colors, but not so with image resolution. Video display driver circuits and display memories must all have much higher I/O bandwidth, an expensive attribute, for higher resolution.

Complete low-end (512x480x4 bit) image systems can be had at reasonable prices of $4 - $6k. The high end systems cost in the $70k range for 1024 x 1280 x 12 bits. The cost of the work station thus is a very strong function of the display resolution. Low-end graphics needs are well met by monochrome displays that sell for $2k - $5k (all based on early 1984 prices).

17.5 Work Station Characteristics

The word "workstation" is defined in this discussion to mean a single user environment that provides data access and display capabilities. As such, the term accommodates devices that range from a simple terminal with local or remote slow speed (1000 character/second) connections to a micro-based processor with local disk storage capable of stand alone operation. Device capabilities range from simple monochrome line graphics through very high resolution multicolor image display stations with hardware image processor assists. Prices also cover a substantial range from $3k for the low-end graphics terminals to $100k and more for the high end image systems.

Which workstation is chosen for a particular project is largely a function of the type and volume of data that must be accommodated and the resources that are available. If several classes of data, as described earlier, are to be processed, then clearly the system must be sized for the most demanding application. Funding constraints may require compromises where a compatible mix of systems are selected to deal with a range of data types.

17.5.1 Hardware Categories

The following Sections describe several Categories of Workstations.

1. Graphics

Most of the graphic work station requirements can be met by graphics terminals connected to minicomputers or by microprocessor based work stations. Recently low cost graphics terminals have become available which can serve many of the scientists day to day needs. Terminals like these are the lowest resolution divides which are suitable for use in planetary science studies, and represent the lowest level workstation for non-imaging data.

2. Low-End Image/Graphics

Microprocessor based systems also can satisfy the graphics requirement. Even small systems now have low resolution graphics boards which give comparable resolution to the graphics terminals mentioned above. Some microprocessor systems also can be used for low-end image processing applications. These low-end systems feature moderate display resolution (256x256 or 240x320) a minimum of 16 colors or shades of grey and 10 or 20 megabytes of local disc storage. This configuration may communicate with the planetary data network for queries and data extraction, manipulation of subsets of digital images (line plots, contrast stretching, etc.) and development of software and algorithms for image analysis.

The low-end image system is configured to be a minimum system used for display of image segments with the processing power to do rudimentary image processing tasks (stretches, etc.). It trades off processing speed and convenience for low cost.

3. Mid-Range Image/High End Graphics

Medium power microprocessor based work stations like the SUN system can meet the high end graphics requirement and the mid-range imaging requirements. They have high resolution graphics (1152x900 in black and white and 480x640 in color with up to 256 colors). A stand alone system with 1 megabyte of memory and 50 megabytes of disk storage costs about $25 with educational discount. These systems can also be networked together, using Ethernet protocols, to a common file server. This provides distributed computational power an allows cost sharing of the more expensive disks, tapes, printers and other peripherals.

4. High-End Image

The high-level workstation provides 1024x1280 resolution with 12 pixels, color or monochrome image display with graphics overlay capability, large local disk storage (300-500 MByte) and specialized hardware capabilities, such as an array processor and mass storage devices (digital videodisk, etc.).

17.5.2 Performance

True real-time response is unlikely for all but the simplest scientific image processing task. Mainframe or super-mini computers typically execute instructions faster than a workstation, but large systems which are shared by many users often prove slower in apparent response to a specific user. To improve workstation performance, several paths may be tried. The standard microcomputer families have shown steady increases in speed as their manufacturers respond to competitive pressure. Second, an associated processor, such as a floating point micro circuit is becoming common for advanced micro computers. Third, an array processor unit may be put on the workstations' computer interconnection bus. Such a unit can offer the speed, for floating point calculations, of a super-minicomputer. Last, special purpose custom integrated circuits or video-rate processors may be produced for specific operations, such as geometrical transforms.

17.5.3 Benchmarks

The following tables from "The Micro-Computer Workstation", W.K. Erikson, L.B. Hoffman, W.E. Donovan (NASA/AMES) show the results from running some standard benchmarks. All 68000 times are based on a 10MHz version with fast memory. To place things in perspective, the HP 3000 Series III is a high-end 16-bit minicomputer ($100K system cost). The SEL 32/77 is a mid-range super-minicomputer ($15K), the DEC VAX 11/780 is a high-end superminicomputer ($300k) and the CRAY 1S is a state-of-the-art supercomputer ($6M). All times mentioned are wall-clock times, with one user on the machine. The rate of change in the field is such that these figures will soon be obsolete, but they do provide a snapshot in time of the capabilities of these processors.

These are all run on unloaded processors, and are thus unfair to the 6800-based workstation, since the larger systems must run many programs simultaneously to justify their costs.

Single Precision Whetstones (Double precisions, not vectorized)

68000 Software Floating Point (SFP).................... 45,000/Second
68000 Hardware Floating Point (HFP)................... 120,000/Second
HP 3000 Series III (HFP) 220,000/Second
SEL 32/77 Firmware Floating (FFP) 500,000/Second
DEC VAX 11/780 HFP 1,150,000/Second
Cray 1S 15,600,000/Second

All 68000 HFP times are estimated from 8 MHz 68000 values

Note: the Whetstone benchmark is a standard benchmark written in Fortran used to evaluate the floating-point capability of computer systems. The more Whetstones a second, the better.

Ethernet Transmission Times (68000 based workstation)

Time to transmit
1,000 bytes .... 0.85 seconds
10,000 bytes .... 1.6 seconds
100,000 bytes .... 7.3 seconds
1,000,000 bytes .... 65 seconds
40,000,000 bytes .... 45 minutes (estimated)

Note: the values above were obtained with a stopwatch and reflect the actual time elapsed, all overheads included. The two workstations involved were connected via 1,000 feet of Ethernet cable strung mostly underground between two buildings.

17.6 Example Workstations

Workstations come in a variety of shapes and sizes. The following section describes a low-end graphics capability and three basic classes of stand alone workstation and their areas of application. The second part of this section describes several realizations of these workstations using current technology.

No recommendations are implied by the choice of hardware. There are several dozen manufacturers making workstations, some with excellent built-in network support and attached (array) processors.

Table 1 contains general hardware specifications for three categories of stand alone systems: low, medium, and high end. The specifications are based on 1983 technology and by no means cover the myriad of processor and bus combinations.

The low-end graphics system consists of a raster or vector terminal which can be attached to an existing micro, mini or main frame. It represents the most cost effective way to augment existing computer capabilities. If stand alone capabilities are required, the mid-range system, without the image display, provides a cost effective solution.

The low-end image system is configured to be a minimum to be a minimum system used for display of image segments with the processing power to do rudimentary image processing tasks (stretchees, etc.). It trades off processing speed and convenience for low cost.

The middle system is configured to be used by the "average" scientists and has the capability to handle all image processing tasks. The availability of virtual memory and reasonable processing power makes these systems the functional equivalent of larger systems. This system does not include high resolution display or hardware compute assists but these options can be added as needed. For purely graphics applications that don't require color display the image buffer and monitor can be eliminated to effect a cost savings.

The high end system provides a complete high resolution image processing facility with compute power equal to a mid-range super-minicomputer. It is suitable for intensive image processing applications and the attached array processor allows even 2-D FFTs and filtering operations to be done in a timely fashion.

Table 1

HARDWARE SPECIFICATIONS
(IMAGING AND NON-IMAGING)

	LOW-END NON-IMAGING ONLY	LOW-END IMAGING ONLY	MIDDLE-END	HIGH-END
Processor		8086/8088	68000	68010
Clock Speed		5MHz	10MHz	12MHz
Memory		256 K	1Mb.	2Mb.
Disk Storage		10 Mb	80 Mb.	450 Mb.
Tape Storage		none	9 TRK1600 BPI	9 TRK1600 BPI
Display Resolution	640 X 480	240x320	512x512	1024/1280
Bit Depth	1	4	8	24
Input devices		Joystick	Mouse/Graph.Tab	Mouse/Graph.Tab
Hardcoy	Electrostatic Printer	None	35 MM Film	Matrix Camera
Printer		DOT Matrix	DOT Matrix	Dot Matrix
Monitor		Monochrome	Medium Res RGB	Hi-res RGB
Hardware Options				Array Processor

Table 2 is a hardware configuration table for existing systems. The table is divided into three categories: low, medium, and high end stand-alone systems. The table is by no means complete, but it dies cover the range of reasonable configurations. In addition, the options table gives approximate costs on items that are not essential but can increase the speed or convenience of a workstation. These hardware configuration are for stand-alone systems. Groups of work stations may be tied together in a high-speed local area network thus reducing the cost per workstation by sharing expensive peripherals, i.e. disk, tape, printers and cameras.

Table 2

HARDWARE CONFIGURATION

	Low Non Imaging Only	Low Imaging Only	Medium	Medium	High
Name	*	IMBPCXT	Sun	Micro-Vax	Jupiter 12
Processor		8088	68010	VAX	68010
OP Sys		MS-DOS	UNIX	VMS	UNIX
Memory		256KB	2MB	1MB	2MB
Disk Storage		10MB	80MB	56MB	474MB
Tape Storage		none	9/1600	9/1600	9/1600
Display Resolution	640 X 480	240x320	512X512	512X512	1280x1024
Bit Depth		4	8	8	12
Communication Baud Rage	9600	1200	9600	9600	9600-56K
Input device		Joystick	mouse	mouse	mouse
	Optional		BITPAD	BITPAD	BITPAD
Image Hardcopy	Optional	none	35mm	35mm	MATRIX CAM
Printer		DOT MATRIX	DM	DM	DM
Monitor		Monochrome	Medium	Medium	High
			RES RGB	RES RGB	RES RGB
Array Processor		none	none	none	yes
Terminal		n/a	1100X800	VT240	1024X800
			integral		integral
		IBMPC	Multibus	Q-BUS	Q-BUS
Cost	3K	7K	32K	32K	70K

* The low-end non-imaging workstation is a medium resolution graphics terminal attached to a minicomputer. These are available from a wide variety of vendors for example: TEKTRONICS 4006, VT 240, HEWLETT-PACKARD 2623A, VT-100 type terminals with graphics board.

Hardware Options Table

1.	High resolution color monitor	$ 3-7k
2.	Digitizer Pad (11" x 11")	$ 0.8k
3.	Ethernet cable interface	$ 1.5k
4.	Modem (1200 baud)	$ 0.5k
5.	Printer (dot matrix)	$ 1.0k
6.	Array Processor (Multibus or Q bus)	$ 5.0k
7.	Floating Pt. Processor	$ 1.0k
8.	Analog Video Disk + RS-232	$ 1.0k
9.	Video camera	$ 3.5k
10.	35mm camera	$ 0.3k
11.	Mainframe Bus Adapter	$ 2.0k
12.	1Kx1Kx8 bit image plane	$ 4.0k
13.	Frame Grabber	$ 0.5k
14.	Scanner Digitizer	$ 20.k
15.	Pen Plotter	$ 4.0k

APPENDICES TO RADIO SCIENCE

APPENDIX A

This appendix is a partial list of members of the planeary radio science community, their affiliations, and interests. For those who have produced data at least one instrument is indicated. Interests are not meant to be exclusive; many researchers are active in more than one field.

Scientists who were contacted directly and contributed to preparation of this report are denoted by *. Those who participated in the Radio Science Splinter Group at the PSASS Workshop are denoted by (WS).

Investigator (Affiliation)	Instrument	Interest
	CELESTIAL MECHANICS	
*J. Anderson (JPL)	Spacecraft	Masses
F. B. Estabrook (JPL)	Voyager	Gravity waves
*B. Reasenberg (SAO)		Gravity fields
I. Shapiro (SAO)		Celestial mechanics
*B. Sjogren (JPL)	spacecraft	Gravity anomalies
M. Standish (JPL)		Ephemerides
F. Sturms (JPL)		Ephemerides
*S. Synott (JPL)		Satellite motions
	RADAR (ACTIVE RADIO) ASTRONOMY	
*D. Campbell (Arecibo)	Arecibo	Venus radar maps
J. B. Cimino (JPL)		Venus atmospheric occultations
*P. Clark(WS) (Murray St.)		Lunar, Mercury radar maps
T. Croft	Pioneer Venus	Interplanetary plasmas
J. Cuzzi (Ames)	Voyager	Saturn's rings
G. Downs (JPL)	Goldstone	Mars
C. Elachi (JPL)		Radar imaging
V. Eshleman (Stanford)	spacecraft	Atmospheres
*P. Ford(WS) (MIT)	Pioneer Venus	Venus surface
*J. Garvin (Brown)		Surface properties
T. Gehrels (Arizona)		Comets
T. Gold (Cornell)		Planetology
R. M. Goldstein (JPL)	Goldstone	radar
C. Hamilton(WS) (JPL)	Spacecraft	
*J. Harmon (Arecibo)	Arecibo	Mars, mercury

J. Head (Brown)	VRM	Venus geophysics
J. Holberg (USC)	Voyager	Saturn's rings
W. Hubbard (Arizona)		atmospheres
D. Hunten (Arizona)		atmospheres
*R. Jurgens(WS) (JPL)	Goldstone	Venus radar maps
P. Kamoun (MIT)	Arecibo	Comets
W. Kaula (UCLA)	Apollo	Laser altimetry
*A. Kliore (JPL)	spacecraft	atmospheres
A. L. Lane (JPL)	Voyager	Saturn's rings
G. Lindal (JPL)	spacecraft	occultations
J. Lissauer (Ames)	Voyager	Saturn's rings
M. Malin (ASU)	Goldstone	Venus geophysics
E. Marouf (Stanford)	Voyager	Saturn's rings
H. J. Moore (USGS)		Lunar, Mars radar
*P. Mouginis-Mark (H.I.G.)		Venus, Mars radar
*S. Ostro (Cornell)	Arecibo	Asteroids
*A. Peterfreund(WS) (Brown)		Surface properties
*G. Pettengill (MIT)	Pioneer Venus, Arecibo	radar
R. Phillips (LPI)	Apollo	Radar sounder
J. Pollack (Ames)		atmospheres
L. E. Roth (JPL)		Mars topography
C. Sagan (Cornell)		Planetology/ S. Saunders (JPL) VRM Planetology
*G. G. Schaber (USGS)		Surface properties
*R. A. Simpson(WS) (Stanford)	Arecibo, spacecraft	Mars radar
*B. Singer (H.I.G.)		Mars surface
*S. Solomon (MIT)		Geophysics
*D. Sweetnam (WS)(JPL)	spacecraft	Occultations
*T. Thompson (JPL)	Arecibo	Lunar radar maps
*G. L. Tyler (Stanford)	spacecraft	Bistatic radar
J. F. Vesecky (Stanford)	spacecraft	Solar wind
R. Woo (JPL)		atmospheres, ionospheres
S. S. C. Wu (USGS)		Topography
S. Zisk (Haystack)	Haystack	Lunar radar maps

RADIO (PASSIVE) ASTRONOMY

*J. Alexander (Goddard)		Jupiter decametric radiation
M. A. Allen (JPL)	VLA	
V. Boriakoff (Cornell)	Arecibo	Interplanetary plasma
W. J. Borucki (Ames)	spacecraft	lightning
F. Briggs (Pittsburgh)	NRAO	Saturn's rings
J. Caldwell (SUNY)	VLA	

W. Coles (UCSD)	UCSD	Interplanetary plasma
E. Danielson (Cal Tech)		
*I. De Pater (Arizona)	VLA	Radio mapping
M. D. Desch (Goddard)	Voyager	Radio emissions
*J. Dickel (Illionis)	VLA	Jupiter
J. Fix (Iowa)		
M. Gordon (NRAO)	Kitt Peak	Radio Astronomy/*S. Gulkis JPL
	VLA	Jupiter
B. Irvine (U. Mass)	VCRAO	Comets
W. Jaffe (Space Telescope)	VLA	Outer Planets
M. Janssen (JPL)		Venus atmosphere
T. V. Johnson (JPL)	Galileo	Outer Planets
K. J. Johnston (NRL)	VLA	Asteroids
M. L. Kaiser (Goddard)	Voyager	Radio emissions
S. Keihm (PSI)		
D. L. Matson (JPL)	Voyager	Outer Planets
*D. Muhleman (Cal Tech)	VLA, Owens Valley	Atmospheres, surfaces
K. S. Noll (SUNY)	VLA	Uranus
*F. P. Schloerb (U. Mass)	FCRAO	Comets
P. Shelus (Texas)	McDonald	Laser ranging
E. Silverberg (Texas)	McDonald	
J. Warwick (Colorado)	Voyager	Radio emissions
J. Welch (U.C. Berkeley)	Hat Creek	

APPENDIX B

PLANETARY DATA SETS

The following catalog of planetary radio science data sets is in a VERY preliminary state. Quality of the catalog varies considerably among its divisions. For example, only the best known and most widely used lunar radar data sets are included, while most of the entries under "Earth-Based Radio Observations" were culled from summaries of observing programs published in the Bulletin of the AAS and may not even represent viable data sets. The listing under "Earth Based Radar Observations - Mars," on the other hand, is almost complete. Considerably more work will be needed if and when PSASS is implemented to identify further the condition of these and other data sets.

Listings are brief and contain the following information:

1) Investigators - either reporting on or conducting the observations;

2) A three-entry code giving observing wavelength (cm), the spacecraft and/or observatory involved in the observations, and the data product.

3) Observing dates.

4) Usually a reference publication, but sometimes a specific measurement objective.

5) A code giving the status of the data set.

Several of the entries are given as abbreviations; see the next few pages for explanations.

Two supplements are included. Supplement B1 gives NSSDC radio science data sets. Supplement B2 is a list of observatories which might have taken planetary data. The latter is intended to point future detectives toward data sets which were not found in this search.

We have not attempted to include Soviet ground based radio observations; those have been conducted on at least the Moon and Venus (Kuz'min) and the Galilean satellites (Pariskii) but would be difficult to acquire. Nor is our effort for other countries very complete; the interested reader is referred to Supplement B2.

Observatory Codes

A	Arecibo Observatory (PR)
Ap	Apollo spacecraft
Bell	Bell Labs (NJ)
CL	Clark Lake Radio Observatory (CA)
DSN	Various stations of NASA Deep Space Network
EC	El Campo (TX)
EISCAT	European Incoherent Scatter facility
Ex	Explorer spacecraft
FC	Five Colleges Radio Observatory (MA)
G	Goldstone (CA) DSN station
H	Haystack Observatory (MA)
HC	Hat Creek (CA)
KP	NRAO Kitt Peak (AZ)
Luna	USSR moon series spacecraft

M	Mariner spacecraft
Mars	USSR Mars series spacecraft
McD	McDonald Observatory (TX)
MH	Millstone Hill (MA)
MP	Max Planck (Germany)
N	Nancay (France)
P	Pioneer spacecraft
Pleas	Pleasanton (CA) radar
PV	Pioneer Venus orbiter
PVp	Pioneer Venus probe(s)
SU	Stanford University (CA)
UCSD	Univ. California at San Diego
UF	Univ. Florida Radio Observatory
USSR	Unspecified earth stations in the USSR
UT	Univ. Texas Radio Observatory
Vik	Viking orbiter spacecraft
VikL	Viking Lander spacecraft
VLA	NRAO VLA (NM)
Voy	Voyager spacecraft

Data Types

Images	two-dimensional maps (something vs position)
R	radar ranging (power vs time)
RD	radar range-Doppler data (power vs time vs frequency)
S	spectra (power vs frequency)
T	spacecraft tracking data (range or Doppler residuals)

3D three-dimensional maps (power and altitude vs position)

Check the bottom of each catalog page for more specific information on data types within each division.

Data Status

1 Could be easily incorporated into PSASS now.

2 Relatively easy to incorporate; would require some tidying up and documentation.

3 Worth incorporating but would take time to recover formats and documentation (format information and documentation material is believed to exist)

4 Major effort required to recover, but it could probably be done.

5 Recovery unlikely or not worth the trouble

6 Data destroyed or otherwise known to be lost (e.g., recycling of tapes)

ND Suffix ND indicates data not presently in digital format.

NSSDC Data already at NSSDC. See Supplement 81 for data type.

EARTH-BASED RADAR OBSERVATIONS - MERCURY

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(S)	Data Status
Zohar + Goldstein	12.5/G/RD	1970-74	AJ, 79, 85	2(?)
Downs	12.5/G/RD	1981	not published	2
Harmon + Campbell	12.6/A/RD	1978-83	Bull AAS, 15, 837	2(?)

EARTH-BASED RADAR OBSERVATIONS - Venus

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Campbell	70/A/Images	pre-1975		2(?)
Campbell + Burns	12.6/A/Images	1975-83	JGR, 85, 8271	2(?)
Jurgens et al.	12.9/G/3D	Mar-Apr 77	JGR, 85, 8282	2(?)

NB: There are many more Venus data sets.

Data Formats: Images generally give radar reflectivity vs position on the surface. Venus 3-D images by Jurgens et al. also give elevation vs position. There is additional data in the form of elevation/reflectivity/roughness triplets vs (latitude, longitude) along linear ground tracks. Some data may exist in "depolarized" mode.

EARTH-BASED RADAR OBSERVATIONS - Moon

Observer(s)	Wavelength(cm) Observatory	Date(s)	Reference(s)	Data Status
Thompson	70/A/Images	11/66-9/69	The Moon, 10, 51	1(?)
Thompson	750/A/Images	Mar 70	Icarus, 36, 174	1(?)
Zisk	3.8/H/Images		The Moon, 10, 17	1(?)
Evans + Pettengill	3.6/Pleas/R 68/MH/R 784/EC/S	Sep 61 11/61-4/62 Jan - 2/62	JGR, 68, 423	
Evans + Hagfors	23/MH/R	Feb - 3/65	JGR, 71, 4871	
Shelus + Silverberg	*/McD/R			

*Optical laser ranging to reflectors placed at Apollo landing sites.

NB: There are MANY more data sets. Most activity was pre-1970, however, and it is likely that those data sets would be difficult to recover.

Data Formats: Images generally give radar reflectivity vs position on the surface. Data from Evans et al. is received power vs time; these data may not still exist, or may not exist in digital form. Data of Shelus and Silverberg is in unknown format; this is an active data set, however, so its conditions is believed to be good. Some data sets include depolarized data or images.

EARTH-BASED RADAR OBSERVATIONS - Mars

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Carpenter	12.6/G/S	Apr - May	unpublished	3ND
Pettengill et al.	3.8/H/R,S		AJ, 74, 461	
Goldstein et al.	12.5/G/R	May- 6/79	Radio Sci., 5, 475	
Rogers et al.	3.8/H/R,S	May-7/69	Radio Sci., 5, 465	
Downs et al	12.6/G/RD	1971	Icarus, 18, 8	1(?)
Pettengill et al.	3.8/H/R	Jul- 9/71	Icarus, 28, 22	
Pettengill	70/A/RD	1973	unpublished	
Downs et al.	12.6/G/RD	1973	Icarus, 26, 273	1(?)
Pettengill	3.8/H/R	1973	unpublished	3(?)
Simpson et al.	12.6/A/S	8/75 -7/76	Icarus, 33, 102 Icarus, 36, 153	3
Campbell	70/A/R	10/75- 1/76	unpublished	
Downs et al.	3.5/G/RD 3.5/G/S	10/75- 3/76 May - 6/76	Icarus, 33, 441 Icarus, 33, 441	
Simpson et al.	12.6/A/R 12.6/A/S	Jan - 7/78 Apr - 6/78	unpublished JGR, 85, 6610 Icarus, 49, 258	4 3 3
Downs et al.	3.5/G/RD	1978		
Harmon et al.	12.6/A/S 12.6/A/RD	Feb 80 1980	Icarus, 52, 171 EOS, 61, 1020	
Downs et al.	12.6/G/RD	1980	JGR, 87, 9747	
Downs et al.	12.6/G/RD	Feb - 3/82		
Harmon et al.	12.6/A/RD	1982		
Harmon et al.	12.6/A/?	May 1983		

Data Formats: Early data is either power vs time or power vs frequency, giving basic scattering information about the planet. More recent (RD) data can be (has been) sorted to give scattering information (elevation, reflectivity, and roughness) along ground tracks. Some depolarized data may be available.

EARTH-BASED RADAR OBSERVATIONS - Galilean Satellites

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Goldstein & Morris	12.6/G/S	Aug 74	Science, 188, 1211	6
Campbell et al.	12.6/A/S	1975	Science, 196, 650	
Campbell et al.	12.6/A/S	1976	Icarus, 34, 254	
Ostro et al.	12.6/A/S	11/77-2/79	Icarus, 44, 431	

EARTH-BASED RADAR OBSERVATIONS - Saturn's Rings

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Goldstein & Morris	12.6/G/S		Icarus, 20, 260	6
Goldstein et al.	3.5,12.6/A,G/S		Icarus, 30, 104	6
Ostro et al.	12.6/A/S	1977-79	Icarus, 41, 381	

Data Formats: Data are exclusively spectra -- power versus frequency. some show detection only; more recent data may resolve hemispheric differences (as on Galilean satellites). Recent Arecibo data sets include depolarized as well as polarized spectra.

EARTH-BASED RADAR OBSERVATIONS - Asteroids

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
1685 Toro:				
Goldstein et al.	12.6/G/S	Aug 72	AJ, 78, 508	?
Ostro et al.	12.6/A/S	Jul 80	AJ, 88, 565	
1566 Icarus:				
Goldstein	12.6/G/S	Jun 68	Science, 162, 903 Icarus, 10, 430	6
Pettengill et al.	3.8/H/S	Jun 68	Icarus, 10, 432	
433 Eros:				
Jurgens & Goldstein	3.5,12.6/G/S	Jan 75	Icarus, 28, 1	4
Campbell et al.	70/A/S	Jan 75	Icarus, 28, 17	
1580 Betulia:				
Pettengill et al.	12.6/A/S	May 76	Icarus, 40, 350	
1 Ceres:				
Ostro et al.	12.6/A/S	Mar 77	Icarus, 40, 355	
Vesta				
Ostro et al.	12.6/A/S	6 Nov 77	Icarus, 43, 169	

Data Formats: Data are exclusively spectra -- power versus frequency. Some show detection only. Recent data sets may include depolarized as well as polarized spectra.

EARTH-BASED RADAR OBSERVATIONS - Comets

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Encke				
Kamoun et al.	12.6/A/S	Nov. 80	Science, 216, 293	
Grigg-Skjellerup				
?	?/A/S	1982		
IRAS-Araki-Alcock:				
Goldstein et al.	3.6, 12.6/G/S	1983	Bull AAS, 15, 800	4
Campbell et al.	12.6/A/S	1983	Bull AAS, 15, 800	
Saguna-Saigusa-Fujikawa				
Campbell et al.	12.6/A/S	1983	Bull AAS, 15, 800	

Data Formats: Data are exclusively spectra -- power versus frequency. Some show detection only; ore recent data may resolve hemispheric differences (as on Galilean satellites). Recent Arecibo data sets include depolarized as well as polarized spectra.

BISTATIC RADAR OBSERVATIONS - Planetary Surfaces

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Moon:				
Tyler & Simpson	220/Ex35-Su/S	1967	Radio Sci, 5, 263	1
Tyler & Howard	220/ Ap 14,15,16-SU,DSN/ S	1971-73	JGR, 78, 4852 IEEE Trans, AP-30, 438	1 3
Yakovlev et al.	32, 170/ Luna 11,12,14-USSR/			
Venus:				
Kolosov et al.	32/Ven 9,10-USSR/S		IEEE Trans, AP-27, 18	
Croft	13/PVp-DSN/S	Dec 78	GRL, 7, 521	ND(?)
Mars:				
Kliore et al.	12.6/M9-DSN	May-June 72	JGR, 78, 4331	
Simpson & Tyler	12.6/Vik-DSN/S	11/77-3/78	Icarus, 46, 361	3
Lindal et al.	3.6,12.6/Vik-DSN	1976-78	JGR, 84, 8443	

Data Formats: Most reduced data are in the form of spectra -- power vs frequency. Analyzed data which result give scattering properties (dielectric constant and roughness) of the surface. Kolosov et al. have also estimated elevations. Data of Tyler and Howard (in IEEE Trans. paper) are surface tild probability density functions inferred from spectra. Kliore et al. and Lindal et al. have used occultation techniques to determine elevations.

BISTATIC RADAR OBSERVATIONS - Atmospheres, Ionospheres, and Rings
Inner Planets

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Mercury:				
Howard	/M10-DSN/		Science, 185	
Venus:				
Eshleman	/M5-SU/		Science, 158, 1678	NSSDC
Howard	/M10-DSN/		Science, 183	NSSDC
Kliore	3.6,12.6/PV-DSN/	12/78-2/79	JGR, 85, 7957 Icarus, 52, 320	NSSDC
Woo	/PV-DSN/		JGR, 85, 8031	NSSDC
Kolosov et al.	32/Ven 9,10-USSR/S		IEEE Trans, AP-27, 18	
Mars:				
Kliore et al.	/M4/		Science (9/10/65)	
Kliore	13/M6-DSN/			NSSDC
Kliore	13/M7-DSN/			NSSDC
Kliore et al.	/M9-DSN/	May-Jun 72	JGR, 78, 4331	
Lindal et al.	3.6,13/Vik-DSN/	1976-78	JGR, 84, 8443	

Data Formats: Raw data are usually periodic samples of the received waveform. Reduced data typically are retained as spectra -- power versus frequency. These are used to produce temperature-pressure profiles of atmospheres. Statistics of the power spectra are used to infer turbulence parameters of atmospheres and/or ionospheres. Differential phase measurements in the case of two-frequency experiments may be used to infer electron content of plasmas.

BISTATIC RADAR OBSERVATIONS - Atmospheres, Ionospheres, and Rings
Outer Planets

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Jupiter:				
Kliore	/P10,11-DSN/		Science, 183, 323	
Lindal et al.	3.6,12.6/Voyl,2-DSN/		JGR, 86, 8721	
Io:				
Kliore	/P10-DSN/			NSSDC
Saturn:				
Tyler et al.	3.6,12.6/Voy1,2-DSN/	11/80-8/81	Science, 215, 553 Icarus, 54, 160	NSSDC
Titan:				
Lindal et al.	3.6,12.6/Voyl-DSN/	12 Nov 80	Icarus, 53, 348	

Data Formats: Raw data are usually periodic samples of the received waveform. Reduced data typically are retained as spectra -- power versus frequency. These are used to produce temperature-pressure profiles of atmospheres or opacity profiles of rings. Statistics of the power spectra are used to infer turbulence parameters of atmospheres and/or ionospheres.

SPACECRAFT RADAR

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
<u>Moon:</u>				
Kaula	*/Ap15,16,17/			NSSDC
Peeples *et al.*	/AP 17/			NSSDC
Kroupenio	3/Luna 16,17/		COSPAR XV	
<u>Venus:</u>				
Pettengill *et al.*	17/PV/	1978-81	*JGR*, *85*, 8261	NSSDC
<u>Mars:</u>				
Michael	/VikL/	1976		NSSDC

* Apollo laser altimeter.

Data Formats: Data formats within this classification are varied. Pioneer Venus radar data report elevation, reflectivity, and roughness vs position on Venus' surface. Apollo instruments presumably give range to points along the sub-spacecraft track; the radio sounder data are more complex (see Peeples et al.) Viking Lander data are from engineering telemetry.

SPACECRAFT RADIOMETRY

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
<u>Venus:</u>				
Ford & Pettengill	17/PV/	1978-81	<u>Science, 220,</u> 1379	
<u>Mars:</u>				
Kroupenio	3.4/Mars 3,5/	1971, 1974		
<u>Outer Planets:</u>				
Warwick	700+/Voy 1,2/		JGR, 86, 8529+ <u>Science, 215,</u> 582	NSSDC

Data Formats: Pioneer Venus data have been mosaicked to give temperature vs position on Venus' surface. The Voyager data are radio receiver power as a function of time in a large number of frequency bands.

RADIO AND RADAR OBSERVATIONS - Solar Wind

Observer(s)	Wavelength(cm) Observatory Data Product	Date(s)	Reference(s)	Data Status
Coles			Space Sci Rev, 21, 411	
Coles	407/UCSD/			
Coles et al.	/EISCAT/			
Coles + Bourgois	18,21/Nancay/			
Tyler et al.	/M10,Vik-DSN/		Ap J, 249, 318	
Harmon et al.	/A/	1979, 1981	Ap J, 270, 748	

Data Formats: Data retained are generally spectra, showing scintillation of radio sources. Data of Coles are scintillations on natural radio sources, data of Tyler are scintillations on spacecraft transmissions, and data of Harmon are scintillations on earth-based radar echoes from Venus.

EARTH-BASED RADIO OBSERVATIONS - Planets
Terrestrial Planets

Observer(s)	Wavelength Observatory Data Product	Date(s)	Subject(s)	Data Status
Mercury:				
Venus:				
Janssen et al.	1.3,2/VLA/		"weather"	
Allen et al.	mm/KP/		S, C1	
Muhleman + Clancy	mm/KP		CO	
Good + Schloerb	0.3/HC		SO2	
Schloerb + Good	mm/FC,Bell		CO	
Willson			CO	
Mars:				
Muhleman + Clancy	mm/KP		CO	

Data Formats: Data can be continuum observations or specta; either of those types may be in mapped or non-mapped format. Attributes of these data sets are not known.

EARTH-BASED RADIO OBSERVATIONS - Planets
Outer Planets

Observer(s)	Wavelength Observatory Data Product	Date(s)	Reference(s)	Data Status
Jupiter:				
DePater et al.	1.3,2,6,20/VLA/			
Douglas et al.	decametric/UT/			
DePater et al.	11/VLA		Icarus, fall '82	
Galilean Satellites:				
DePater et al.	1.3,2,6,20/VLA/		*	
Berge et al.	2,6/VLA			
Saturn:				
DePater et al.	1.3,2,6,20/VLA		*	
Pettengill + Chapman	20/VLA			
Romis et al.	20/VLA			
Titan:				
Muhleman				
Caldwell + Jaffe			Ap. J., (1980-82?)	
Uranus:				
DePater et al.	1.3,2,6,20/VLA		*	
Caldwell et al.	2,6/VLA			
Neptune:				
DePater et al.	1.3,2,6,20/VLA		*	
Pluto:				
Kellerman et al.	6/VLA			

* DePater Ph.D. thesis and several A+A articles.

Data Formats: Data may be either continuum observations or spectra; either of these types may be displayed in mapped or unmapped format. Attributes of these data sets are not known.

EARTH-BASED RADIO OBSERVATIONS - Asteroids

Observer(s)	Wavelength Observatory Data Product	Date(s)	Reference(s)	Data Status
Johnston *et al.*	2,6/VLA			
Wade *et al.*	2,6/VLA			
Webster *et al.*	2/VLA			

Data Formats: Unknown

EARTH-BASED RADIO OBSERVATIONS - Comets

Observer(s)	Wavelength Observatory Data Product	Date(s)	Reference(s)	Data Status
Kohler:				
Crovisier et al.	1.35/ /		Astron. Astrophy., 97, 195	
Meier:				
Crovisier et al.	1.35/ /		Astron. Astrophy., 97, 195	
Austin:				
Palmer et al.	6,18/VLA/S			
DePater + Ip	2,6,20/VLA		Bull AAS, 15, 805	
Encke:				
Giguere et al.	18/A			
Drake et al.	28/A			
Bradfield:				
Ekelund et al.			Icarus, 47, 431	
Kohoutek:				
Maran et al.			NASA SP-355, 185	
Hobbs et al.			Ap J, 201, 749	
Akabane + Chikada	0.41/ /		Pub Astr Soc Japan, 27, 101	
Bruston et al.	0.14/ /		Nature, 252, 665	
West:				
Hobbs et al.	3.7/ /		Ap J, 218, 573	

Data Formats: Generally spectra -- power versus frequency.

CELESTIAL MECHANICS

Observer(s)	Wavelength Observatory Data Product	Date(s)	Target	Data Status
Anderson	/M2/T		Venus	
Anderson	/M4/T		Mars	NSSDC
Anderson	/M5/T		Venus	NSSDC
Anderson	/M6/T		Mars	NSSDC
Anderson	/M7/T		Mars	NSSDC
Lovell & Shapiro	/M9/T		Mars	
Howard *et al.*	/M10/T		Mars & Venus	
Michael *et al.*	/Vik L/T		Mars	
Anderson	/Voy/T		Saturn	
Shapiro	/PV/T		Venus	NSSDC

* Many radio and radar sets have also been used for celestial mechanics. Randing to Venus is used to develop ephemerides, for example, while Doppler broadening of Mercury and Venus echoes has been used to determine their rotation rates.

Data Formats: Sometimes raw ranging data from spacecraft tracking systems. Sometimes range and/or Doppler residuals.

SUPPLEMENT B2

Planetary Observations from Ground Observatories

The following is a partial copy of the "List of Radio and Radar Astronomy Observatories" published by the National Academy of Sciences and the National Academy of Engineering in March 1983. It has been annotated to indicate 1) whether planetary observations have been made at each facility, 2) where those data might reside, and 3) who might know about them. Where no annotation has been made, we have no information.

Codes are as follows:

1) Has significant planetary work been done at this facility?

   Y = Yes; currently or recently (e.g., past 12 months)
   P = yes, but not recently
   N = never

2) Would data (either in raw or processed form) have been saved?

   O = probably at observatory
   I = probably by investigator
   N = probably not

3) Who would be a good person to contact for specific information about these data?

APPENDIX C

JPL PLANETARY RADAR FACILITY A BRIEF REPORT

The JPL planetary radar facility has acquired new computing equipment and facilities for image display during the past year. We have planned to support some limited on-line image retrieval system, documentation files, and data calibration files. The system hardware consists of a VAX-780 configured with the following:

- 2 6250 BPI tape drives
- 1 800-1600 BPI tape drive
- 1 7 Track 800 BPI tape drive
- 2 25MB cartridge disc
- 1 600 mb hard disc
- 1 writeable control store with double precision hardware
- 1 AED 512 color graphics display
- 4 modem lines
- 4 resident terminals

We have explored the possibility of using the Washington U. BIRP system for cataloging and displaying radar images, however, this may be overkill for our limited data set. We have also explored using DECNET as a networking system, but are not convinced that full network capability is needed, i.e., we can support several remote users directly and are currently doing so.

The JPL radar data set consists of both imaging and non-imaging types of data. Our final image products are map frames containing atleast 250 k pixels of 6 or 8 bits each. All of these images were archived on seven track tape in IPL format. The calibration sites and other non imaging data were all preserved as binary data on seven track tape. Thus, much of our effort to maintain this data set has been directed at tape conversion. We can currently convert all of our tapes for which the original data format is known. Documentation for some formats may not be known unless copies of the data reduction programs are available. Such formats are not available for Venus intermediate data tapes from 1972 through 1975.

Converted images may be displayed on an AED 512 graphics system. We have two software packages that we have developed for this purpose. The first uses only the RS 232 interface to the AED 512, thus an outside user can display images using this program with a 1200 band transfer rate. It's slow, but it works. We have also experimented with the transmission of 6 bit pixels as a ASCII characters over modem lines. Such files can be transferred to the user for display on other systems.

The second program uses the fast parallel interface. The screen can be refreshed in a few seconds with this program. When images are larger than 512 x 512, software exists to scroll through the image using the joystick.

There are a number of limitations with respect to the data that can be kept on line. Since each radar usage occupies atleast a quarter of a megabyte of disc, no more than thirty such images will normally be available at one time. Our normal data processing activity normally uses 500 to 600 megabytes of disc storage, thus the images may be removed from time to time.

Calibration data is difficult to make available in that it requires a dedicated effort to locate the original tapes, reprocess them, and create the data log files. We will attempt to do this for some of the most interesting data sets.

We plan to put the entire Mars data set on-line (non-image type data). Since this is a fully calibrated set, the user should not normally require other calibration data.

Finally, the problem of an adequate catalog is still open. We would like to see some standard format adopted before we invest much effort in this activity. Our initial attempts at this will be in the form of descriptive documentation. Since most of our effort is currently devoted to rebuilding the radar system and rewriting the data processing software, little resources will be available for this activity this year.

APPENDIX D

Institution	Computer(s)*	Programming Language(s)*	Network(s) Available#
Observatories:			
Arecibo	Harris/800	Fortran	
	Harris/6	C	
FCRAO			
Goldstone			
Haystack			
NRAO-Kitt Peak	PDP 11	FORTH	
Owens Valley			
VLA	DEC 10	Fortran	
	VAX 11/780s	AIPS	
Bell Labs			
Clark Lake			
Max Planck			
McDonald			
U Tx RAO			
NRAO - Green Bank	Modcomps	Fortran	
	VAXs	AIP	
Other:			
Universities:			
Arizona			
Arizona State	{PDP 11/45		
	MINIVICAR}		
Brown			
Cal Tech	VAX 11/780		
Colorado	VAX 11/780		
Cornell			
Illinois	VAX 11/780	Fortran	
		AIPS	
Iowa			
Massachusetts			
MIT	{IBM 4321	Fortran	Bitnet
	IBM 370	PL/1	
	VAX 11/780}	C	
Murray State			
Pittsburgh	Dec 10		
Stanford	Eclipse S-250	Fortran	None
	VAX 11/782	Fortran	Ethernet
			Telenet

SUNY	UNIVAC	
Texas		
Wisconsin	VAX	Fortran
		AIPS
UCLA	IBM3033	
UCSD		
Other:		
Other Institutions		
Ames	7600	
Dwingeloo/ Groningen	7600	AIPS
		GIPSY
Goddard	Vax 11/780	
Haw. Inst. Geophys.	{VAX 11/750	
	TI 980B	
JPL	{UNIVAC 1100	
	{PRIME 550	
LPI		
PSI - Pasadena		
PSI - Tucson		
SAO - Cambridge		
USGS - Flagstaff		
USGS - Menlo Park		

* used for planetary work
# linking indicated computer with others

1 Report No	2 Government Accession No	3 Recipient's Catalog No
NASA CP-2343 (Part 2)		

4 Title and Subtitle	5 Report Date
Planetary Data Workshop	October 1984
	6 Performing Organization Code EL
7 Author(s) Dr. Hugh H. Kieffer, Chairman	8 Performing Organization Report No
	10 Work Unit No
9 Performing Organization Name and Address Office of Space Science and Applications Solar System Exploration Division	11 Contract or Grant No
	13 Type of Report and Period Covered Conference Publication
12 Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546	14 Sponsoring Agency Code

15 Supplementary Notes

16 Abstract

This is the proceedings of the Planetary Data Workshop, held November 29-30 through December 1, 1983, at Goddard Space Flight Center, Greenbelt, Maryland. The community of planetary scientists addresses two general problems regarding planetary science data: (1) important data sets are being permanently lost, and (2) utilization is constrained by difficulties in locating and accessing science data and supporting information necessary for its use. These proceedings, in two parts, explore a means to correct the problems, provide science and functional requirements for a systematic and phased approach, and suggest technologies and standards appropriate to the solution.

17 Key Words (Suggested by Author(s))	18 Distribution Statement
planetary science data science data management data base management discipline centers	UNCLASSIFIED - UNLIMITED primary distribution to conference participants; secondary distribution by request Subject Category 91

19 Security Classif (of this report)	20 Security Classif (of this page)	21 No of Pages	22 Price*
unclassified	unclassified	124 (Part 2)	A06

*For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1984

National Aeronautics and
Space Administration

Washington, D.C.
20546

Official Business

Penalty for Private Use, $300

THIRD-CLASS BULK RATE

Postage and Fees Paid
National Aeronautics and
Space Administration
NASA-451

U.S.MAIL

NASA

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return