NASA Technical Memorandum 86250

The Pilot Land Data System: Report of the Program Planning Workshops

Pilot Land Data System Working Group

JULY 1984

NASA

NASA Technical Memorandum 86250

The Pilot Land Data System: Report of the Program Planning Workshops

Pilot Land Data System Working Group
NASA Office of Space Science and Applications
Washington, D.C.

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1984

THE PILOT LAND DATA SYSTEM

REPORT OF THE PROGRAM PLANNING WORKSHOPS

PREFACE

Under the sponsorship of the National Aeronautics and Space Administration's (NASA) Office of Space Science and Applications' (OSSA) Information Systems Office (ISO), the Universities Space Research Association (USRA) assembled a Working Group and coordinated a series of planning workshops during fall 1983 and winter 1984 to draft an advisory report which NASA will use in developing a program plan for a Pilot Land Data System (PLDS). The purpose of the PLDS is to improve the ability of NASA and NASA-sponsored researchers to conduct land-related research. The goal of the planning workshops was to provide and coordinate planning and concept development between the land-related science and computer science disciplines, to discuss the architecture of the PLDS, requirements for information science technology, and system evaluation. This report presents the findings and recommendations of the Working Group.

The goal of the pilot program is to establish a limited-scale distributed information system to explore scientific, technical, and management approaches to satisfying the needs of the Land Science community. The PLDS would pave the way for a Land Data System to improve data access, processing, transfer, and analysis, thus fostering an environment in which land sciences information synthesis can occur on a scale not previously permitted because of limits to data assembly and access.

This document was prepared by the Working Group for Pilot Land Data System Planning, composed of the following individuals:

Science Working Group (SWG)

Mr. Ted Albert, U.S. Geological Survey

Dr. Glenn Bacon, IBM Corporation

Dr. Daniel Botkin, University of California at Santa Barbara

Dr. John Estes, University of California at Santa Barbara
(PLDS Planning Workshop Project Director)

Ms. Janet Franklin, Universities Space Research Association
(PLDS Planning Workshop Coordinator)

Dr. Roger Holmes, Allegheny International Inc.

Dr. Edward Kanemasu, Kansas State University

Dr. Robert Ragan, University of Maryland

Dr. Robert Singer, University of Hawaii

Dr. Jeffrey Star, University of California at Santa Barbara

Dr. Sylvan Wittwer, Michigan State University

Centers Working Group (CWG)

NASA HQ

Mr. Michael Devirian, Information Systems Office, OSSA (PLDS Program Manager)

Mr. Alexander Tuyahov, Earth Observations Division, OSSA

NASA/Ames Research Center (ARC)

Mr. James Brass
Mr. David Morse
Ms. Susan Norman

NASA/Goddard Space Flight Center (GSFC):

Dr. Philip Cressy
Dr. Robert Price
Dr. Paul H. Smith

Jet Propulsion Laboratory (JPL):

Mr. Fred Billingsley
Dr. Tom Farr
Mr. Jose Urena (Global Resources Information System representative)

NASA/Johnson Space Center (JSC):

Dr. Jon Erickson
Mr. Robert Musgrove

NASA/National Space Technology Laboratories (NSTL):

Dr. Kenneth Langran
Mr. Ronnie Pearson

Additional Working Group members and contributors to the report included:

Mr. William Campbell, GSFC
Mr. William Erickson, ARC
Dr. Alex Goetz, JPL
Mr. Michael Goldberg, GSFC
Dr. Ken Green, SAR/GSFC
Dr. Harry Jones, ARC
Mr. William Likens, ARC
Dr. Dale Lumb, ARC
Ms. Elizabeth Middleton, GSFC
Mr. Wayne Mooneyhan, NSTL

Ms. Leslie Morrissey, TGS/ARC
Dr. Robert Murphy, NASA HQ
Dr. H.K. Ramapriyan, GSFC
Mr. Charles J. Robinov, USGS
Mr. Larry Roelofs, CTA/GSFC
Dr. Howard Stauffer, ARC
Ms. Valerie Thomas, GSFC
Mr. Sid Whitley, NSTL
Dr. Kerry Woods, UCSB

We would like to acknowledge the following people for their comments, contributions, and help in preparing the report:

Dr. Ray Arvidson, Washington University
Dr. Howard Hogg, NASA HQ
Dr. Chris Johannsen, University of Missouri, Columbia
Ms. Herminia Manapasal, USRA
Dr. Mitchell Rambler, NASA HQ
Dr. Mark Settle, NASA HQ
Ms. Donna Smith, RMS/GSFC
Dr. James Vette, GSFC

EXECUTIVE SUMMARY

There is a trend in scientific research in general, and more specifically in National Aeronautics and Space Administration programs to ask research questions which are multidisciplinary in nature and global in scale. Researchers at agencies and institutions across the nation and around the globe are attempting to improve our understanding of global carbon cycling; the relationship between land energy balance and biophysical conditions, and their interrelationship with climate; global and regional geologic and geomorphic structure and process; and, to identify early indications of change in global elemental cycles, climate, hydrology and environmental quality.

Satellite remote sensing offers the land science community interested in such questions a unique and essential tool. These systems can provide data of a type and on a scale previously unattainable. Yet, looking forward to the capabilities of Space Station and the Earth Observing System (EOS), full realization of the potential of satellite remote sensing has consistently been handicapped by inadequate information systems. This must not be allowed to continue. Recent studies and activities, and the experience of the participants in Planning Workshops, suggest that the full potential of remote sensing will not be achieved without expanded efforts to effectively integrate remote sensing and information science technologies. Such an approach must not stop at the ground receiving station, but must fully integrate all aspects of the information systems needs of NASA and NASA-sponsored researchers.

Under the sponsorship of the NASA Information Systems Office (ISO), the Universities Space Research Association (USRA) assembled a Working Group and coordinated a series of workshops to examine the need for and basic characteristics of a project which would incorporate the latest technological advances in information systems in a coordinated attack upon the needs of a broad range of land scientists: a Pilot Land Data System (PLDS). The overall task of program definition was carried out by the PLDS Working Group, which included science discipline, information systems, and management personnel.

Based on the work conducted under the planning activity, members of the PLDS Working Group conclude that:

- There is a need to improve the ability of NASA and NASA-sponsored scientists to locate, access, process and analyze remotely sensed and other land science data;

- rates, volumes, and types of remotely sensed data severely tax current data and information systems; and

- unless the ability to handle these and other land science data is established now, effective use of data from future systems (e.g., Space Station) will be severely impacted.

Further, the Working Group recognized that:

- o Land scientists under NASA sponsorship have complex, high volume, multidisciplinary information requirements;
- o Satisfying these requirements will enable researchers to better address important, multi-disciplinary science questions, and can lead to improved understanding of many land processes;
- o The technology exists to enable land science research to function more effectively.

The goal of the pilot program discussed herein to establish a limited-scale distributed information system to explore scientific, technical and management approaches to satisfying the needs of the land science community. The PLDS will improve scientific data access, processing, transfer, and analysis. The PLDS will also foster an environment in which land sciences information synthesis can occur on a scale not previously permitted. The proposed PLDS can be viewed as a means of increasing scientific productivity through a more effective use of information science technology.

The development of a PLDS represents a challenge due to the number and size of relevant data acquisition, networking processing, and analysis systems, and, the need to interconnect scientists in a number of institutions across the country who are currently employing a variety of hardware and software systems. Experience in many science disciplines has shown that effective and efficient use of data must be based on a solid data system foundation. PLDS must be and a network structure implemented in such away that enhances science potential and fosters cooperation with other agencies and research institutions.

Experience of workshop participants in developing information systems as tools for science, and the dynamic nature of advances in computer science and information systems technologies, lead to the adoption of a number of general guidelines for PLDS planning and implementation. These guidelines include:

- o PLDS will be designed specifically to serve the data and information systems needs of NASA, NASA-sponsored, and NASA-related scientists working on land science research projects.
- o PLDS would be a research and "proof-of-concept" tool. When fully implemented, it would support NASA land research programs, and validate system attributes in support of the global-scale land science research community. The PLDS will be expected to form the basis of a full scale land data system.
- o System goals must be defined jointly by both the land and information science communities in terms of major earth science issues to be resolved, and feasible tools with which to resolve them. Progress, and indeed the goals

themselves, must be re-examined regularly in light of NASA program evolution and new achievements in science and technology.

- PLDS will remain most viable when systems operations and management involves researchers with a long term commitment to the use of the data and to sharing their data with others for the purpose of conducting science research (CODMAC 1982).

- Pilot system developmemt should proceed through integration and testing of available, well-understood ("low-risk") technology, exploiting system components in place at participating institutions, whenever possible. Use of science scenarios based on on-going research to drive system planning and implementation is appropriate.

Based upon the experience gained by personnel in PLDS planning activities and in accordance with the guidelines seen above, the Working Group strongly recommends that:

- A Pilot Land Data System be implemented beginning in fiscal year 1985 to link NASA and NASA-sponsored land researchers;

- The initial system be a limited-scale, modular, distributed information system;

- The system have the strong continuing and cooperative involvement of NASA and NASA-sponsored land and information scientists;

- An advisory committee of land and information scientists be constituted to review PLDS progress, to provide advice and guidance and to periodically report to NASA Headquarters Information Systems and Earth Science and Applications management on PLDS progress; and finally,

- Information Systems Office and Earth Science and Applications Division personnel closely coordinate to insure that the land science scenarios chosen to drive the PLDS design are as representative as possible of the range of data and information systems requirements of the land resources community.

TABLE OF CONTENTS

	Page
Preface	iii
Executive Summary	vi
Acronyms	xvi
1. Introduction	1
2. Science Issues and Drivers	4
2.0 Background	4
2.1 Scientific Research Goals	4
2.2 Approach	5
2.3 Success Criteria	6
3. Science Project Drivers	7
3.0 Background	7
3.1 Science Scenario 1. Vegetation Biomass and Productivity, and Large Area Inventory	8
3.2 Science Scenario 2. Biogeochemical Cycling in Forests	8
3.3 Science Scenario 3. Land Surface Climatology	9
3.4 Science Scenario 4. Hydrological Modeling and Soil Erosion/Productivity Modeling	10
3.5 Science Scenario 5. Multispectral Analysis of Sedimentary Basins	11
3.6 Science Scenario 6. Monitoring Environmental Change	12
3.7 Summary of Science Needs	13
4. Land Data System Overview	19
4.0 Introduction	19
4.1 LDS Characteristics	19

4.2 LDS Support of Users 20
4.3 LDS Concept 22
4.4 An Intelligent LDS 24
4.5 PLDS to LDS Transition 26
5. Pilot Land Data System Development 27
5.0 Introduction 27
5.1 FY 85-FY 87 Goals 27
5.2 PLDS Start-up Phase 28
5.3 PLDS Evaluation 30
5.4 Long-Term Goals 30
6. Conclusions and Recommendations 33
6.1 Conclusions 33
6.2 Recommendations 34
APPENDICES I.0.1
I: SCIENCE SCENARIO APPENDICES I.0.2
I.0 Introduction and Overview of Science Scenario Appendices I.0.2
I.1 Science Scenario 1: Vegetation Biomass and Productivity, and Large Area Inventory I.1.1
I.2 Science Scenario 2: Biogeochemical Cycling in Forests I.2.1
I.3 Science Scenario 3: Land Surface Climatology I.3.1
I.4 Science Scenario 4: Hydrological Modeling and Soil Erosion/Productivity Modeling I.4.1
I.5 Science Scenario 5: Multispectral Analysis of Sedimentary Basins I.5.1
I.6 Science Scenario 6: Monitoring Environmental Change I.6.1
II: TECHNICAL APPENDICES II.0.1
II.0 Introduction to Technical Appendices II.0-1
II.1 DATA MANAGEMENT II.1-1

II.1.0 General Data Management Concepts and Considerations II.1-1

II.1.1 Finding Data II.1-4

II.1.2 Accessing Data II.1-4

II.1.2.1 General Data Access Considerations II.1-5

II.1.2.2 On-line Directory and Catalog Design II.1-5

II.1.2.3 Data Storage and Compression II.1-6

II.1.3 Building Versatile Data Bases II.1-7

II.1.3.1 General Data Base Considerations II.1-7

II.1.3.1.1 Geocoded Data Structures II.1-7

II.1.3.1.2 Data Base Design and Maintenance II.1-8

II.1.3.1.3 Data Validation II.1-9

II.1.3.2 Advanced Database Management System Development II.1-9

II.1.3.3 Advanced Hardware Technology II.1-11

II.2 COMMUNICATIONS II.2-1

II.2.1 Introduction II.2-1

II.2.2 Structure and Regulations Regarding NASA Communications II.2-1

II.2.3 PLDS Functions Requiring Communictions II.2-2

II.2.4 Classes of Data Traffic II.2-3

II.2.5 Communication Technologies for the PLDS II.2-3

II.2.5.1 Data Transport II.2-3

II.2.5.2 Communication Protocols and Control II.2-7

II.2.5.3 Interconnecting LANs AND LDNs II.2-8

II.2.6 Implementation Approaches for PLDS Communications II.2-8

II.2.7	Network Control and Administration	II.2-9
	II.2.7.1 Operations Management	II.2-10
	II.2.7.2 Accounting and Billing	II.2-10
	II.2.7.3 Subscriber Services	II.2-10
II.3	DATA INPUT/OUTPUT AND INTERFACE SUBSYSTEMS	II.3-1
II.3.1	Introduction	II.3-1
II.3.2	Standard Formats for Digital Data	II.3-1
	II.3.2.2 Interchange Format Structure Basis	II.3-2
	II.3.2.3 Mission-independent Formatting	II.3-2
	II.3.2.4 Mission-dependent Formatting	II.3-3
	II.3.2.5 Layered, Onion-skin, Formats	II.3-3
	II.3.2.6 Standard Format Data Unit	II.3-4
II.3.3	Standard Formats for Other Data Types	II.3-4
	II.3.3.1 Vector Data	II.3-4
	II.3.3.2 Catalog Data	II.3-5
	II.3.3.3 Tabular Data	II.3-5
II.3.4	Data Input Types and Reformatting	II.3-5
	II.3.4.1 Existing Digital Data	II.3-5
	II.3.4.2 Map Data	II.3-5
	II.3.4.3 Field and Sample Data	II.3-6
II.3.5	Editing and Quality Checking	II.3-6
	II.3.5.1 Raster Data	II.3-6
	II.3.5.2 Vector Data	II.3-6
	II.3.5.3 Vector-to-Raster Conversion	II.3-6
II.4	USER INTERFACE	II.4-1
II.4.1	Introduction	II.4-1
II.4.2	Design Considerations	II.4-2

II.4.3 Hardware Requirements, Guidelines, and Standards II.4-2

II.4.4 Software Requirements, Guidelines, and Standards II.4-3

II.4.5 User Levels and System Configurations II.4-4

II.4.6 Technological Developments Required II.4-9

II.4.7 Implementation Strategy II.4-10

II.5 DATA ANALYSIS AND MANIPULATION II.5-1

II.5.1 Introduction II.5-1

II.5.2 Rectification and Registration II.5-1

II.5.2.1 Geocoding Considerations II.5-2

II.5.2.2 Performance Requirements II.5-2

II.5.2.3 The Registration Process II.5-3

II.5.2.4 Development Issues II.5-4

II.5.3 Data Manipulation and Software Sharing II.5-4

II.5.3.1 Software Exchange and Non-local Use Considerations II.5-5

II.5.3.2 Allocation to Various Processors In Network II.5-6

II.5.4 Geographic Information Systems II.5-6

II.5.5 Intensive Computational Systems II.5-7

II.5.5.1 Attached Array Processors II.5-7

II.5.5.2 Computer Architecture II.5-9

II.5.5.3 Supercomputers II.5-10

II.5.6 Artificial Intelligence and the PLDS System Design II.5-11

II.6 SYSTEM SUPPLIED SUPPORT SERVICES AND ADMINISTRATION II.6-1

II.6.1 Overview II.6-1

II.6.2 General Accounting II.6-1

II.6.3 Security II.6-1

II.6.4 Access Charges II.6-2

II.6.5 Data Charges II.6-2

II.6.5.1 Value Added Charges II.6-2

II.6.6 Electronic Mail II.6-3

II.6.7 Summary II.6-3

REFERENCES REF-1

LIST OF FIGURES

Figure	Title	Page
3.1	Functions of a System to Support Land Science	14
4.1	LDS Functional Overview	21
4.2	LDS Subsystem Overview	23
4.3	LDS Communication and Information Analysis Overview	25
5.1	PLDS Development Phases	31
5.2	Potential Nodes on the PLDS Network	32
I.2.1	Locations of Co-investigators in Intensive Study Sites for Science Scenario 2	I.2-7
I.3.1	Data and Analysis Network for Science Scenario 3	I.3-8
I.4.1	Study Area for Science Scenario 4:Washita River Basin, Oklahoma	I.4-2
I.4.2	Data Network for Science Scenario 4	I.4-7
II.2.1	Evaluation Methodology for Transport System Alternatives	II.2-5
II.4.1	Typical Level I User Configuration	II.4-6
II.4.2	Typical Levels II and III User Configurations	II.4-7
II.4.3	Network Concept	II.4-8

LIST OF TABLES

Table	Title	Page
3.1	Summary of the Functional Requirements of the PLDS Planning Science Scenarios	16
3.2	Ordered Ranking of Information System Functions Requiring PLDS Support by Science Scenarios	18
I.0.1	Institutions Involved in Science Scenarios	I.0-3
I.0.2	Data Requirements for Science Scenarios	I.0-5
I.1.1	Summary of Data Needs - Science Scenario 1	I.1-6
I.2.1	Sampling Parameters for Biogeochemical Cycling Study - Science Scenario 2 Science Scenario 2	I.2-5
I.2.2	Summary of Data Needs - Science Scenario 2	I.2-6
I.3.1	Summary of Data Needs - Science Scenario 3	I.3-6
II.1.1	DMBSs Currently Used at NASA Centers	II.1-2
II.4.1	Suggested Specifications for User Workstations	II.4-12

ACRONYMS

ACSM	American Congress on Surveying and Mapping
ADABAS	Adaptable Data Base System
AGRISTARS	Agricultural and Resources Inventory Surveys Through Aerospace Remote Sensing
AI	Artificial Intelligence
AID	Agency for International Development
AIS	Airborne Imaging Spectrometer
AOI	Area of Interest
AP	Array Processor
ARC	Ames Research Center
ARPANET	Advanced Research Projects Agency Network
ARS	Agricultural Reporting Service (USDA)
ASAP	Advanced Scientific Array Processor
ASWS	Advanced Scientific Workstation
AVHRR	Advanced Very High Resolution Radiometer
AVIRIS	Airborne Visible and Infrared Imaging Radiometer
b	bits
B	Bytes
BMDP	Biomedical Statistical Software Package, P Series (UCLA Medical Center)
bpi	bits per inch
bps	bits per second
CCSDS	Consultative Committee for Space Data Systems
CCITT	Comite Consultatif Internationale de Telegraphie et Telephonie
CCT	Computer Compatible Tape
CIE	Classified Image Editor (ARC Software)
CODMAC	Committee on Data Management and Computation
COSPAR	Committee on Space Research
CPU	Central Processing Unit
CSC	Computer Science Corporation
CSNET	Computer Science Network
CTA	Computer Technology Associates, Inc.
CUNY	City University of New York
CWG	Centers Working Group
DBMS	Data Base Management System
DEC	Digital Equipment Corporation
DEM	Digital Elevation Model
DMA	Defense Mapping Agency
DDR	Data Definition Record
DMS	Data Management Subsystem
DSTP	Data Systems Technology Program (MSFC)
EDC	EROS Data Center
EDITOR	ERTS Data Interpreter and TENEX Operations Recorder
ELAS	Earth Resources Laboratory Applications Software
EOD	Environmental Observations Division (OSSA)
EPIC	Erosion-Productivity Impact Calculator
EROS	Earth Resources Observations Systems
EWS	Engineering Work Station
ESMR	Electronically Scanned Microwave Radiometer
FFT	Fast Fourier Transform
FIPS	Federal Information Processing Standards
FPS	Floating Point Systems

FTS	Federal Telephone System
FY	Fiscal Year
GEOSAT	The Geological Satellite Committee
GHZ	Gigahertz
GIPSY	General Image Processing System (Virginia Polytechnic Institute and State University, Blacksberg)
GIS	Geographic Information System
GOES	Geostationary Operational Environmental Satellite
GWS	General Work Station
GSA	General Services Administration
GSFC	Goddard Space Flight Center
HCMM	Heat Capacity Mapping Mission
HIRS	High Resolution Infrared Sounder (NOAA-7, 8)
IAMAP	International Association of Meteorology and Atmospheric Physics
IBIS	Image Based Information System
IBM	International Business Machine Inc.
ICPS	Intensive Computational Processing System
IDIMS	Interactive Digital Image Manipulation System
IMP	Interface Message Processors
IMSL	International Mathematics and Statistical Library (Fortran Subroutines)
INORAC	Inquiry, Order and Accounting System
I/O	Input/Output
IPS	Image Processing System
IPWS	Image Processing Workstation
ISLSCP	International Satellite Land Surface Climatology Program
ISO	International Standards Organization
ISO	Information Systems Office (NASA, OSSA)
ISS	Information Sciences System
I^2S	International Imaging Systems
JASIN	Joint Air Sea Interaction Experiment (PODS)
JPL	Jet Propulsion Laboratory
JSC	Johnson Space Center (NASA)
LAN	Local Area Network
Landaat	Land Satellite
LARS	Laboratory for Applications of Remote Sensing (Purdue University)
LARSYS	LARS System Software
LDN	Long Distance Network
LIDAR	Light Detection and Ranging
LISP	List Processing
MetSat	Meteorological Satellite
MFLOPS	Million Floating Point Operations per Second
MIDAS	Microcomputer-based Data Analysis System
MIMD	Multiple Instructions Multiple Data
MIPL	Multi-mission Image Processing Laboratory (JPL)
MIPS	Million Instructions per Second
MLA	MultiLinear Array (Shuttle Experiment)
MPP	Massively Parallel Processor
MSFC	Marshall Space Flight Center
MSS	MultiSpectral Scanner (Landsat)
MSU	Microwave Sounding Unit (NOAA-7, 8)
NAS	Numerical Aerodynamic Simulator (ARC)

NASA	National Aeronautics Space Administration
NASCOM	NASA Operational Communications
NASPN	Numerical Aerodynamic Simulator Processing System Network
NCIC	National Cartographic Information Center (USGS)
NHSCF	NASA High Speed Computing Facility (GSFC)
NOAA	National Oceanic and Atmospheric Administration
NPS	National Park Service (USDI)
NSC	Network Systems Corporation
NSTL	National Space Technology Laboratories
NURE	National Uranium Resource Evaluation
NWS	National Weather Service
OAST	Office of Aeronautics and Space Technology (NASA)
ORNL	Oak Ridge National Laboratory
ORSER	Office for Remote Sensing of Earth Resources
OSI	Open Systems Interconnection
OSSA	Office of Space Science and Applications (NASA)
OSTA	Office of Space Tracking and Data Systems (NASA)
PCDS	Pilot Climate Data System
PE	Processing Element
PI	Principal Investigator
Pixel	Picture element
PLDS	Pilot Land Data System
PODS	Pilot Oceans Data System
PPDS	Pilot Planetary Data System
PSC	Program Support Communications
PSCN	Program Support Communications Network
RETMA	Radio, Electronics, and Television Manufacturers Association
RFP	Request for Proposals
RIPS	Remote Image Processing System
RMS	Republic Management Systems
SAR	Science Applications Research, Inc.
SAR	Synthetic Aperture Radar
SAS	Statistical Analysis System
SCPC	Single Channel Per Carrier
SCS	Soil Conservation Service (USGS)
Seasat	Sea Satellite
SLAR	Side-looking Airborne Radar
SIR-A,-B	Shuttle Imaging Radar
SMIRR	Shuttle Multispectral Infrared Radiometer
SMMR	Scanning Multifrequency Microwave Radiometer
SMS	Synchronous Meteorological Satellite
SPOT	Satellite Probatoire d'Observation de la Terre (France)
SPSS	Statistical Package for the Social Sciences
SSAR	Seasat Synthetic Aperture Radar
SUN	Stanford University Network
SWRRB	Simulator for Water Resources in Rural Basins
SWG	Science Working Group
TAE	Transportable Applications Executive
TDMA	Time Division Multiple Access
TGS	Technicolor Government Services
TIMS	Thermal Infrared Multispectral Scanner
TIROS	Television and Infrared Observation Satellite
TM	Thematic Mapper
TMS	Thematic Mapper Simulator

UCLA	University of California at Los Angeles
UCSB	University of California at Santa Barbara
UIS	User Interface Subsystem
UN/FAO	United Nations Food and Agricultural Organization
UNEP	United Nations Environmental Program
USDA	United States Department of Agriculture
USDI	United States Department of Interior
USGS	United States Geological Survey
USRA	Universities Space Research Association
USLE	Universal Soil Loss Equation
VICAR	Video Image Communication and Retrieval
VISSR	Visible and Infrared Spin Scan Radiometer
VLSI	Very Large Scale Integration
WMO	World Meteorological Organization

1 INTRODUCTION

Satellite remote sensing offers tremendous potential for the Earth Sciences. The realization of this scientific potential is currently limited by inadequate information systems. Today, the National Aeronautics and Space Administration (NASA) and NASA-sponsored scientists have varying levels of difficulty accessing, transferring, processing, and analyzing remotely sensed and other scientific data. Advances in data management, networking, and analysis techniques now make it possible to develop a data system to meet the land scientists' most critical expressed information systems needs to archive, locate, transfer, integrate, and manipulate data in the volumes and at the time scales dictated by the increasingly complex nature of their research. Work towards improved data systems for land scientists must truly involve the land sciences community in order to realize the full potential of future missions, especially in a Space Station era.

Under the sponsorship of the Information Systems Office (ISO), the Universities Space Research Association (USRA) assembled a Working Group and coordinated a series of workshops for the purpose of examining the need for and basic characteristics of a Pilot Land Data System (PLDS). The overall task of program definition was carried out by the PLDS Working Group, which included discipline science, information systems, and management personnel.

This report discusses their conclusions, and describes the recommendations for the structure and implementation of a Pilot Land Data System. The system described is a limited scale, modular, distributed information system which can demonstrate the potential of existing information science technology to meet the most immediate needs of the land research community, while providing a sound technical basis for a future, fully operational Land Data System.

The goal of the pilot program is to establish a limited-scale distributed information system to explore scientific, technical, and management approaches to satisfying the needs of the land science research community. The PLDS will pave the way for a Land Data System to improve data access, processing, transfer, and analysis, thus fostering an environment in which land sciences information synthesis can occur on a scale not previously possible because of limits to data assembly and access. Accomplishing this goal will require interactive research and development, with land and information scientists working closely together to understand the needs of the users of scientific data as they seek to move multidisciplinary NASA land science research forward. The success of the PLDS will be measured in the future by its contribution to increased scientific productivity and improved understanding of features and processes of interest to the land sciences research community. As such, the PLDS must provide the land science and applications oriented user with a powerful, friendly, and cost-effective computing environment for conducting land science research. The environment must support the full spectrum of information functions necessary to conduct land science investigations including location, acquisition, processing, and transfer.

It was recognized from the outset of this effort that developing a PLDS represents a particular challenge which arises from the range of science disciplines and investigators involved, the number and size of possible relevant data bases, their spatial and georeferenced nature, and the variety of relevant data acquisition, networking, processing and analysis systems. A major issue has been the need to interconnect institutions and scientists currently employing a variety of hardware and software systems. The PLDS will be a distributed system. These factors make the task of designing and implementing the network configuration particularly complex. Also, due to the multidisciplinary and inter-institutional nature of land sciences research, the PLDS, and the science issues that structure it, must be defined in such a way as to foster cooperation among NASA Centers, and with other agencies and research institutions.

Experience in many science disciplines (e.g., planetary sciences, oceanography, and climatology) of trying to conduct research involving the processing of large data sets from diverse sources, has shown that effective and efficient use of data in all disciplines must be based on a solid data system foundation. The Information Systems Office has initiated several data system programs to support specific science discipline areas, e.g., the Pilot Oceans Data System (PODS), Pilot Planetary Data System (PPDS), and Pilot Climate Data System (PCDS). The Pilot Land Data System is a further step in this effort to meet the data access, processing and analysis needs of the science user community served by NASA, and to evaluate the potential of existing and newly developing technologies to serve those needs. It is the first effort of its kind that so thoroughly coordinates a large and diverse group of information and discipline scientists and management personnel from the outset of the planning process.

The complexity of environmental processes that affect the Earth requires a multidisciplinary approach to the understanding of natural phenomena and their dynamics. The integration of the PLDS and subsequent Land Data System with other discipline data systems can provide the foundation for the development of a multidisciplinary capability (a Global Resources Information System). This could facilitate truly integrated research involving all of the global data sets potentially available from satellite remote sensing and other conventional sources to address global processes involving the land, air and water.

Experience in developing information systems as tools for science, coupled with the dynamic nature of advancements in computer science and information systems technologies, suggested the adoption of a number of guidelines to ensure successful PLDS planning and implementation. The guidelines listed below represent a mix of both basic philosophical and more pragmatic considerations used to guide the work which is reported here. They are:

1. Data bases tend to remain most viable when maintained by an active group of researchers with a long term commitment to the use of the data, and to sharing the data with others for the purpose of conducting

scientific research (CODMAC 1982). The PLDS architecture and selections of data archive sites will be guided by that consideration.

2. The PLDS will be designed specifically to serve the data and information systems needs of NASA, NASA-sponsored, and NASA-related scientists working on land science projects.

3. As a Pilot, the system described here represents a research and "proof-of-concept" tool. When fully implemented, it will support NASA land research programs, and will validate the system attributes needed to support a global-scale land science research community.

4. Long-term system goals must be defined jointly by both land and information science communities in terms of major earth science issues to be resolved, and feasible tools with which to resolve them. Continued progress toward these goals, and indeed the goals themselves, must be reexamined regularly by the joint science community to ensure close coordination with NASA program evolution, and to take advantage of pertinent achievements in science and technology.

5. Testing and validation of the elements of the PLDS must build upon ongoing research programs, and on system components in place at participating institutions. The use of specific science scenarios based on ongoing land science research assures that the PLDS meets scientific research needs.

6. Pilot system development should proceed through integration, testing, and evaluation of available, well-understood ("low-risk") technology. Close coordination with NASA computer science research (OAST) and communications research (OSTA) must provide the mechanism for development of advanced technologies for incorporation in future upgrades of the PLDS.

The PLDS Working Group identified a number of long-range land science objectives whose successful pursuit (let alone accomplishment) requires the development of an advanced, distributed Land Data System (see Section 2). Land scientists in the workshops described six research scenarios drawn from current research programs (see Section 3). Together, these scenarios established well-validated requirements for data system support in the near-term, and led to a clear expression of user requirements to drive overall system design. Long-term system design concepts were developed and refined in response to these requirements (see Section 4), and major PLDS subsystems and their phased implementation were detailed in Section 5. The conclusions and recommendations of the Working Group are given in Section 6. Specific scientific and technical information developed in support of PLDS definition is given in Appendices I and II, respectively.

2. SCIENCE ISSUES AND DRIVERS

2.0 BACKGROUND

The launch of Landsat 1 stimulated a decade of major advances in the science and technology of remote sensing. These, as well as comparable advances in information sciences, have fundamentally changed the nature of land science. Traditionally, field studies in land sciences have been limited to understanding the behavior of one or two variables within a small area. This was due in large part to the inability to obtain, manage, and interpret large volumes of data, which severely restricted the land scientist's ability to understand and explain the role of factors controlling land processes.

Via the Landsat program, the land science community has come to a point where it is possible to conduct research at a scale large enough to examine the interactions of the critical natural processes that define a "real world" system. But, so far, full realization of the scientific potential of satellite remote sensing has been handicapped by inadequate information systems. Knowledge of available data and information is constrained by limited access to archives and a lack of networked data systems. The ability to access and exchange data and algorithms is hindered by a lack of format standards. Access to appropriate computational resources is often lacking at the institutional or laboratory level. Often, scientists are required to devote a significant portion of their time and energy to data acquisition and preparation. Better integration of remote sensing and information technologies offers the potential to overcome these barriers.

2.1 SCIENTIFIC RESEARCH GOALS

A thorough understanding of land and environmental processes, and the effect of human activity on these processes is required, so that accurate predictive models can be developed to allow the human population to better anticipate environmental events rather than simply to react to them. In recognition of the need to improve understanding of large scale land processes, there is a movement in scientific research in general, and specifically in NASA earth science programs, to ask research questions which are multidisciplinary in nature and global in scale (NASA 1983a, NASA 1983b; Gwynne 1982). This section identifies the critical research problems in the environmental and land science areas, the solutions for which are very difficult or impossible without the application of remote sensing and information systems techniques. The resolution of such large-scale science issues calls for the establishment of interdisciplinary research teams and inquiries. It is expected that the science issues will drive the evolution of the information systems and that new advances in information science will, in turn, create a new perspective for looking at critical problems in the land science domain.

There are many land related and environmental problems of both material and global significance with economic, human health, and environmental impacts. The ultimate objective of the scientific community is to correctly understand the factors involved in land processes and provide a sound predictive modeling capability. Some of the most critical scientific goals that have been identified by the PLDS Science Working Group are very similar to those outlined in the NASA Global Habitability Land Related Research Issues Report (NASA 1983a). Additional critical goals identified here are more geologic in nature. These are all long-term goals in land sciences research and will require many years of concentrated work. However, opportunity exists to make significant, short-term progress on many of these goals, through a program of well-organized and well-supported studies. Some of the most important goals in the land science area are to:

1. Establish methods by which a global carbon budget model may be developed and monitored.

2. Assess the effects of acid rain on biological productivity and soil nutrient availability.

3. Detect the presence and amounts of pollutants and/or toxic substances in the soil the atmosphere, and fresh water resources.

4. Establish the relationship between the land energy balance and land biophysical conditions and their interrelationships with climate.

5. Improve the accuracy of models used for short-term prediction of the local availability, movement, and quality of water, including snow and ice.

6. Identify the early indicators of change in global element cycles, climate, hydrology, and environmental quality.

7. Define the areal extent and spatial distribution of current biomass and productivity of the major biomes.

8. Improve the accuracy of models used for the assessment and prediction of land degradation.

9. Improve the science base for nonrenewable resource assessment.

10. Advance the understanding of global and regional geologic and geomorphic structure and process.

11. Develop improved methods for assessing and monitoring geologic hazards.

2.2 APPROACH

Significant time and effort will be needed to accomplish the scientific goals outlined above. Central to the effort must be an efficient computer processing, data transfer and data management system. Such a system, which in the present report will be termed a "Land Data System," will be critical if the science goals are ever to be approached. While it is relatively easy to conceive of the general operation of a Land Data System, there is no current prototype for it. The technology for each element of a Land Data System is understood, but experience in the integrated communications and data handling requirements must be developed before NASA can proceed toward implementing a global scale system. A well-defined pilot data system, serving a relatively small group of scientists at several locations, is a necessary first step.

The PLDS must be developed as an instrument to improve the ability of scientists to do land-related research and its design will be driven by the needs of the science community. As a pilot program, the PLDS cannot meet all possible needs. Therefore, in the planning process a limited number of science scenarios were selected from research projects already existing or proposed within the NASA land science community. These research scenarios, summarized in Chapter 3 and presented in detail in Appendix I, were selected to be representative of the types of projects whose success would be important to the ultimate success of the goals listed in Section 2.1. Through the use of these scenarios, a majority of the generic information system functions and requirements becomes apparent while a narrow focus, appropriate to a pilot system, is retained. In addition, the use of these scenarios helped to illuminate the needs specific to particular science disciplines, important in determining the overall PLDS requirements.

The discipline science members of the Working Group produced a generalized list of functions required for a complete data analysis system to support land science research. These functions are listed in Table 3.1. These were divided into those that are functions of a communications and data network and those that are a part of the analysis process. Functions were then ranked according to those which required support from a PLDS by the largest number of scenarios. The ordered ranking is shown in Table 3.2. A summary of science needs is provided in Section 3.7.

2.3 SUCCESS CRITERIA

In a pilot study such as the PLDS, there must be periodic benchmarks by which progress can be measured. The key driver behind the PLDS is that none of the scenarios presented in Chapter 3 and Appendix I can be completed at an optimum scale without the proposed data interpretation, management and networking systems. Beyond that foundation, there must be a mechanism to evaluate the degree of success and rate of progress so that adjustments can be made during the program.

Technical measures (e.g., data transmission rates and volumes) can be used to evaluate the success of some aspects of the system on a quantitative basis. While scientific benchmarks may not lend themselves to the same degree of quantification, periodic evaluations (e.g., peer review) can be employed to judge pilot program progress in meeting science objectives.

3. SCIENCE PROJECT DRIVERS

3.0 BACKGROUND

There is a trend in scientific research in general, and specifically in NASA research programs, to ask research questions which are multidisciplinary in nature and global in scale. Research projects discussed below reflect this trend. Many of these projects emphasize the interaction of the land surface with the atmosphere and oceans through the hydrologic cycle, climate, and biogeochemical cycling. Conducting this research requires large and complex data sets and teams of multidisciplinary scientists, often working at remote locations.

As stated in section 2, representative science projects which require support by a PLDS were generated by the members of the Working Groups in the planning process. These research scenarios were used as a mechanism to uncover generic system functions and requirements. In addition, these scenarios helped to illuminate needs specific and unique to a given science discipline, a critical element important in determining overall system requirements.

The science projects employed as representative examples to drive the PLDS planning workshop activities are entitled:

1. Vegetation Biomass and Productivity, and Large Area Inventory

2. Biogeochemical Cycling in Forests

3. Land Surface Climatology

4. Hydrologic Modeling and Soil Erosion/Productivity Modeling

5. Multispectral Analysis of Sedimentary Basins

6. Monitoring Environmental Change

The following sections contain brief descriptions of these science projects. The descriptions are followed by a summary of the relevant needs of the projects with comments on the similarity or uniqueness of those needs. More detailed descriptions of the science projects are contained in Appendix I. Neither the brief paragraphs below nor the appendices are intended to be complete descriptions of the research, but are used to highlight the data access and processing needs of a broad range of land related research activities. The PLDS should be implemented in conjunction with ongoing research projects which will drive the design and implementation in the same way these example scenarios have served to drive the planning. The projects used in implementation could be chosen from among all NASA-sponsored research, including the scenarios described below.

3.1 Science Scenario 1. Vegetation Biomass and Productivity, and Large Area Inventory

Terrestrial vegetation is involved in the biogeochemical cycling of carbon and other elements, impacts the hydrologic cycle, and affects land-climate interactions. Present estimates of the distribution of global terrestrial biomass contain large uncertainties, which limit the ability to model and monitor global processes of serious ecological consequence such as carbon dioxide effects on climate and acid rain effects on vegetation. Better information is needed to characterize terrestrial vegetation on a global scale. This project is aimed at understanding the distribution of terrestrial biomass and productivity on a global scale and vegetation biophysical characteristics and plant processes. Developing this understanding requires a dual approach. The first step is to employ sensor systems to directly measure biomass, leaf area index, and net primary production of terrestrial vegetation. A second step is to stratify the landcover of very large areas to perform multi-stage sampling, and statistically characterize biomass and productivity within selected strata.

Such a study is currently being conducted on representative vegetation types of the boreal forest, rangeland, and cultural vegetation of North America. The primary sources of remotely sensed data are; AVHRR, Landsat MSS and TM, medium and low altitude aerial photographs, and radiometric measurements from helicopter and truck platforms using a Barnes radiometer and a C-band scatterometer. In addition, extensive field measurements of biomass and leaf area index are being acquired in the field and need to be moved to the laboratory and analyzed in real or near-real time.

The project requires knowledge of and access to a wide variety and large volume of data: field measurements, remotely sensed measurements and images, and other digital and analog data, such as topographic data, air photos, and small-scale vegetation maps. Acquisition of these data and their subsequent analysis must be accomplished quickly and efficiently, particularly during the field season, by a large and geographically dispersed team of investigators at the University of California, Santa Barbara (UCSB), Purdue University, Kansas State University, and NASA/Johnson Space Center (JSC).

3.2 Science Scenario 2. Biogeochemical Cycling in Forests; An Integration of Remote Sensing, Modeling and Field Analysis

A well coordinated interdisciplinary research effort is proposed in nutrient cycling of nitrogen-limited forests. This research will integrate data synthesis, nutrient theory, process-level modeling, laboratory chemical analysis, and both field and remote sensing techniques to explore two hypotheses:

- that total canopy concentrations of nitrogen, phosphorus, and carbon can be used to characterize the elemental cycling dynamics of these forests, and

- o that these variables can be measured by both laboratory and remote sensing techniques using infrared spectroscopy.

A mechanistic model is proposed that ties together water relationships and carbon synthesis with an explicit treatment of nutrient flow. A synthesis of extant data from well-established research sites will be used to develop and test the model. The leaf chemistry research is tightly related to the model and represents a new thrust in remote sensing of vegetation. The principal variables being pursued to fulfill modeling needs are total canopy nitrogen, phosphorus, and carbon. These variables are expressed radiometrically by characteristic reflectance and infrared absorption spectra associated with the vibrational and rotational excitation modes of chemical bonds involving these elements. A two-pronged approach is required: 1) laboratory characterization with independent chemical analysis to establish these spectra; and 2) both field portable and airborne, high- spectral resolution infrared spectrometry of leaf samples and canopies evaluated against the laboratory spectral analysis.

This program of research will be carried out by a team of ecological researchers using new analytical techniques that promise great potential for biogeochemical research. The data requirements for this undertaking are extensive. Networking between investigators at Ames Research Center, JPL, and six other locations (at universities and national agencies), is essential for the exchange of data and the sharing of hardware and software resources.

3.3 Science Scenario 3. Land Surface Climatology

The land surface system interacts in a complex and dynamic manner with the atmospheric system through processes of energy, mass, and momentum exchange to produce weather, and long-term climate. A need exists to develop a better understanding of the nature and scope of influence of the land surface on weather and climate. An improved understanding of land surface climatology processes and interactions can best be achieved through the development and validation of terrestrial and climatological process models which require many diverse types of data.

Scientific investigations in land-surfaces climatology are now being supported through a new international program, the International Satellite Land Surface Climatology Project (ISLSCP), conducted under the auspices of COSPAR and the International Association of Meteorology and Atmospheric Physics. The Goddard Space Flight Center will be participating in this project, and has already received funds to conduct workshops aimed at defining specific pilot experiments to be performed.

The overall scientific objective of the Land Surface Climatology Project is to develop a better understanding of the processes occurring within, and the interactions among, the earth's biospheric, edaphic, hydrologic, and atmospheric systems, and to determine their role in influencing or governing climate over land surfaces.

This complex research program in Land Surface Climatology requires the development of data bases, and the establishment of electronic data transfer and networking connections between the land data base and distributed computational systems at physically separate facilities.

The development of a capability for preprocessing data acquired from archives or other data bases linked to the PLDS, and electronic linkages among the PLDS participants, will greatly facilitate this research. Technological developments in data storage and automated full-scene processing will greatly assist with data flow through the PLDS and with the preparation of data for analysis undertaken at each institution, even on systems not directly linked to the PLDS. In this same line, this project and the PLDS could benefit from the development of data standards to the extent practical. The definition of generic data formats, projections, and file structures could provide an impetus leading towards greater compatibility among institutions.

3.4 Science Scenario 4. Hydrological Modeling and Soil Erosion/Productivity Modeling

Serious gaps in scientific knowledge continue to limit the quality and efficiency of models which measure the relationships of hydrology, soil erosion and sediment yields, and the effects of erosion on soil productivity. Inaccuracies in the results of these models frequently lead to incorrect policy decisions that produce significant personal and economic losses on an annual basis. Remote sensing has created new sources and types of data, and recent developments in computer and communications technologies provide opportunities for scientists to translate data from multiple sources into hydrologic and soil erosion/productivity information not previously available. This new information has led to the development of a new family of simulation models that offer great potential for meeting our forecasting objectives and providing the improved research capability necessary for a better understanding of the basic processes.

These models have been developed and tested with historical data. However, many questions concerning the utility of the models and the scientific validity of some of their formulations have not been examined because it has been impossible to obtain and interface many of the critical data elements. The existence of a PLDS, interfacing an extensive data set with a distributed scientific community, is the only mechanism that can allow a comprehensive evaluation of this new generation of remote sensing centered hydrologic and soil erosion/productivity models.

There are a number of hydrologic and soils-related projects being conducted by NASA, USDA and university organizations in the Little Washita River Basin, Oklahoma. There is an excellent historical data base and a well-designed data collection system. Missing is an efficient means of distributing the data among all of the users, operational software to merge multiple data planes in order to derive critical information, a system that will allow the interfacing of NOAA data bases for the regions surrounding the Little Washita, and a mechanism to efficiently obtain, interpret and distribute digital satellite imagery.

Under the AgRISTARS Land Resource Program, NASA-NSTL is presently conducting two research projects (Conservation Practices Inventory, and Soil Erosion Modeling) in the Little Washita Basin. The central thrust of these projects is to improve the scientific base that will allow better predictive modeling in the area of hydrology and soil erosion/productivity. The data sets that must be handled include near-real-time remote sensing, ground base hydrological and meteorological instrumentation networks, and a variety of digital archived data. Scientists and information resources at NASA, USDA, NOAA, and university facilities would be involved.

While NASA is using the PLDS to develop its expertise in network distribution of land science data, the scientific hydrology community will, for the first time, be addressing a series of extremely important science issues in a research arena that provides for real-time access to both adequate data, and computer support. The breadth of the user community, the variety of the data sets, and the distribution requirements make this science scenario an excellent case study for examining the type of problems that will be involved as NASA progresses toward global scale information systems capabilities. Solving the infrastructural, technical and user need problems that will be encountered in this relatively small area will provide the experience base that is absolutely critical if NASA is to be successful with its global scale strategy.

3.5 Science Scenario 5. Multispectral Analysis of Sedimentary Basins

Instruments and techniques for analysis of remote sensing data have improved over the last few years, but there have been few concerted efforts to apply the variety of new techniques to a single geologic problem. The Multispectral Analysis of Sedimentary Basins project is an outgrowth of the GEOSAT project, in which a few test sites were studied with a variety of remote sensing systems and techniques in order to assess their utility for geologic remote sensing. The Basins project is designed to use new techniques for analysis of remote sensing data obtained by a variety of sensors at many wavelengths for geologic analysis of a major sedimentary basin.

Sedimentary basins are large (>100x100 km) structures that occur throughout the world and that often contain economically significant amounts of oil, gas, coal, and other resources. In addition, sedimentary basins provide a record of the depositional and tectonic history of an area. The keys to efficient exploitation of the nonrenewable resources of a sedimentary basin are a knowledge of the distribution of geologic units both at the surface and within the basin, and an understanding of the evolution of the basin with time.

The objectives of this project are:

a) to evaluate the utility of remote sensing data for mapping subtle variations in sedimentary lithology,

b) to apply remote sensing data to geologic mapping of a large sedimentary basin (Wind River Basin, Wyoming),

c) to compare remote sensing data to conventional field mapping data,

d) to combine remote sensing data of surface properties with geophysical data of subsurface properties to generate a 3-dimensional representation of a basin, and

e) to employ findings to constrain models of basin formation and evolution.

The Multispectral Analysis of Sedimentary Basins project involves a number of NASA-funded investigators at JPL and the University of Hawaii. Spacecraft and aircraft remote sensing data, geophysical field and seismic data, and field and laboratory spectral reflectance measurements will be acquired, calibrated, and registered to a digital topographic base map to provide a multidimensional database. This places a very heavy load on preprocessing functions such as calibration, registration, and overlay. With existing resources, analysis tasks must be performed separately at the research nodes with inconvenient transfer of data and intermediate results by mailing of magnetic tapes. Interactive processing between nodes is highly desirable but presently out of the question. Routine transfer of text and newly developed algorithms is also necessary. Many aspects of this research program would be greatly facilitated by a Pilot Land Data System, which would quantitatively and qualitatively enhance the scientific results.

3.6 Science Scenario 6. Monitoring Environmental Change

Monitoring of environmental change is one of the most cost-effective uses of Earth satellites. The ability to view the same area repetitively at a consistent rate and with uniformly calibrated sensors allows users to determine the rate, direction, and magnitude of change of various types of Earth features for land, water, and environmental assessment. Two methods are generally used. The first involves classification and mapping of the desired features, subsequent reclassification, and identification of the features to determine their change. The second involves simply the measurement of change in one or more of the parameters detected in satellite images (such as albedo) and determination of the type of feature and condition that has changed. Both methods can be used to detect rapid changes in the state of features, but the second approach is better for the characterization of small rates of change in the condition of features.

The major objective of this project is to develop a test bed to evaluate methods for mapping and monitoring environmental change in a cost-effective manner for large geographic areas. Those methods can then be routinely implemented by agencies and institutions that are responsible for environmental monitoring. This objective can be achieved by a thorough assessment of presently available spacecraft, aircraft, and ground monitoring methods; by the design of new and improved methods involving new sensors, new methods of data processing, and by an evaluation of the effectiveness of these methods. In order to produce timely and meaningful results, data for large areas must be acquired and assembled (rectified, registered, mosaicked) quickly. This would require improved data access via networked data bases and catalogs, and simplified data retrieval, and access to computer resources over a network for computationally-intensive registration, mosaicking, and other algorithms. This project would also be facilitated by efficient scientific communication between remote investigators. Electronic mail and computer conferencing would improve research management and scientific interaction. Institutions that should be involved include the U.S. Geological Survey, NASA, the U.S. Department of Agriculture, the Environmental Protection Agency, and the Federal Emergency Management Agency.

3.7 SUMMARY OF SCIENCE NEEDS

This section contains a summary of the information system support needed by these scenarios in order to accomplish their research goals. Figure 3.1, which shows a general model for an information system to support land sciences research, was used as a guide for prioritizing the research functions that require PLDS support. The purpose of this diagram is to generalize the steps involved in performing this type of research, and to identify which functions could be supported by a PLDS in order to enable the researcher to do the project more cost effectively and efficiently. These steps were broken down into processing functions (data input, preprocessing, processing, analysis, output production), and functions required for a networking and data access system. Table 3.1 summarizes the results of analyzing each project in this manner and uses a numeric code in prioritizing functional requirements for the PLDS.

The number used for assigning priorities were defined as follows:

1 - enable the scientists to do the research.

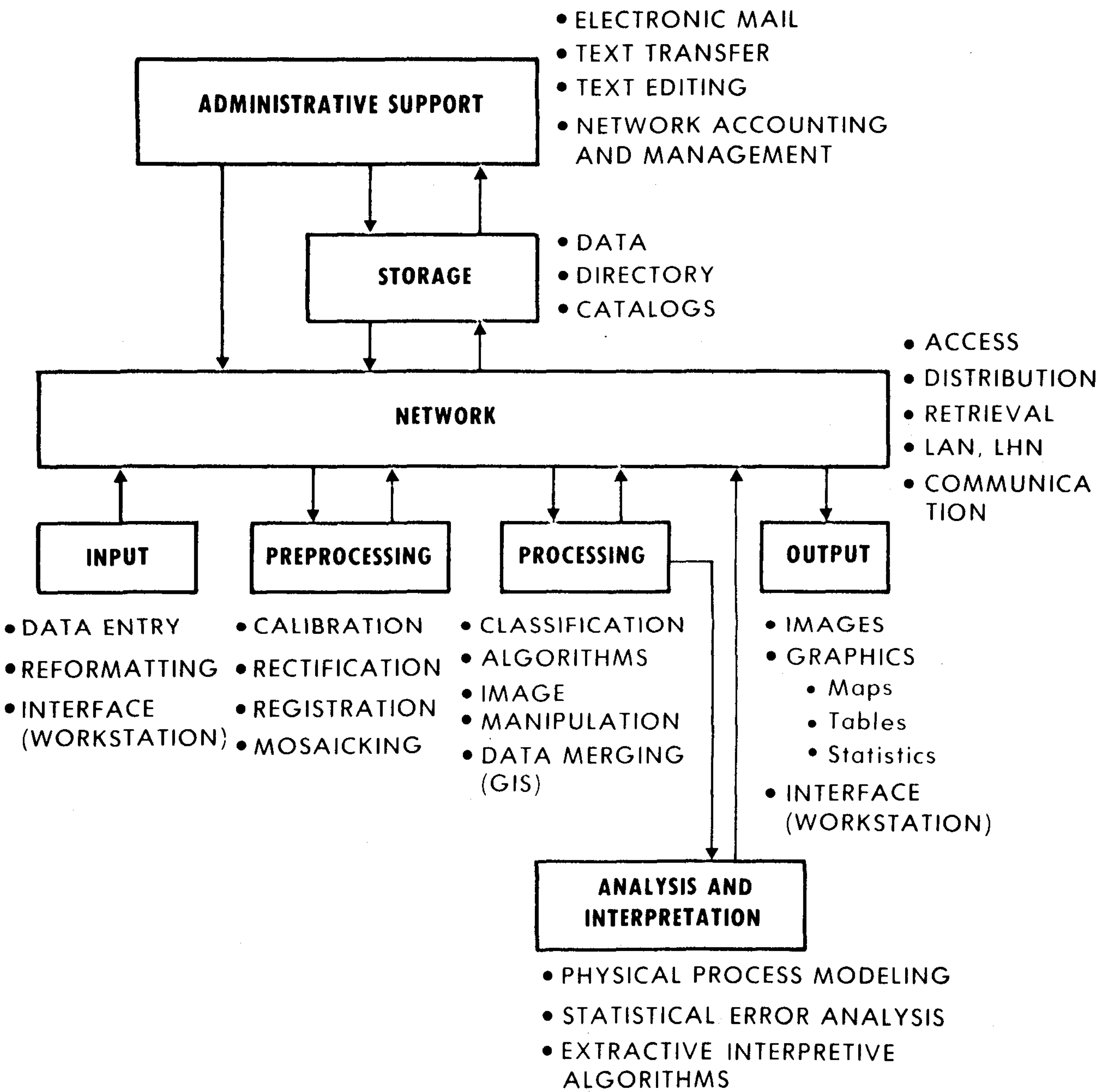
2 - enhance the scientists' ability to do the research.

3 - Research could be accomplished now without PLDS support for this function, but it is a service that would be useful.

4 - Do not need support for this function from the PLDS (accomplished effectively with current capabilities).

Table 3.2 shows the functions in order of the priority which they were given in Table 3.1. These science scenarios require support from a PLDS in data storage, input, preprocessing and distribution (data and catalog access, and communication links to allow sharing of data, algorithms, CPU and peripherals) on a high priority basis. Lower priority is assigned to support for analysis and output, and network administration functions (although some of these items may be implicitly required to support the functions that were assigned higher priority). Appendix I.0 gives more specific information on the data sources required for the research scenarios, and the institutions

FIGURE 3.1 FUNCTIONS OF AN INFORMATION SYSTEM TO SUPPORT LAND SCIENCE RESEARCH



involved, and shows how this type of information can be used to elucidate the specific architecture of a pilot system.

Success of these scenarios depends to a large degree on efficient access to and sharing of data and resources (hardware/software) among the investigators. The PLDS could facilitate the development of linkages among research centers not currently possible. The linkages would provide a mechanism for the timely exchange of data, information, software, and the access to, and sharing of computational resources, such as remote CPU-intensive systems, and peripheral devices. Each project involves a wide array of diverse data sources making data management for these research scenarios very complex. A data management system capable of storing, correlating, retrieving and sharing these data among scientists is a key element in the processing flow. Preprocessing tasks, such as reformatting, data encoding, rectification and registration of these diverse data sets would also be facilitated by the PLDS through a cooperative sharing of facilities and software.

Table 3.1

Summary of Functional Requirements of the PLDS Planning Science Scenarios

	Science Scenario						
	1	2	3	4	5	6	Total
1. User Node Processing Functions							
1.1 Input							
Data encoding	1	2	2	1	2	2	10
Data reformatting	1	2	1	1	2	2	9
1.2 Preprocessing							
Data Calibration	2	3	1	1	2	2	11
Image Registration	1	1	1	1	1	2	7
Image Mosaicking	1	1	3	2	1	2	10
1.3 Processing							
Multi-source geocoded data overlay	1	1	1	1	1	1	6
Image and statistical processing (software sharing)	1	1	2	1	3	1	9
1.4 Analysis							
Statistical analysis	2	2	4	2	4	4	18
Modeling	1	2	4	2	4	4	17
1.5 Output							
Image	1	2	3	1	3	4	14
Statistical (tabular)	2	2	3	2	4	4	17
Tables and figures (graphic)	2	2	3	2	4	4	17
Storage media - CCT, disc	2	3	3	2	4	2	16

Table 3.1, Cont.

	Science Scenario 1	2	3	4	5	6	Total
2. Network Functions							
2.1 Storage							
Directory (Catalog [2])	1	1	3	1	2	2	10
Catalog	2	1	2	1	2	2	10
Data	2	2	1	2	3	2	12
2.2 Distribution							
Access to archive data	1	1	1	1	1	2	7
Networking of processing	1	1	1	1	3	2	9
Shared peripherals for output	2	2	2	1	4	2	13
2.3 Administrative Support							
Electronic mail	2	1	3	3	2	4	15
Text transfer	2	1	3	3	2	4	15

Table 3.2

Ordered Ranking of
Information System Functions
Requiring PLDS Support by the Science Scenarios

	Total
Multi-source geocoded overlay	6
Access to archived data	7
Image registration	7
Data reformatting	9
Software sharing	9
Networking of processing	9
Mosaicking of images	9
Directory of information	10
Calibration of data	10
Data encoding	10
Data storage	12
Shared peripherals for output	13
Image output production	14
Electronic mail	15
Text transfer (compatible text editing)	15
Output storage media	16
Tabular output production	17
Graphic output production	17
Modeling	17
Statistical analysis	18

4. LAND DATA SYSTEM OVERVIEW

4.0 INTRODUCTION

A Land Data System (LDS) must provide the land science and applications users with a powerful, friendly and cost-effective computing environment for conducting research. The environment must support the full spectrum of information functions necessary to conduct land sciences investigations including location, acquisition, processing, analysis, and transfer of data.

The overall goal of an LDS could be described as follows:

> To provide a powerful and responsive system to support land science research, to facilitate understanding of the land resource complex through mapping, inventory, monitoring, predicting, and modeling, and to provide the sharing of and access to land-related data sets and advanced techniques and processing capabilities by scientists in a variety of disciplines and locations.

Currently, land sciences research is characterized by a vast array of geographically dispersed users with varying levels of technical capability operating in a more-or-less independent manner. To establish and validate the potential of and functional design for a comprehensive Land Data System to serve that community, a prototype, a Pilot Land Data System, must be developed. LDS is distinguished from PLDS in scale, scope and by the experimental/developmental nature of the PLDS. This section provides an overview of the concept of a Land Data System. The proposed development of the PLDS will be described in Section 5.

4.1 LDS CHARACTERISTICS

Required characteristics of LDS include:

- Ability to use the LDS efficiently with a user-friendly interface requring minimum training on and/or understanding of the total system.
- Systematic archiving and maintenance of relevant primary and derived scientific data.
- Access to data management tools that will allow researchers to rapidly review and select data needed to support research.
- Rapid and simple access to all archived data necessary to conduct scientific research.
- Simple access to existing bibliographic information systems.

- Provision of a full history of origin, calibration information, quality assessment, and processing steps for all data.
- Ability to have data registered, calibrated, projected, and otherwise modified as a service, with minimal scientist interaction required.
- Ability for system to modify, correct or change data into a format compatible with the LDS.
- Ability to rapidly transfer scientific and technical information among nodes routinely and easily.
- Ability to provide sufficient processing power to the scientist so that the research can be performed in a timely manner.
- Improvment of technology for local processing and display capabilities at research nodes.
- Access to remote computers and computer peripherals by users for scientific analysis.
- Ability to access software tools from other nodes in support of scientific research projects which could then be implemented in the local computing environment.

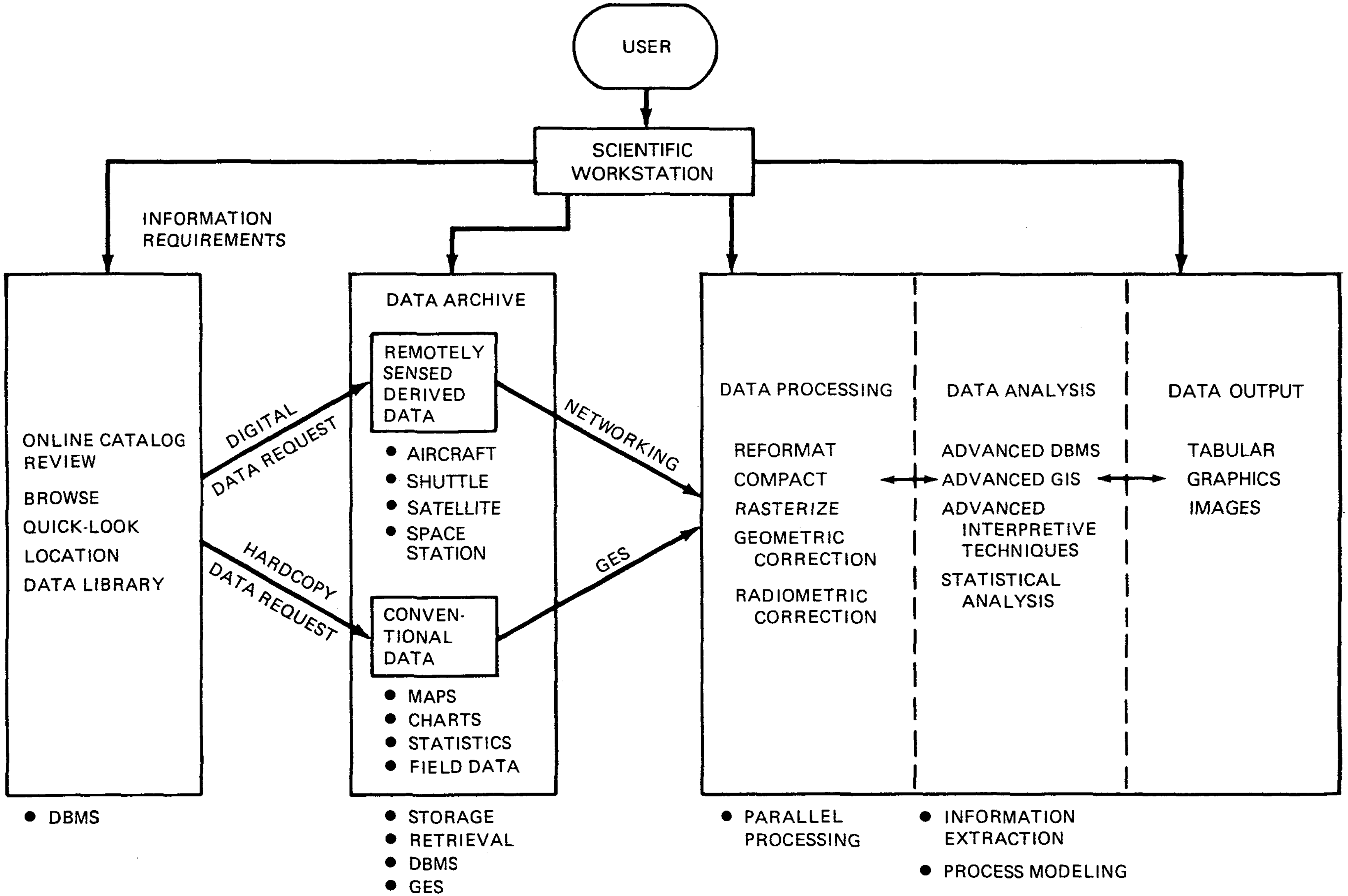
Successful implementation of a system exhibiting these characteristics can completely change the character of land science research. Such a system would enable multidisciplinary, multi-institutional research which is not now practical. It could allow experiments to be conducted in near-real time commensurate with experimental designs.

4.2. LDS SUPPORT OF USERS

As currently envisioned, LDS would be a computer system with a distributed architecture, intelligent attributes, and value-added services designed to support land sciences in the coming decades. In concept, the LDS would support the most technically demanding computer operations with minimal user knowledge of, or experience on, the system. The overall goal of the system would be to reduce the information processing burden on scientists without compromising their ability to conduct scientific investigations.

The LDS could employ powerful microprocessor workstations as the user interface for the scientist. Workstations would be linked, using high speed digital communications, to supercomputers, background and foreground processors (e.g., array processors, and data base machines) and advanced data management systems. A functional representation of the LDS is presented in Figure 4.1.

FIGURE 4.1 — LDS FUNCTIONAL OVERVIEW



4.3 LDS CONCEPT

Based on an analysis of existing systems that support land science research, the long-term goals identified in this study, the needs of the users, and new and emerging technologies, a preliminary conceptual system design has been formulated. This design consists of a set of subsystems which support the functions listed in Section 4.1. The LDS would consist of five major subsystems (see Figure 4.2):

- Data Management Subsystem,
- Communications and Networking Subsystem,
- Intensive Computational Processing Subsystem,
- User Interface Subsystem, and
- Input/Output Interface Subsystem.

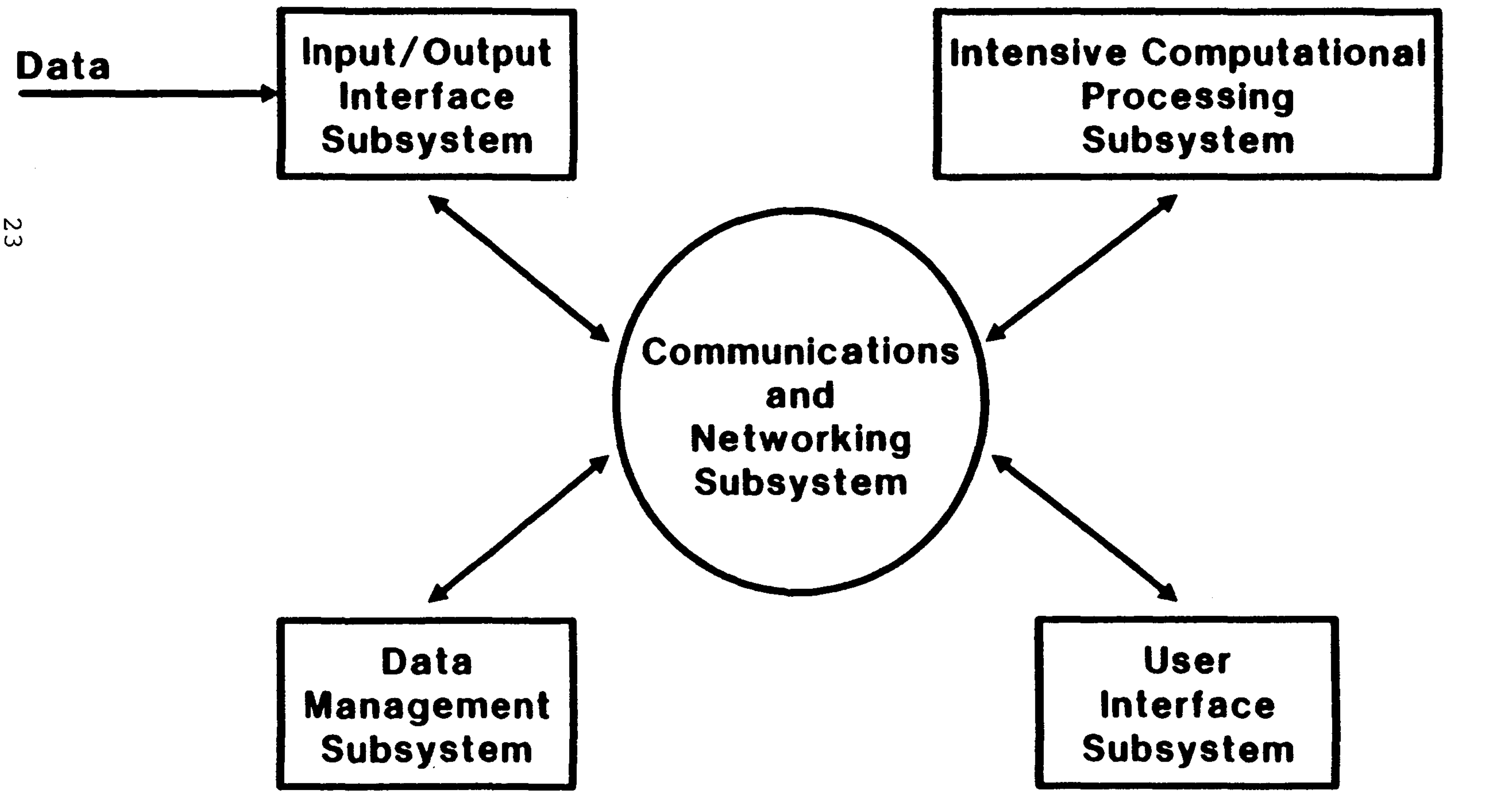
The Data Management Subsystem would perform the function of providing all data and information about the data to the scientists. This subsystem would be designed to permit the scientist user interaction with little prior familiarity with the system. Users would communicate with the subsystem using natural language. Operation and management of the subsystem and many data management functions would employ knowledge-based engineering and expert system technology derived from the knowledge and experience of data system designers, users, and managers. The Data Management Subsystem would also have the ability to store and update large amounts of data and support many users concurrently (see Appendix II.1).

The Communications and Networking Subsystem would provide a mix of wide-band, high-speed, and low rate digital communications to be used as is appropriate. This subsystem would support inter and intra-institutional communications and would facilitate a near-real-time interface between the User Interface Subsystem, the Data Management Subsystem, the Intensive Computational Processing Subsystem, and the Input/Output Interface Subsystem. Such communications would be supported by several technologies, including packet switching network communications, local area networking, and satellite communications (see Appendix II.2).

LDS data manipulation and analysis would occur on a number of the subsystems. The Intensive Computational Processing Subsystem would provide a service for supporting processing where the power of advanced large-scale computers is required. Subsystem design would be based on present NASA Center and institutional resources, new supercomputer technology, and background and foreground processors, as well as expert systems technology for management and control operations. In addition, this subsystem must have the capability to assist the scientist user in performing technically demanding tasks such as image and data interpretation and pattern recognition (see Appendix II.5).

FIGURE 4.2

LDS
SUBSYSTEMS OVERVIEW



The User Interface Subsystem would be composed of a range of workstation types with capabilities depending upon user needs and resources. These workstations will be connected to the other subsystems by means of the Communication and Networking Subsystem and will interface to individual processing facilities at the user nodes, or serve as free-standing processing stations. Workstations will be based on microprocessor and minicomputer technology. These microprocessors could provide a working environment that until recently could only be achieved with the supermini or mainframe type systems (see Appendix II.4).

The last subsystem, the Input/Output Interface Subsystem, would connect the overall LDS with the outside computer and data/information world (see Appendix II.3). It will be the function of this subsystem to perform reformatting, modification, and data manipulation, and to allow the overall LDS to communicate with other systems, institutions, and data archives/depositories.

A potential configuration for in some detail an LDS node is shown in Figure 4.3. This shows a node at a NASA Center in some detail, but nodes are also expected to be located at universities and other institutions or agencies.

4.4 AN INTELLIGENT LDS

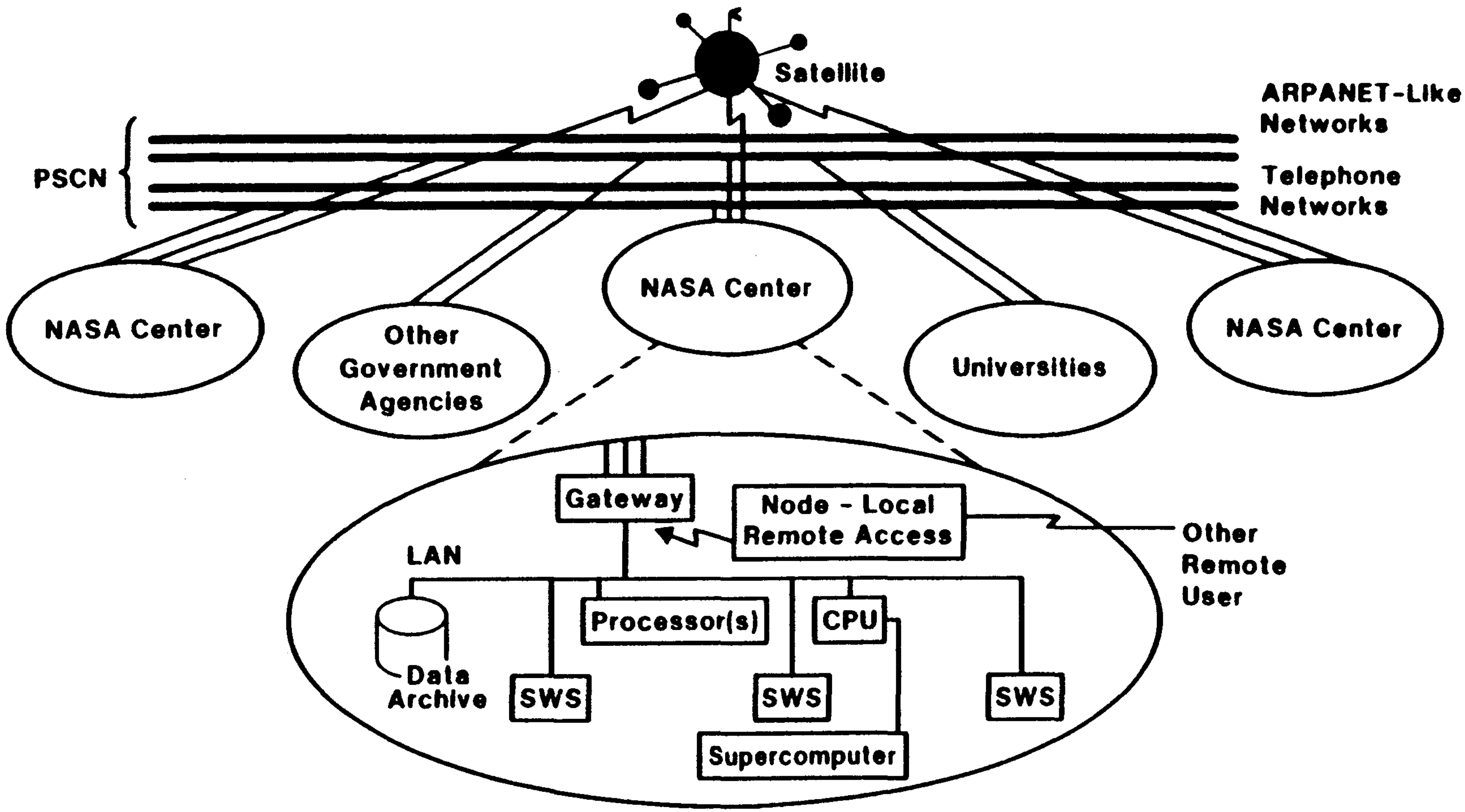
Creating a system that can intelligently assist the scientist can begin to reduce the data processing burden. The technological area on which the system intelligence development will be based is commonly known as Artificial Intelligence (AI). Subdomains of AI that could be considered for an LDS include knowledge-based engineering (expert systems), and natural language processing.

Knowledge-based expert systems could provide functional intelligence to LDS to first, support independent operations of the system, and second, provide the user with automated capabilities in specific technical areas to perform tasks that typically require detailed expertise. The expert systems could also contain the symbolic descriptions that characterize the definitional and empirical relationships in a specific land resources knowledge domain and the procedures for manipulating these descriptions to successfully solve problems. Expert systems would be constructed to assist in the performance of a number of generic tasks for the land scientist including:

- data access,
- data transfer,
- data manipulation,
- data analysis, and
- system diagnostics.

FIGURE 4.3

LDS COMMUNICATIONS AND INFORMATION ANALYSIS OVERVIEW



PSCN – Program Support Communications Network
SWS – Scientific Work Station

4.5 PLDS TO LDS TRANSITION

When fully implemented, the PLDS will be capable of supporting a subset of the NASA and NASA-related land science community and will establish the basis for transition to a Land Data System. Systems concepts for the phases of PLDS, and transition to LDS, are illustrated in Figure 5.1. The transition between PLDS and LDS could occur starting in CY 1989. A more detailed discussion of some near-term efforts that will be performed to support the PLDS are presented in Section 5. Given the need for improved information systems for land science and using the LDS concept detailed here as a foundation, Section 5 provides recommendations, guidance and phasing for implementing enhanced capabilities that support land science research, such as the scenarios addressed at PLDS workshops (see Section 3 and appendix I).

5. PILOT LAND DATA SYSTEM DEVELOPMENT

5.0 INTRODUCTION

Pilot Land Data System development during the period FY 1985 through 1987 will be important in establishing and verifying the design characteristics for a fully implemented PLDS. This portion of the report describes what the PLDS system should do by the end of FY 87 and provides specific recommendations for 1985 and 1986. Recommended actions provide engineering guidance in support of the Working Groups' expressed long-term goals and operational requirements.

PLDS development should be based on three important principles. First, the system should be built on existing capabilities wherever possible. This keeps costs down and can permit concept testing and an opportunity for near-term assessment of potential scientific return. Second, a structured system engineering effort should be initiated at the onset of the project to assure that even during early development, long-range goals are fully taken into account. Finally, new technologies should be regularly reviewed and integrated where appropriate within budgetary constraints.

This approach provides a minimum start-up risk while drawing on the experiences of NASA Centers, universities, and other agencies, and can also provide a focus for NASA and NASA-funded research in computer science and system design. The rate of progress from a baseline of available technology towards PLDS goals will be scaled by dollar and personnel resources availability, and management interest. It is important to note that a good start already exists, and the land science investigators will continue their efforts of the last decade to improve data management and communications; the PLDS provides a formalism and a focus for further progress.

5.1 FY 85-FY 87 GOALS

A list of the functional requirements generated by the Working Group (Section 4.1) can be summarized as a series of goals for the PLDS by the end of fiscal year 1987. These goals are:

- Establish the capability for scientists at various NASA Centers, universities, and other agencies to communicate quickly and easily with respect to land science matters from their own local work sites.

- Build a directory and catalogs for data sets distributed among NASA Centers, universities, and other agencies. Establish and perform a curation function for these data sets where necessary.

- Establish a data management system capability designed for efficient local and remote use.

- Demonstrate the capability for scientists to remotely access, use, and transfer data.

- Demonstrate that remote requests for value-added services (calibration, registration, rectification, projection, analysis, and outputs) can be answered easily in a timely way.

- Demonstrate that a scientist at one location can remotely access and use hardware and software at another location.

- Demonstrate the portability and expandability of the system.

Human factors considerations are also important. Monitoring of user patterns and user concerns should be built into the system in some formal way. This is important to the evolutionary development of a system which is responsive to the needs of the land science users.

5.2 PLDS START-UP PHASE

In the fiscal years 1985 and 1986 significant progress can be made toward goals outlined above. While a different set of research scenarios may be used to drive the pilot implementation, based on the scenarios used in the planning workshops, the following short-term recommendations by the Working Groups can be outlined:

It is essential to establish and test communication capabilities by identifying, making gateways to, and exercising existing communication channels among Ames Research Center, Goddard Space Flight Center, Jet Propulsion Laboratory, Johnson Space Center, National Space Technology Laboratories, several universities, and other agencies. The bulk of this work will likely consist of building local area network capability between a node at a NASA Center, university, or other agency, and the workstations at the scientists' work sites. Once the existing channels are established and exercised, it will be necessary to identify problems and establish procedures for connecting them.

It is necessary to begin building a dispersed data library, identify and obtain access to existing data sets, and to build data catalogs at ARC, GSFC, JPL, JSC, NSTL, and selected universities. For major existing data sets this access could be through standard commercial or existing scientific channels. Further, the PLDS should either establish a NASA Centers' team or designate a Center to establish links with, and catalogs of, data sets at EROS Data Center, USGS, NOAA, and other appropriate agencies and institutions. Again, these links, in some cases, may take the form of tapping into existing scientific data acquisition channels. A workable link with one of these data sets should be established as a demonstration prototype. Scientists and technologists at each NASA Center should be selected to form a curation unit and develop a central directory of catalogs. Finally, protocols and formats for adding to data sets at NASA, and for linking to data sets outside the NASA domain at universities and other agencies must be developed.

Establishment of a data management system accessable from a user workstation should begin with a review of existing data management systems with special emphasis on image and spatial data management. This process should then proceed with a team composed of personnel from NASA Centers, universities, and other agencies to adopt or modify existing data management systems identified through the review process discussed above.

Each NASA Center in the program and selected universities should set up or adapt existing workstations. These workstations could be selected from a number of systems ranging from the large personal computers to small mainframe computers. Even more significant for user satisfaction will be the data management software that may be accessed from the workstation. Some emphasis on natural language interfaces could be important here.

Once the network, data library, data management systems, and workstations are in place, experiments based on selected science scenarios can be performed exercising the PLDS capability to remotely access, use and transfer data to and from selected participating NASA Centers, universities, and other agencies. It is the judgment of the Working Group that to get to the point of testing the PLDS utilizing science scenarios will take approximately one year from the inception of a PLDS program. Actual testing could be concentrated in the one to one and a half year time frame from program start-up. Figure 5.1 shows this and the further development steps to be discussed below. Figure 5.2 shows some of the potential nodes of the PLDS envisioned at this time; others may appear in response to program developments.

Following this mid-point milestone of access, use, and transfer of data to serve land science purposes, the next effort should be on the development of remote value-added services capability. Experiments that will demonstrate calibration, registration, and rectification of diverse data sets in quick response to remote requests appear to be the top priority distilled from the science scenarios. To achieve this, it will be necessary to review local site capabilities in these areas, to select the most promising sites for two-node experiments, and to expand to multiple nodes the provision of value-added services. Of lower priority are analysis and output services, which could be added if users' needs warranted. An important aspect of this portion of the PLDS is the establishment of management policies for response priorities, protocols, and prices or funds transfer guidelines.

Once the delivery of value-added services has been successfully demonstrated between scientists at different nodes on the network, it will be appropriate to expand to progress to direct use of the equipment and software at a remote site. A scientist will then be able to perform value-added services using capabilities located elsewhere. In addition, experiments to demonstrate the remote use of hardware and software for analysis as well as access, use, and transfer of data from libraries should be planned as defined by the selected science scenarios. Finally, management policy issues will be an important aspect in all phases of the program.

Throughout all PLDS phases there will be the opportunity and necessity to demonstrate software portability and system expandability. It is suggested that expansion to other data, nodes, and facilities be built into the PLDS plan from the outset and that software portability be a constant goal of the program. Since satisfaction with the capabilities and ease of use of the PLDS is crucial to its success, frequent meetings of active users should be held. Initially, quarterly meetings are suggested. In addition, an on-line electronic mail service for direct communication of problems and providing assistance would also be important. In this way, grievances and frustrations can be quickly aired and solutions expedited.

5.3 PLDS EVALUATION

Once a set of representative science scenarios are selected to drive development of and be served by the PLDS, scientists directly involved can present their achievements and problems to other scientists for comment. As a part of PLDS progress review meetings, developers and users should exchange critical evaluations of the status of and results concerning the systems and science scenarios. This can be accomplished both through the meetings, peer review sessions, and by publication of results in reviewed scientific journals. Particular attention should be paid in such presentations to comparisons of scientific productivity before/after PLDS, keeping in mind that during initial PLDS development, research may become temporarily less efficient as the project comes up the learning curve. Scientists not directly participating in the PLDS program should be encouraged to test, on a non-interference basis, system capabilities as they develop. This mechanism could be used by PLDS and NASA management to provide independent viewpoints hopefully devoid of ownership biases.

5.4 LONG TERM GOALS

Many of the technologies which will be employed in the PLDS are in such a state of rapid development that it is difficult to predict the future for these areas. The PLDS should be developed with a full awareness of the volatility of these technologies so that building from existing systems will not lead to built-in obsolescence. "Upward compatibility," "open-sided technology," and "evergreen technology" are a few of the phrases used in the industry to describe an avowal not to get caught in a technological dead end. The PLDS should make the same resolution as it heads for the mid-1990's system.

FIGURE 5.2 EXAMPLE PLDS NODES & SITES

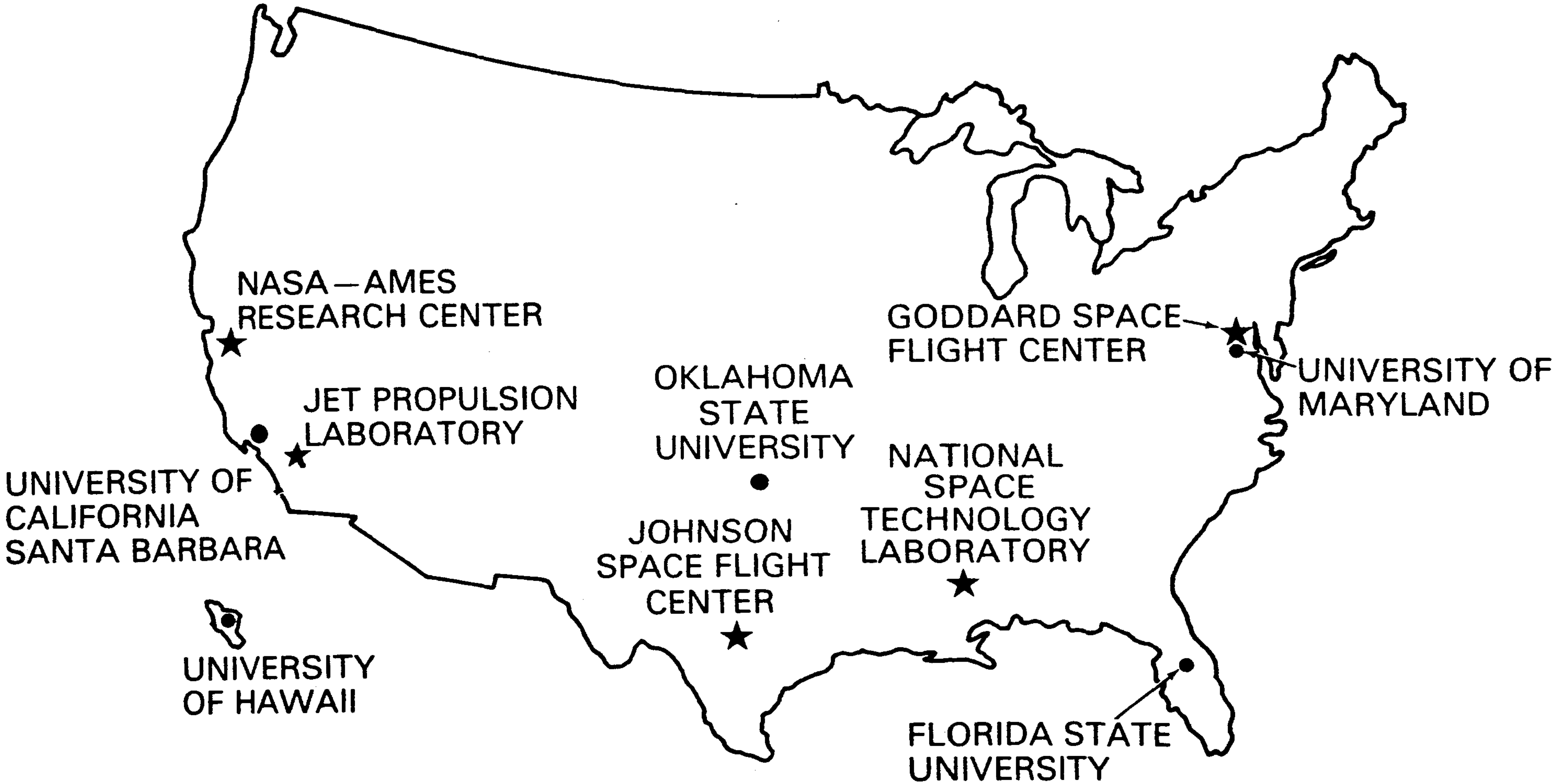
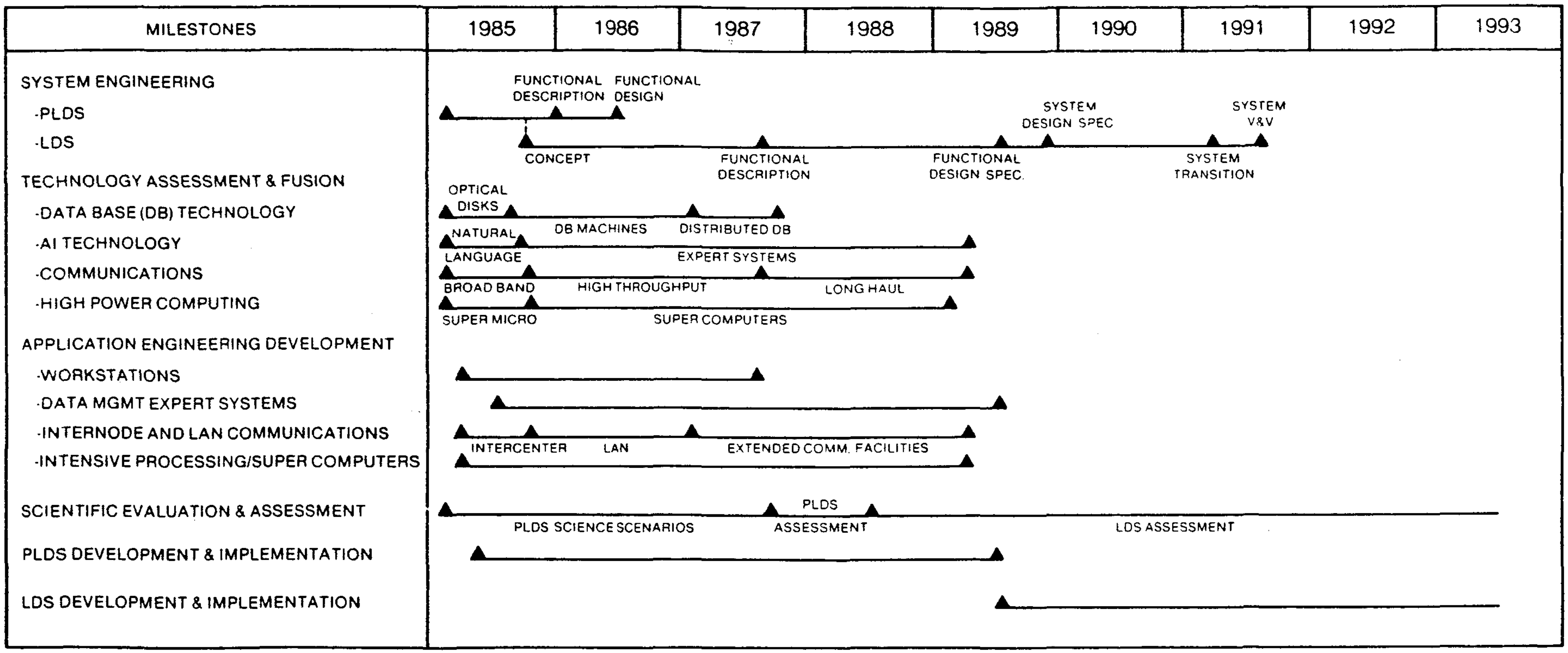


FIGURE 5.1 PLDS/LDS DEVELOPMENT SCHEDULE



6.0 CONCLUSIONS AND RECOMMENDATIONS

Satellite remote sensing systems offer a truly unique tool to the land science community. These systems can provide scientists with data of a type and on a scale previously unattainable. Yet, looking forward to the capabilities of Space Station and the Earth Observing System (EOS) we are aware that full realization of the potential of satellite remote sensing has consistently been handicapped by inadequate information systems. This must not be allowed to continue. Recent studies and activities, and the experience of the participants in the PLDS Planning Workshops, suggest that the full potential of remote sensing will not be achieved without expanded efforts to effectively integrate remote sensing and information sciences technologies. Remote sensing is a critical technology for land science, whose use has been inhibited by the lack of a total systems approach. Such an approach must not stop at the ground receiving station, but must fully integrate all aspects of the information systems needs of NASA and NASA-sponsored researchers.

Information sciences technology is developing rapidly, as is remote sensing. There is a natural complementarity between remote sensing and information science and technologies. There is great willingness on the part of the respective discipline scientists to collaborate, as has been more than adequately demonstrated in these workshops. Improvement and integration of these technologies is a necessity if the full potential of satellite remote sensing for the land sciences is to be achieved.

6.1 CONCLUSIONS

Based on the work conducted under the planning activity the conclusions of the Working Groups are that:

- There is a need to improve the ability of NASA and NASA-sponsored scientists to locate, access, process and analyze remotely sensed and other land resource science data;
- rates, volumes and types of remotely sensed data severely tax current data and information systems; and
- unless the ability to handle these and other land science data is established now, effective use of data from future systems (e.g. Space Station) will be severely impacted.

Further, it is recognized that:

- Land scientists, under NASA sponsorship, have complex, high volume, multidisciplinary information requirements;
- These requirements, if satisfied, will enable researchers to better address important, multi-disciplinary science questions, and can lead to improved understanding of many land processes;

- The technology exists to begin to enable land resources scientists to function more effectively; and, finally,

- The proposed PLDS can be a means of increasing scientific productivity through a more effective use of information science technology.

6.2 RECOMMENDATIONS

Based upon these conclusions, the Working Group strongly recommends that:

- A Pilot Land Data System be implemented beginning in FY 1985 to link NASA and NASA-sponsored land researchers;

- The initial system be a limited-scale, modular, distributed information system;

- The system have the strong continuing and cooperative involvement of NASA and NASA-sponsored land and information scientists;

- An advisory committee of land and information scientists be constituted to review PLDS progress, provide advice and guidance and periodically report to NASA Headquarters Information Systems and Earth Science and Applications personnel on PLDS progress; and finally,

- Information Systems Office and Earth Science and Applications Division personnel closely coordinate to insure that the land science scenarios chosen to drive the PLDS design are as representative as possible of the range of data and information systems requirements of the land resources community.

APPENDICES

I. SCIENCE SCENARIO APPENDICES

I.0 INTRODUCTION AND OVERVIEW OF SCIENCE SCENARIO APPENDICES

The following sections are detailed descriptions of the land sciences research projects that were used to develop the PLDS concept during the planning workshops. (See Sections 2 and 3 for an explanation of how these scenarios were used for PLDS planning.) Each description provides a brief outline of the background, objectives and approach for each project, and then delineates its data processing requirements, with specific emphasis on how these requirements could be met by a PLDS. These scenarios are not intended to be complete statements of the proposed or ongoing research. Rather, they emphasize the information and communication requirements of the projects, the present constraints on the ability to conduct the research, and methods that the PLDS can bring to bear to begin to alleviate some of the problems.

Before describing the individual scenarios, summary information is presented in this section on the data needs and communication links required by the six scenarios as a whole. This information is needed in order to construct the appropriate communication links among investigators at a variety of institutions service centers, data archives, and other nodes on the PLDS. It is useful in selecting the most important data sets to include initially in the high level information directory, and the data management system. This method will be used in the actual system design and implementation planning of the PLDS.

A matrix of institutions and science scenarios are given in Table I.0.1. This summary illustrates the need to interact with other institutions, including federal agencies other than NASA (primarily USGS, NOAA, and USDA), state and international agencies, national laboratories, and universities, both to obtain data and collaborate on research.

Table I.0.2 summarizes the data requirements for the projects. The data required are diverse in source and form, but can be generally grouped into three categories: 1) digital data from air- and space-borne sensors, 2) digital and analog thematic land surface data, and 3) field measurements to be used in conjunction with remotely sensed and other georeferenced data. Note that several key data sets are used by almost all of the projects, namely digital image data from AVHRR, MSS, and TM satellite sensors, and digital terrain or topographic information.

Table I.0.1

Institutions Involved in Science Projects

Key:

S = Scientific interaction (collaboration)
D = Data interaction (exchange data)

	Science Scenario					
	1	2	3	4	5	6
1. NASA						
ARC		S,D		D		S
GSFC	D		S,D	S,D	D	S
JSC	S,D	S,D		D		
JPL		S,D	S,D	D	S,D	S
NSTL			S,D	S,D		
2. Federal Agencies						
AID			S			
DMA		D	D	D		
NOAA	D	D	S,D	D		S,D
NWS			D	D		S,D
NPS (DOI)		S,D		S,D		
USDA -		D		S,D		
ARS-S. Plains Lab				D		
Nat'l Water Data Bank				D		
SCS			D	D		
US Hydrology Lab			S,D	D		
USGS -	D	D	S,D	S,D	D	S,D
EDC						
Water Resources Division						
National Laboratories -						
ORNL	S	S,D	S			
Woods Hole	S					S,D
3. State Agencies				D		S,D

Table I.0.1, Continued

	Science Scenario 1	2	3	4	5	6
4. Universities						
CUNY - Hunter	S					S
Clark U.			S			S
Florida State U.		S,D				
Iowa State U.				D		
Kansas State U.	S,D					
Oklahoma State U.		S,D	S,D	S,D		
Oregon State U.		S,D				
Purdue - LARS	S,D					S
Stanford U.		S				S
SUNY - Binghamton	S,D		S			
Texas A & M			S	D		
U.C.L.A.		S,D				
U.C.S.B.	S,D	S,D				
U. of Hawaii					S,D	
U. of Kansas			S	D		
U. of Maryland			S	S,D		S
U. of Montana		S,D				
U. of Missouri				S,D		
U. of Oklahoma				S,D		
5. International						
African Regional Commission			S			
UN/FAO			S,D			
UNEP			S			S,D
WMO			S,D			

Table I.0.2

Data Requirements

	Science Scenario						
	1	2	3	4	5	6	#
1. Digital Sensor Data							
1.1 Spacecraft							
AVHRR	X	X	X	X		X	5
GOES-VISSR	X			X			2
HCCM		X			X		2
LANDSAT-MSS	X	X	X	X	X	X	6
LANDSAT-TM	X	X	X	X	X	X	6
MLA	X	X	X	X		X	5
NIMBUS 5,6 -ESMR			X				1
NIMBUS 7 -SMMR			X				1
SEASAT					X		1
SIR-A,-B	X	X		X	X		4
SPOT	X					X	2
HIRS/MSU			X				1
1.2 Aircraft							
AIS, AVIRIS	X	X	X	X	X		5
DAEDELUS	X	X					2
LAPR-2	X		X				2
LIDAR		X		X			2
Radiometer	X	X					2
SAR	X			X	X		3
Scatterometer					X		1
SLAR	X	X		X	X		4
TIMS, AVHIR		X		X	X		3
TMS	X		X	X			3
2. Photography							
Aerial Photographs	X	X	X	X		X	5
Large Format Camera	X	X	X				3

Table I.0.2, Continued

	Science Scenario						
	1	2	3	4	5	6	#
3. Digitized Data							
Evapotranspiration	X	X	X	X		X	5
Fire history	X	X		X		X	4
Geology	X	X	X	X	X	X	6
Gravity and magnetic					X		1
Ground Water	X		X	X		X	4
Land Cover	X	X	X	X		X	5
Soil	X	X	X	X			4
Stream network	X	X		X		X	4
Topographic (terrain)	X	X	X	X	X	X	6
4. Field and Lab Data							
Biogenic gas concentration		X					1
Biomass samples	X	X	X			X	4
Forest dimension measures	X	X		X			3
Geochemical data					X		1
Geologic spectra					X		1
Gravity data					X		1
Hydrologic data	X			X		X	3
Land cover	X		X	X		X	4
Magnetic data					X		1
Meteorological data (ppt., albedo, ST)	X	X	X	X		X	5
Micro-meteorological data	X	X	X	X			4
Radiometer measures	X	X	X	X			4
Seismic data					X		1
Site description (ecological)	X	X	X	X			4
Soil (type, texture, depth)	X	X	X	X		X	5
Reflection seismic					X		1
Vegetation transects	X	X	X	X			4

I.1 SCIENCE SCENARIO 1: VEGETATION BIOMASS AND PRODUCTIVITY, AND LARGE AREA INVENTORY

Daniel Botkin, John Estes, Kerry Woods, and Janet Franklin
University of California, Santa Barbara

I.1.1 PROJECT BACKGROUND

This project developed out a series of workshops directed by Dr. Daniel Botkin for the NASA Office of Life Sciences, designed to improve understanding of the potential of remote sensing to facilitate research in global ecology. A principal finding of these workshops concerned the need for accurate mapping and estimates of the areal extent and biophysical characteristics of global surface cover types. It was the consensus of many of the scientists attending that current surface cover maps for the vast majority of the globe are, to various degrees, inaccurate and subjective. These researchers acknowledged that current estimates of pools and fluxes in many biogeochemical cycles, which depend on these estimates of surface cover as sources and sinks, could contain serious flaws.

Conventional surface cover maps can be generated in a variety of ways and include information from many sources. Accuracy assessments of such products are difficult. Current estimates of biophysical properties (e.g., biomass, productivity) are derived from these maps and extrapolation from intensive studies on very small areas. Thus the validity of the estimates is suspect. Alternatively, NASA-developed satellite technology offers the potential for the generation of globally consistent data sets from which surface cover information can be generated and the accuracy of this information more easily verified. The project described below was initiated to do this.

The project consists of two parallel efforts involving investigators at NASA Johnson Space Center; University of California, Santa Barbara (UCSB); Laboratory for Applications of Remote Sensing, Purdue University (LARS); Kansas State University; City University of New York, Hunter College; State University of New York, Binghamton; and Oregon State University.

I.1.2 OBJECTIVES

The purpose of this research is to improve understanding of vegetation characteristics and processes, such as biophysical characteristics (leaf area index, biomass, net primary productivity, canopy temperature, and albedo), and plant physiological processes (evapotranspiration, photosynthesis, and respiration).

The goals of the project are: 1) development of methods to measure directly, by remote sensing, biomass and net primary production of terrestrial vegetation, and 2) to employ satellite imagery, primarily Landsat AVHRR data, for assessing and improving the current representational accuracy of major accepted sources of continental -scale land cover information. The boreal forests of North America are being used as a test system.

I.1.3 APPROACH

Two approaches are being used to meet these goals. The ability to infer key vegetation characteristics from remotely sensed data is principal to the economy of large-scale research. Therefore, in the first approach, close-range spectral signatures of vegetation, collected from a low altitude platform, are correlated with laboratory measurements such as leaf reflectance. These data are used as a basis for higher-altitude measurements from aircraft and spacecraft where atmospheric conditions attenuate and distort the characteristics of these signatures. An evaluation of this ability continues to be a major objective of this work.

The first steps in the first approach are: (1) estimating biomass and net primary production on tests sites in boreal forests; (2) using helicopter remote sensing to measure reflectance from these sites (using an eight band Barnes radiometer); (3) relating the remote sensing measurements to the ground measurements; and (4) using aircraft-mounted Thematic Mapper Simulator and Landsat satellite measurements to distinguish categories of vegetation in the boreal forest.

The second approach employs both manual interpretation and machine classification of Landsat and AVHRR data to stratify vegetation and other land covers into broad, physiognomic categories (based on vegetation structure) suitable for global comparisons. Aerial photographs, field reconnaissance and other sources will be used for accuracy verification. The derived maps will be used to examine the accuracy of existing small scale vegetation maps, upon which current estimates of global land cover are based. This approach provides both a comparison for current information sources, and an assessment of the methodology of very large area vegetation mapping. Preliminary results from test sites located in East Africa (Kenya and Tanzania) and North America (Minnesota and New Hampshire) show that Landsat imagery can be used to detect cover class boundaries at a scale appropriate for global land cover assessment, and to improve the accuracy of existing global estimates. However, because of the resources that would be required to process data for the entire land surface of the earth, an appropriate strategy would be to use coarser resolution data for primary stratification in a multistage sampling approach. Even using coarse-resolution data and a statistical sampling approach, assembling the required data is a very large task and could be greatly facilitated by a PLDS.

I.1.4 PROCESSING FLOW - DATA INPUT AND ANALYSIS

In support of this project and other research, an Information Sciences System (ISS) is being developed at JSC, which will provide computation and information management support for a highly dynamic and diverse set of processing and data requirements which result from biospheric research needs. A facility is being provided for processing, accessing, and exchanging research results. The integrated services of the Information Sciences System include the activities of data entry and preprocessing, image processing and display, data management (physical and electronic), and computer/laboratory operations management. In the context of this report, this system can be viewed in two ways: 1) it can serve as a model for certain components of the PLDS, such as the DBMS, and 2) it can be seen as a model for user nodes to which the PLDS would interface, thus enhancing the capabilities of both systems.

The Data Base Management System (DBMS) is capable of storing, correlating, and retrieving widely disparate types of data. The DBMS currently operated by Information Sciences is called ADABAS, an acronym for ADaptable DAta BAse System. Under ADABAS, key parameters for each data entry are made relational by storing data location pointers on a list in a separate file on the computer where quick access and correlation is possible.

The processing and management of data from each of the instrumentation and measurement systems involved in this project are outlined in the following paragraphs. A more complete description can be found in Wheelock (1983).

A. Sample Site Descriptive and Tree Dimension and Leaf Area Data - Data forms, with manually entered data, are shipped from the boreal forest test sites to JSC where they are keypunched, transferred to magnetic tape, and input to the Information Science System (ISS) computer as a disc file. Screening programs detect bad data. Listings of files are forwarded to field-workers for proof-reading against original forms, and correction. After corrections are made, files are placed in ADABAS. A copy of the data set locator information is placed in a data base directory file where any user may determine its existence as cross-referenced by data-type, location by site identification, state and county, latitude, longitude, satellite path and row, etc. Physical samples of leaves and pine needles are airshipped to the Laboratory for the Application of Remote Sensing (LARS) at Purdue University, where a subsample is packaged and airshipped to JSC. Once at JSC, the leaves are measured for reflectance, transmission, and needle surface area. These measurements are then loaded into a disc file on the ISS computer. Once again, and for all the following data sets, pertinent locator information is placed on file in the data base catalog/directory system. A 35-mm camera is used to record sampling procedures and photographs of canopy types. The film is shipped to JSC where it is processed into slides and stored in a physical data library where an extensive collection of maps, documents, photographs, etc., is retained.

B. Helicopter and Base Station - 70-mm color aerial photography taken over each test site is shipped to JSC for processing and duplication. The original film is held in archive, and the duplicate film, made available to users, is stored in the data library. Radiometric data from the helicopter and ground based multi-spectral radiometers are off-loaded from the instruments onto magnetic tape cassettes and shipped to JSC with a copy of the flight log and base station data collection report. Meteorological data are off-loaded from the portable environmental station onto cassette tape and transported to JSC. Cassettes of radiometric and meteorological data are fed through an electronic interface directly into disc files on the

ISS computer. Radiometric and meteorological data are then transferred to LARS electronically for calibration and copying.

The solar radiometer and ground pyronometer data are manually entered on data collection forms and output on a strip chart, respectively and shipped to LARS for coincident processing with the helicopter and ground radiometric data and the base station meteorological data. At the completion of LARS processing, the data set is shipped back to JSC and placed in a disc file.

C. Aircraft Data - The aircraft are based at the Ames Research Center.

Photography - Aerial film is processed and duplicated at the ARC Photographic Laboratory. A copy of the film and a flight summary report are forwarded to JSC for analysis and archive.

NS001 - Thematic Mapper Simulator (TMS) - TMS scanner data tapes, also downloaded at ARC, are preprocessed at the ARC computer laboratory where 1600-bpi Computer Compatible Tapes (CCTs) are generated. These tapes are shipped to JSC where they are immediately duplicated and the originals placed in archive. At this point, the TMS data are placed on a disk file, radiometric corrections applied, and a 3-band false color image tape prepared for filming. A single channel of data is also extracted and a black-and- white image tape is prepared. The image tapes are sent to the ISS film recorder where the false color and black and white images are recorded as transparencies.

D. Landsat-4/Thematic Mapper - Tapes received from Goddard Space Flight Center (GSFC), are immediately copied and the originals placed in archive. The copies are put in the Data Tape Library and made available to users. Locator information for these full-scene tapes is placed in the data base directory. Subscenes (512 x 512 pixels) extracted from the full frames are hard-copied, stored in an operations reference file, and cross-referenced to full-scene locator data.

E. Landsat-4/Multispectral Scanner (MSS) - After full-scene tapes arrive from the EROS Data Center, they are immediately copied and the originals archived with locator information published in the data base directory. As with TM data, a subscene image log and hard copy operations file are developed. When a user selects an area of interest (AOI) by specifying the latitude and longitude of the center point along with the image size, the AOI is extracted and sent to the ERSYS registration processor. ERSYS is used because the EROS tapes are fully corrected geometrically and radiometrically. A 4-band universal format CCT is output to the tape library and locator information is published. If a hard copy of the registered image is desired, a 3-band false color or single band black and white image tape is generated, sent to the film recorder, developed, and logged into the data library.

F. Advanced Very High Resolution Radiometer (AVHRR) - AVHRR full-frame data tapes from the NOAA Data Processing and Satellite Control Center are copied upon arrival and the originals placed in archive with locator information published. Geometric and radiometric corrections are applied to the full scenes prior to extracting subscenes or filing the corrected CCTs in the data library for the user community. An operations reference file of subscene polaroid prints is generated. CCTs of these subscenes are placed in the data tape library. (See Table II.1.1 for a summary of data needs for this project.)

I.1.5 CURRENT LIMITATIONS IN COMMUNICATIONS AND DATA FLOW

The data management and processing tasks for this project are very complex and involve diverse data sources and distributed investigators. This limits the rate and efficiency of analysis and magnitude of effort in several ways:

- The necessity of transferring sets between institutions for proof-reading, registration, etc. (especially site data of types A & B) has been a major limiting factor on the speed and efficiency of data analysis. At several stages, forms, listings, or tapes must be mailed between institutions and formats converted. Analysis of some data has been delayed by several months, impeding planning for further work. In addition, considerable human resources are consumed in essentially non-productive work.

- The size and completeness of this study and others like it are limited by the ability to access and calibrate large data sets. Ancillary data - topographic, meteorologic, historic, etc. - may be crucial in understanding patterns studied. Independent acquisition of such data is impractical, but existing data bases are very difficult or time-consuming to obtain. Ready access to (or even knowledge of) data from parallel studies in other areas could be very valuable for verification of generality of patterns.

- Discovering, obtaining, registering, and analyzing remotely-sensed data other than those gathered specifically for this project has been of such difficulty that valuable types of data may be unused due to lack of knowledge of their existence, or resources for making them usable.

I.1.6 SCENARIO FOR PLDS SUPPORT OF RESEARCH

Most of the functions proposed for a PLDS would aid this project in some way. Some of the most important areas of support might be in:

A. Data Input: Direct transmission of field data (both vegetation and radiometric) between field sites and processing centers (JSC, LARS, UCSB) could cut time to processing by an order of magnitude (from months to a few days). Entry or conversion of ancillary data

TABLE I.1.1
DATA SUMMARY: SCIENCE SCENARIO 1
VEGETATION BIOMASS, PRODUCTIVITY AND INVENTORY

AREA COVERAGE: Primary sites in Superior National Forest, MN and Konza Prarie Presenve, KS.

ECOLOGICAL REGION: Boreal forests, tall grass prairies.

OBJECTIVES: 1) Develop techniques for estimation of biomass and net primary productivity of natural vegetation using remotely sensed and collateral georeferenced data sets, 2) Develop large area inventories and estimates for these parameters.

INSTITUTIONS: JSC, U.C. Santa Barbara, Kansas State U., LARS-Purdue, SUNY-Binghamton, CUNY-Hunter College, Oregon State U.

DATA TYPE	QUANTITY	REPETITIVE COVERAGE	SOURCE
SATELLITE			
LANDSAT-MSS	IC: 3 scenes=$8x10^8$B	3/year	EROS
	LA: 40 scenes=$1.5x10^9$B	update 5-10 years	"
-TM	IC: 3 scenes=$8x10^8$B	3/year	"
	LA: 10% subsample of MSS =$1x10^{12}$B	update 5-10 years	"
NOAA-AVHRR	2 swaths, 3bands, 10 dates =$5x10^8$B		NOAA
AIRCRAFT			
TMS	IC: $1x10^5$B/quad	5/year	ARC
photos	IC: 10^3 frames	5/year	ARC, USFS
	LA: 10^3 frames	update 5 years	" "
AVIRIS	$2x10^8$B/pass	?	JPL
Helicopter Radiometer	100 sites	7/year	JSC
C-Band	100 sites	7/year	JSC
DIGITIZED MAPS			
Terrain	(20 m grid=$4x10^6$B/quad)	-	USGS
	IC: 10 quads, $4x10^7$B		
	LA: 10% of TM, $3x10^8$B		
Geology	(50 m grid=$6x10^5$B/quad)	-	USGS/SCS
	IC: 10 quads, $6x10^6$B		
	LA: 10% of TM, $5.5x10^7$B		
Fire history/ Land use	(50 m grid=$6x10^5$B/quad)	-	USGS/USFS
	IC: 10 quads, $6x10^6$B		
	LA: 10% of TM, $5.5x10^7$B		
FIELD			
Radiometer	100 plots	-	field, JSC, UCSB
Forest Dimension	100 plots, 10^5B	-	field, JSC, UCSB
Site Characteristics	100 plots, 10^5B	-	field, JSC, UCBS
Meteorological	100 plots	-	field, NOAA

KEY: IC: Intensive Coverage of Study Sites
LA: Large Area Inventory, entire boreal area of North America
B: Bytes
quad: 7½' quadrangle map; approximately 11x13.5 km

(topographic, soils, climatic) to acceptable format would add to the potential of the project.

B. Preprocessing: Registration (band-to-band and sensor-to-sensor) and common formatting of sensor data (from TM, MSS, AVHRR, scatterometer, radiometer, AVIRIS, etc.) would be of tremendous value and high priority. Efficiency of work would be vastly improved if this could accomplished within a few weeks of data acquisition. Also of value (but less important) would be the capacity to digitize photographs with interactive input from remote Principal Investigators (PIs).

C. Analysis: Analysis facilities are reasonably sufficient, but efficiency of analysis could be increased if real-time interaction between centers and remote investigators in the analysis process were possible.

D. Storage and cataloging: A directory, with documentation of parallel and ancillary data sets held within NASA and elsewhere, would be of great value and is of high priority.

E. Distribution and Networking: Access to data sets referred to in D, and ability to overlay them in common format is high priority. Data, besides being in compatible format, must carry documentation of quality and type. Time scale for such access should be on the order of a few days. Networking of CPU and availability of peripherals at NASA centers to provide access by remote investigators would be valuable.

I.2 SCIENCE SCENARIO 2: BIOGEOCHEMICAL CYCLING IN FORESTS; AN INTEGRATION OF REMOTE SENSING, MODELING, AND FIELD ANALYSIS

James Brass
NASA/Ames Research Center

I.2.1 PROJECT BACKGROUND

A new emphasis on earth science is developing within NASA that is global in scale and concerned with decadal time periods. At these scales, the atmosphere, biosphere, and hydrosphere act as an integrated system. New paradigms are required to address science issues at these scales. A crucial program component, and an essential element of the earth sciences, is biogeochemical cycling.

A program of research to develop a quantitative understanding of biological productivity and biogeochemical cycling of carbon, nitrogen, sulfur and phosphorus has been developed. Initially, this investigation will be limited to forested ecosystems since they account for approximately 90 percent of the world's net terrestrial primary productivity and a commensurate proportion of the relevant exchanges between the biosphere and atmosphere through the processes of biogeochemical cycling.

Development of an understanding of the cycling of carbon, nitrogen, phosphorus, and sulfur for terrestrial ecosystems will require globally aggregated models of the state and fluxes between compartments of these elements, as well as explicit treatment of nutrient dynamics using process level models. Remote sensing can play a significant role in these efforts if meaningful information related to nutrient processes can be extracted. Many of the existing models of ecosystem functions have been developed from site specific information. Remote sensing offers the possibility of accounting for spatial heterogeneity.

We recognize that key observable canopy variables may be useful in the characterization of nutrient cycling dynamics at a regional scale. These variables can be directly measured using a combination of unique laboratory studies coupled with high resolution spectrometry in the field and from airborne platforms. By using these remotely sensed canopy variables, a new class of canopy-driven process models which treat nutrient cycling explicity can be developed.

The overall goals of the coordinated effort are to:

1. Characterize biogeochemical cycling and biological productivity for the study sites (with the possibility of expansion to new sites and new cooperators), emphasizing interactions between the biosphere and atmosphere;

2. Determine techniques for collecting, combining and analyzing data that is descriptive of biogeochemical cycles and record these data in a standard format; and

3. Define the relationships between remotely sensed data, land surface characteristics and variables of reflective biogeochemical cycling processes.

In order to synthesize a research program devoted to biogeochemical cycling and biological productivity, a number of regional studies must be conducted to gain insight to a global approach. The intent of the current proposal is site intensive and complementary with on-going work at Johnson Space Center (Minnesota Boreal Forest test site) and Goddard Space Flight Center (Africa Grassland test site). It is anticipated that the three projects will provide the necessary insights into the global syntheses of models of biogeochemical cycles and biological productivity.

I.2.2 OBJECTIVES

There are three objectives in this research:

1. To develop semi-mechanistic "canopy-driven" models, based on remotely-sensed data, of total upper canopy nitrogen, phosphorus, and carbon, and driven by known surface meteorological and water relationships, to predict and characterize nutrient cycling, including photosynthesis, production, and decomposition processes.

2. To establish a basis for measuring upper canopy nitrogen and phosphorus, and the major compound distribution of carbon, using a combination of laboratory, in situ, and remotely sensed infrared spectroscopic techniques.

3. To develop the framework for regional, and eventually global, estimation of biogeochemical cycling and productivity.

In the past, ecological models of productivity have been built on traditional forestry methods emphasizing stem growth and population statistics. The development of biogeochemical models based on canopy variables, while potentially very attractive, has been limited by the difficulty in obtaining sufficient canopy data. By combining data from an advanced high resolution infrared spectrometer and other sensors (including the Fourier transform infrared interferometer, Airborne Imaging Spectrometer, TM, AVHRR, and radar), we believe the potential exists to measure canopy properties remotely.

I.2.3 APPROACH

The goal of the proposed research is to develop and understand the relationships between forest canopy characteristics and forest energy and nutrient dynamics (through both measurements and models), and then to apply remote sensing technologies to permit efficient large-scale measurements of canopy characteristics over a broad range of forests. In order to develop these relationships, it is necessary to examine as wide a spectrum of forest canopies as is practical, while holding down the number of sites (due to the expense of field work).

Rates of net primary production and nutrient pool sizes are generally well correlated with biome or latitude for sites with sufficient water. Boreal forests are the least productive and have the slowest decomposition rates, followed by cold temperate, warm temperate, and tropical forests. The patterns of nutrient cycling and the potentials for nutrient loss to the atmosphere or hydrosphere, however, are more closely associated with site fertility than with biome. Forest ecosystems on infertile sites differ widely in productivity, but their mineralization and immobilization dynamics and the resistance of their nutrient cycles to perturbation are generally distinctly different from those in more fertile, sites regardless of biome. Consequently, to evaluate relationships among canopy characteristics and ecosystem dynamics and to develop models for these relationships, data from sites selected to represent interactions of biome and soil fertility will be used.

Field data collection will occur in five major biomes, from the boreal forest of Alaska to tropical forests of Costa Rica. These sites have been intensively studied in the past, providing valuable historical data for the productivity/biogeochemical cycling project. Additional test sites in Wisconsin, Tennessee, and California will be used to examine the ecosystem variability between biomes.

The research is divided into two major tasks. The first task involves canopy sampling and measurement. This step will include the analysis of leaf distribution by tree height and species and leaf surface modeling again stratified by species into height. The second task, to develop correlations between the spectral characteristics of the upper canopy and the chemistry of the leaves, will demand the majority of the resources in this project. These correlations will provide the basis for canopy driven models especially suited to nutrient processes and driven by canopy variables that can be measured with remote sensing techniques.

The approach of this research is threefold:

1. Determine the correlation between the infrared reflectance spectra of leaves taken from the field (upper canopy) and from plants grown in the laboratory, and the chemical properties of these same leaves.

2. Use correlation analysis to determine the relationship between various leaf chemical measures, above-ground biomass, and rates of biodegradation.

3. Combine field data (five sites) with infrared reflectance spectroscopy and chemical analysis of upper canopy leaves.

Current co-investigators include:

1. Ames Research Center -- Peterson, Lawless, Whitten
2. Jet Propulsion Laboratory -- Rock
3. Florida State University -- White
4. Stanford University -- Vitousek

5. Oak Ridge National Laboratory -- Emmanuel, Johnson
6. University of Wisconsin -- Aber
7. University of Montana -- Running
8. National Park Service -- Parsons, Graber
9. University of California, Los Angeles -- Rundel

Future co-investigators:

1. University of California Santa Barbara -- Botkin, Estes
2. Johnson Space Center -- Pitts
3. University of California Davis -- Rolston
4. Oregon State University -- Schrumpf

The data requirements for this project are summarized in Tables I.2.1 and I.2.2. In general, the data layers will cover the entire spectrum, from very high resolution data (one meter or less) to low resolution data sets (many kilometers). Much of the data exists in a tabular format and will require encoding for automated analysis. Registration of data layers to a common base will be needed, necessitating a major effort. Detailed data requirements will include both historical and newly acquired ground data.

Data quantity from the intensive site studies will be small in areal extent, but will number in the thousands in terms of individual point measurements. Only when the analysis is expanded to the global scale, estimating the range of ecological conditions and providing input to atmospheric and terrestrial models, is it anticipated that large storage and processing systems will be necessary.

I.2.4 CURRENT LIMITATIONS IN COMMUNICATIONS AND DATA FLOW

Currently the funded project has nine co-investigators spread across the entire country. Data will be collected by most of the researchers, with analysis being done at each location. Therefore, communication and data transfer will be a critical issue in the course of this project. It is anticipated that an initial network will be established between ARC and Florida State, Stanford, Oak Ride National Laboratory, the Universities of Wisconsin and Montana, the National Park Service-Sequoia, and the Jet Propulsion Laboratory (Fig. I.2.1). Additional nodes for the network would include U.C. Santa Barbara, Oregon State University and Johnson Space Center.

Communication bottlenecks which the PLDS could assist in overcoming include:

1. Incompatibility of hardware and software among co-investigators for transferring text, data, and software.

2. Lack of capability to send large data sets between institutions via computer links, and to access remote computer resources for distributed processing.

3. Lack of accepted terminology for describing data sets, standards for data validation, and protection of proprietary data.

Table I.2.1

Sampling Parameters for Biogeochemical Cycling Study
(Science Scenario 2)

Parameters	Alaska	Wisconsin	S.E. U.S.A.	Sequoia	La Selva
Canopy area and mass	+				
Green foliage C,N,P, lignin conc.	+		+		
Litterfall mass, C,N,P, lignin conc	+	+	+	*	+
Aboveground net primary production	+	+	+	*	x
Aboveground biomass	+	+	+	*	x
Forest Floor mass, C,N,P, lignin conc.	+	+	+	*	+
Soil characteristics (bulk density, texture, % OM, etc.)	+	+	+	*	+
Ecosystem hydrologic paramters (throughfall, stemflow, leaching)	+	+	+		+
Decomposition rates		+	+	*	+
Nutrient turnover and availability		+			*
Microclimate paramters (air temp., soil temp, precipitation quantity and quality)	+		+	*	*

\+ Published data

* Research-in-progress

x Data available, but must be compiled

Table I.2.2 Summary of Data Needs -- Science Scenario 2

BIOGEOCHEMICAL CYCLING IN FORESTS

AREA COVERAGE: Five primary sites located in Alaska, Wisconsin, Tennessee, California, Costa Rica

ECOLOGICAL REGIONS: Subarctic, Warm Continental, Hot Continental, Mediterranean, Subtropical

OBJECTIVES: (1) Develop semi-mechanistic "canopy driven" models based on remotely sensed data of total canopy nitrogen, phosphorus and carbon quality; (2) establish a basis for measuring upper canopy N and P and the major compound distribution of carbon with remote sensing technologies, and (3) develop the framework for regional and eventually global estimation of biogeochemical cycles.

INSTITUTIONS: ARC, JPL, Florida State University, Stanford, Oak Ridge National Laboratory, University of Wisconsin, University of Montana, University of California-Los Angeles, National Park Service

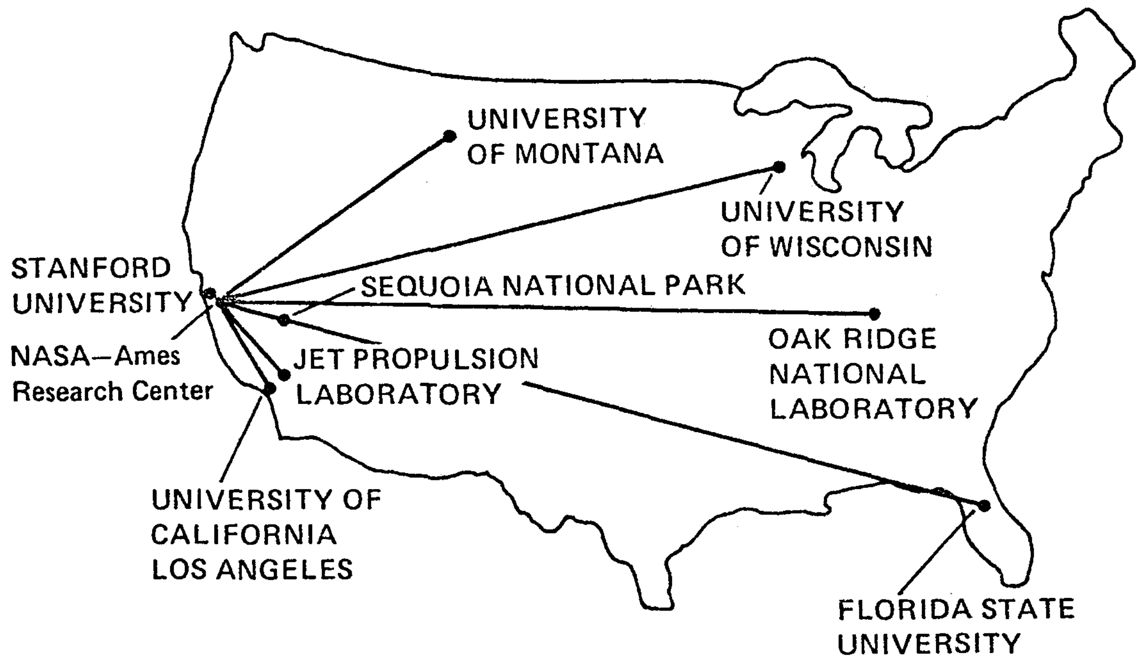
FUTURE COINVESTIGATORS: University of California-Santa Barbara, University of California-Davis, Oregon State University, Goddard Space Flight Center

	SATELLITE			AIRCRAFT		OTHER DIGITAL
	LANDSAT		NOAA 6, 7, 8			
Data Type	MSS	TM	AVHRR	AVIRIS	RADAR	DIGITAL TERRAIN
Quantity	15 scenes	15 scenes	3 scenes	15 strips	15 strips	5 7.5' quads
Repetitive Coverage	3/season	3/season	3/season	seasonal	seasonal	no
Location of Data Source	EROS Data Center		NOAA	JPL	JPL	USGS

	FIELD				
Data Type	IMAGE SPECTROMETER	DIMENSIONAL DATA	SOILS LITTER	LEAF CHEMISTRY	METEOROLOGICAL
Quantity	15,000-20,000	10,000 points	2,000 points	1,200 points	10,000 points
Repetitive Coverage	seasonal	TBD	seasonal	seasonal	seasonal
Location of Data Source	Field	Field/ Historical	Field/ Historical	Field/ Historical	Field/ NOAA

FIGURE I.2.1

LOCATIONS OF CO-INVESTIGATORS IN INTENSIVE SITE STUDIES IN BIOGEOCHEMICAL CYCLING PROJECT



4. Need for computer network to handle multi-users in communication, file/data transfer and processing modes, including update procedures for data and accounting procedure for "outside" users.

I.2.5 SCENARIO FOR PLDS SUPPORT OF RESEARCH

As stated earlier, it is anticipated that a network will be developed with at least nine nodes spread throughout the country. Rapid data transfers will be important due to the collection of ground data by many team members in different locales. It is imperative that data transfers be timely (overnight at most), and not tie-up system or network resources during normal processing times.

In most cases, this project will not require high speed data analysis but will require timely data capture and reasonable data encoding procedures. The capability to rapidly encode the field data following collection will be essential given the large data volume involved. Word processing and electronic mail facilities at each institution are a high priority to expedite project documentation. Both issues must be addressed in any network set up among the co-investigators. In addition, speed may be important -- a printer should not be tied up for many hours dumping a text file being sent from one location to another.

Additional communication and network requirements needed to support the processing and analysis in this project are:

1. Remote use by investigators of large main frame computers located at ARC;

2. Ability to send and receive data files from one node to another;

3. Ability for investigators to use software packages where they exist or transfer them to their local computing facility; and

4. Standardized data capture (encoding) procedures such that all facilities can use digitized data from all investigators.

I.3 SCIENCE SCENARIO 3: LAND-SURFACE CLIMATOLOGY

Robert D. Price and Elizabeth M. Middleton
Laboratory for Earth Sciences
NASA/Goddard Space Flight Center

I.3.1 SCIENTIFIC BACKGROUND

The land surface system interacts in a complex, dynamic manner with the atmospheric system through processes of energy, mass, and momentum exchange to produce weather, and long-term climate. A need exists to develop a better understanding of the nature and scope of influence of the land surface on weather and climate. Understanding the impact of land-surface changes on climatology will ultimately provide insight into the long-term impact of both land surface and atmospheric change on the habitability of this planet.

Processes occurring within the terrestrial system which influence climate involve the type and extent of vegetation cover, soil type and moisture content, topography and latitude, and nature of land use. For example, the atmosphere derives a large portion of its energy from reflected and emitted radiation over land, which depends upon vegetation type and soil type, through albedo and surface roughness. The atmosphere also derives a large portion of its moisture from evaporation and transpiration which depend upon soil moisture and vegetation cover. The spatial and temporal variations in these processes produce variations in the temperature, precipitation, and wind in the atmosphere. In turn, variations in the weather and climate produced by the atmosphere alter the distribution and state of the biosphere and processes in the hydrologic cycle.

Modeling provides one of the best methods for improving understanding of land-surface climatology processes and interactions. The development of terrestrial and climatological process models requires many types of data for parameterization of physical coefficients, for initialization of the modeling simulation runs, and for validation of modeling results. A land data system, which could be used to store, access, manipulate, and analyze data, would certainly facilitate, and in some cases enable, the study of land-surface climatology.

A new program, the International Satellite Land Surface Climatology Project (ISLSCP), is being conducted under the auspices of COSPAR and the International Association of Meteorology and Atmospheric Physics to support scientific investigations in land-surface climatology. The goal of this research project is to develop an understanding of the processes by which the atmosphere and land systems interact through the exchange of mass, energy, and momentum. The Goddard Space Flight Center will be participating heavily in this project, and has already received funds to conduct workshops to define specific pilot experiments to perform.

I.3.2 OBJECTIVES

A. General Scientific Objective

The overall scientific objective of the land-surface climatology project is to develop a better understanding of the processes occurring within, and the interactions among, the earth's biospheric, edaphic, hydrologic, and atmospheric systems, and to determine their role in influencing or governing climate over land surfaces. In order to accomplish this objective, the principal experimental efforts for ISLSCP will be organized and conducted in three parallel activities, each of which depends on the development of a supporting data base and data analysis system. The first major activity will be to conduct a retrospective analysis of existing remote sensing data for selected climatically representative study regions. The objective of this phase of research is to determine to what extent changes in the land surface (which influence climatology) can be determined and measured, and to assess the relative sensitivity of climate to various land processes. The second major activity will be to prepare and validate comprehensive global data sets derived from operational satellites on scales up to $(500\ km)^2$ so as to document the current state of the Earth's land surface for select parameters. The third major activity will be to conduct pilot experiments on specific regional or continental land masses to relate remotely sensed measurements to climatically-sensitive parameters and to validate or modify land-atmosphere interchange models for these study sites.

B. Specific Scientific Objectives

Biospheric, edaphic, hydrologic, and atmospheric system processes and their interrelationships will be thoroughly investigated for several study sites which represent different climatically sensitive regimes. These studies will focus on significant changes in the land surface cover which have occurred during the time frame 1972 - present.

1. Vegetation

Fluctuations in green leaf biomass (monthly, seasonally, and annually) will be related to precipitation and surface temperature on a continental scale, (e.g., in Africa). Biomass will also be related to ecological units (i.e. Holdridge Life Zones) and to processes such as desertification, deforestation, and habitat destruction.

2. Soils

Regional measurements of soil moisture will be related to remotely -sensed surface signatures to determine if a broad range of relative differences in surface moisture conditions can be delineated and monitored using remote sensors.

3. Hydrologic Cycle

The application of remote sensing techniques to provide better estimates of surface water extent and volume, evapotranspiration, precipitation, snow extent and volume, and soil moisture will be explored on a global/regional basis and used where proven feasible.

4. Near-Surface Atmosphere

The capability to remotely detect and quantify climatic variations in the land surface record and to relate such changes to climate process models will be assessed. Measurements will include multi-level sampling of the land surface for albedo, vegetation cover, surface roughness, insolation, ground temperature, precipitation, etc.

I.3.3 APPROACH

A. Technical Plans

The historical land remote sensing record (over approximately the last decade) will be examined to determine if in regions known to have experienced significant variations in their climate, these changes can be detected through land cover change detection. The approach will be to assemble land surface data obtained from land-observing satellites and meteorological satellites into a central data base or, at the very least, institute necessary technology and procedures to gain rapid, easy access to the data at their currently established data storage locations. The data will need to be preprocessed and inter-compared. They will also be compared to collateral data sets in the form of tabular meteorological records, digital topographic data, polygonal land cover and soil designations, and intensive point, area, and transect data from field measurements. This implies that selected ground reference data, such as soil and land use maps, will be integrated into the data base and a geocoded reference structure developed as a means for relating these data sets. Goddard has already begun to assemble satellite (NOAA AVHRR and Landsat MSS) and ground data from Africa and South America.

While analysis of retrospective data will provide some indication of the influence of past land surface changes on climatology, the present physical and biological state of the entire land surface of the Earth is also inadequately described. The best sources of such information on a global scale are data sets derived from operational satellites (e.g., NOAA AVHRR). Such data sets will be assembled into a global data base. Work has begun to compare seasonally-resolved global estimates of green leaf biomass to atmospheric carbon dioxide from 1982 and 1983. Work has also begun on comparing AVHRR-based continental land-cover classifications to a digital $1^{\circ} \times 1^{\circ}$ land cover and land-use data base developed at the Goddard Institute of Space Studies by other means.

Lastly, specific study sites ranging in size between $(10\text{-}500\ km)^2$ will be selected to represent different climatic regimes, and a data base assembled. Pilot experiments which entail collection of field data and remotely-sensed data from several altitudes and sensors will be conducted. These multi-stage coordinated experiments will be designed to determine if specific processes of importance to land-surface climatology for those regions known to exhibit characteristic climatic patterns can be detected through remote sensing of earth surface features such as regional vegetation and soil moisture. This third activity will be supported specifically by collection of data by the experimental Shuttle Multispectral Linear Array (MLA) instrument (all bands) and numerous other data from low-altitude aircraft and field sampling (yet to be defined) aimed at measurements of heat and water fluxes and vegetation condition. Remote sensing measurements will be related to climatically-significant physical parameters where possible. Land-atmosphere process models that represent the exchange of mass, energy, and momentum between the land and atmosphere systems will be developed. Detailed data sets acquired over these specific study sites will be used to either initialize and/or parameterize the models in simulation runs, and/or to verify and validate the results of the models. The response of the land surface in terms of biomass productivity and water budget will all be modeled given the forcing of the climate (precipitation, insolation) and system properties (vegetation cover, surface roughness).

B. Data Needs

Remote sensing data to be examined include visible, near-,short wave-, and thermal-infrared, and microwave radiances at spatial resolutions which range from 30m to 30 km, and temporal resolutions from < 1 day to 18 days. Specifically, the following satellite sensor data have already been identified as having potential utility in this activity: (1) all bands of Landsat MSS and TM instruments; (2) all bands of the MLA instrument; (3) all bands of the TIROS-N instrument; (5) the 37 GHz and 19.4 GHz bands of the NIMBUS 5 and 6 ESMR; (6) the 37, 21, 18, 10.7, and 6.6 GHz bands of the NIMBUS-7 SMMR; and all bands of the NOAA 6 and 7 instruments; and the visible and thermal bands of HCMM.

In addition to satellite data, observations from a variety of spectrometers, radiometers, and cameras, flown on low-, medium-, and high-altitude aircraft will be acquired over selected study sites. Ground measurements of vegetation and soils will also be collected at selected study sites. These measurements include physical sampling of vegetation biomass and soil moisture for laboratory analysis. Another ground measurement includes reflectance in the visible and infrared portions of the spectrum using hand-held and truck-mounted radiometers. Standard meteorological parameters, such as temperature and relative humidity, will be gathered from meteorological stations scattered throughout the study sites.

C. Participating Institutions

Institutions that have been identified for participation in this science project fall into two categories: those that will conduct scientific research, and those that will provide data. The number of scientific investigators involved in the research will increase from about six in FY 85, to perhaps as many as fifty in subsequent years. These research participants will be geographically located at a total of perhaps fifteen domestic and foreign government agencies, universities, and research institutes with the facilities and expertise required for digital processing of remotely-sensed satellite data. Perhaps as many as ten national and international institutions will be involved in providing data, including satellite, aircraft, and field measurements. For rapid, easy access to the data and scientific results, it is expected that, at a minimum, the U.S. institutions should be connected electronically to a data system.

A list of projected participants and data needs in the ISLSCP project are summarized in Table I.3.1.

I.3.4 CURRENT LIMITATIONS IN COMMUNICATIONS AND DATA FLOW

Data of many types and formats will be used in the Land Surface Climatology Project. The needs of the project include locating and retrieving appropriate data, making it available to co-investigators, and transferring it to the computer system appropriate for each data processing step. These steps involve reformatting, preprocessing, processing and information extraction, and developing or verifying process models. Technological support of development is needed in the following areas:

1) catalog of and access to appropriate data;

2) common or universal data standards and formats;

3) storage and on-line processing of large data volumes associated with multi-spectral/multi-dimensional or global data sets;

4) automated preprocessing of image data, including removal of systematic radiometric and geometric distortions and biases;

5) automated preprocessing of image data to obtain georeferenced or map-registered products;

6) preprocessing of array or matrix non-image data typically acquired from field or laboratory, to produce georeferenced data for comparison with image data;

7) reduction of user costs for computer usage for preprocessing, data reduction, and analysis of large data sets;

8) electronic and physical management of all data involved;

9) development of a geobased Land Data Management System; and

Table I.3.1. GSFC Land Surface Climatology Scenario Summary

Area Coverage: 3-5 primary sites located in the U.S. Great Plans, North and Central Africa, Northern South America, and Australia.

Ecological Regions: Tropical rainforest, semi-arid, grassland, savannah

Objectives: 1) correlation of green leaf biomass and albedo estimates from remote measurements of vegetation to precipitation and surface temperature on a regional/ continental scale; 2) the capability to remotely detect and quantify climatic variations through land cover change; 3) determination of regional differences in surface soil moisture conditions with remote sensors.

Institutions, Scientific Collaboration: GSFC, JPL, NSTL, AID, NOAA, USDA, USGS, Oak Ridge National Laboratory, Clark University, Oklahoma State University, SUNY-Binghamton, Texas A&M, University of Kansas, University of Maryland, African Regional Commission, United Nations/FAO, UNEP.

Institutions, Data Interactions: GSFC, JPL, NSTL, Defense Mapping Agency, NOAA, National Weather Service, USDA, USGS, Soil Conservation Service, Oklahoma State University, United Nations/FAO.

DATA EXPECTED FOR EACH STUDY SITE:

DATA TYPE	QUANTITY (N=#data points)	SOURCE
	(since 1972, as available)	
Satellite		
Landsat MSS	1 scene/season	EROS Data Center
Landsat TM	1 scene/season	EROS Data Center
NOAA 6,7,8 AVHRR	1 scene/season	NOAA
NIMBUS 5.6.7 ESMR	1 scene/season	NOAA
NIMBUS 5,6,7 SMMR	1 scene/season	NOAA
SPOT	1 scene/year	Spot Image Corp.
Aircraft	(since 1972, as available)	
TMS	1 strip/season	NSTL
AVIRIS	1-3 strips/season	JPL
LAPR-2	1-3 strips/season	NSTL
Photography	1 set/season	NSTL,JPL
Field		
Photography; Site Descriptions	1 set each/season(N≈1000)	Science Participants
Biomass Samples	1 set each/season(N≈1000)	Science Participants
Vegetation Transects	1/yr.(N≈1000)	Science Participants
Evapotranspiration	daily for 1 mo. (N≈100)	Natl. Weather Service
Meteorological data	daily for 1 mo. (N≈100)	Natl. Weather Service
Hydrologic samples	daily for 1 mo. (N≈1000)	Natl. Weather Service
Micro-meteorological data	daily for 1 mo. (N≈1000)	Natl. Weather Service, USDA
Radiometer measurements	daily for 1 mo. (N≈10,000)	Science Participants
Soil samples	1 set (N≈1000)	Soil Conservation Service
Other Digital or Map Data		
Geology, Land Cover, Terrain	1 set each	USGS,USDA,Defense Mapping Agency
Soils	(N≈10,000)	Soil Conservation Service

10) establishment of communication/data transmission links among the various co-investigators.

These technology bottlenecks could be alleviated through the technology support and development sponsored by the PLDS.

I.3.5 SCENARIO FOR PLDS SUPPORT OF RESEARCH

The proposed configuration for this project is diagrammed in Figure I.3.1. A major goal of the PLDS matches a major requirement of the Land Surface Climatology scenario. That goal is to provide access to both remote data bases and processing services to accomplish the following tasks: data storage and archiving, data preparation (reformatting, registration, etc.), geographic overlay of dissimilar data (e.g., geographic information system development), and data base management. A secondary, long-term goal is to provide these services quickly and in a straightforward and easy-to-use manner. Together, these capabilities would allow scientists to focus on the research, and not waste effort on the details of accessing and preparing the data for use.

The following are steps in the development of the PLDS which would support these goals:

- Provide on-line cataloging, documentation, and "help" facilities to allow the researcher to locate data bases and available processing systems, both hardware and software.

- Remove whatever political, bureaucratic, or other non-technical roadblocks exist to accessing remote data bases or in the remote processing of them.

- Provide communications links to expedite the transfering of data bases to and from the data archive, the processing system(s), and the researcher.

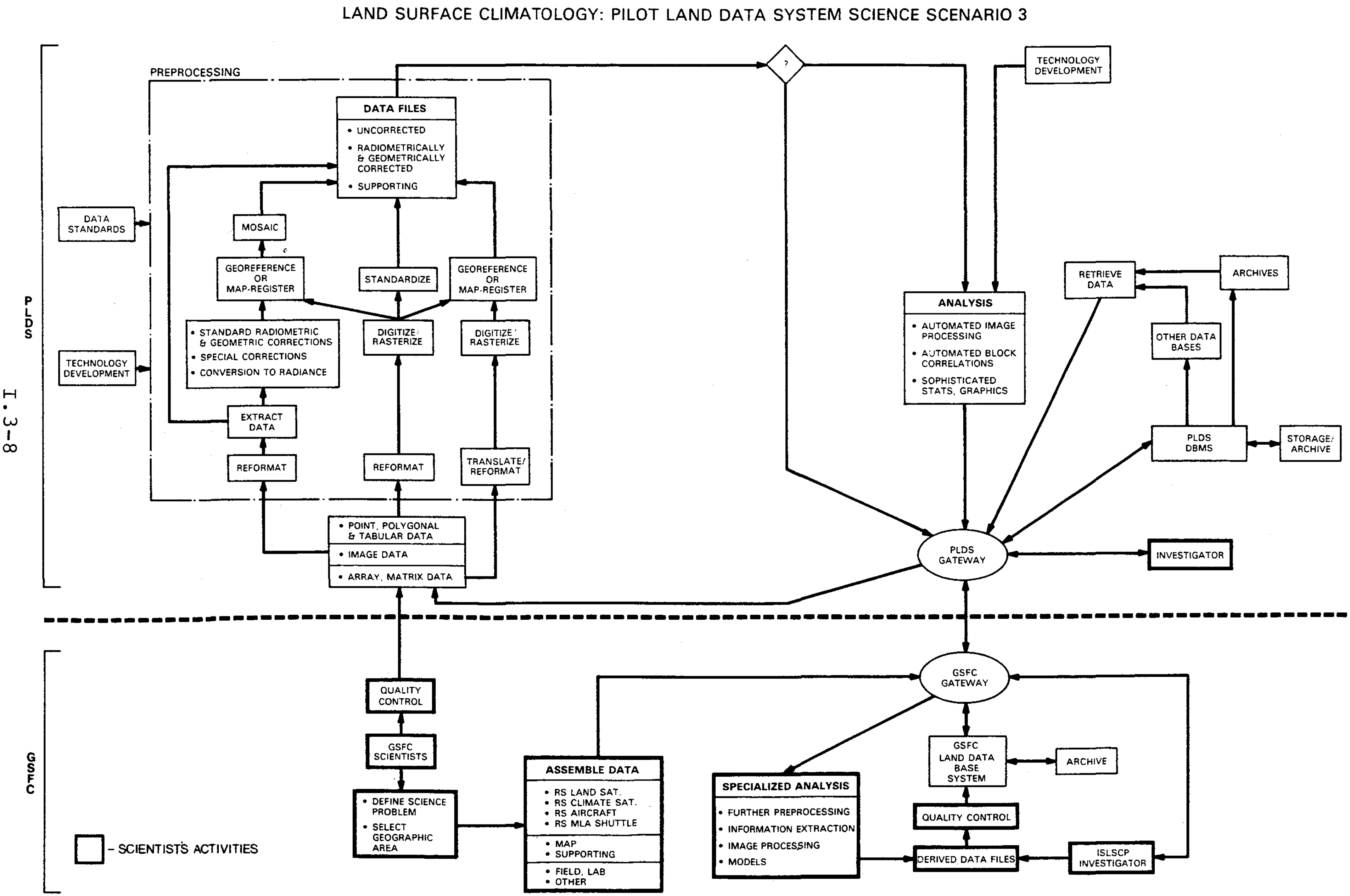
- Establish data formats and standards to reduce the task of reformatting data for different processing environments.

- Provide enhanced access to data bases developed by non-NASA agencies.

- Develop an intelligent user interface to reduce the need for researchers to concern themselves with the locations, formats, and system-specific details of remote processing environments and data bases.

- Establish or provide a communication and data transfer capability among GSFC and its co-investigators at universities and other facilities.

Figure I.3.1 Data and Analysis Network for Science Scenario 3



I.3.6 SUMMARY

The International Satellite Land Surface Climatology Project (ISLSCP), an international scientific effort to study climate and land interactions, has been initiated. This program will undertake complex interdisciplinary research sponsored over the next decade and beyond. It will require substantial technological support for all aspects of the data and information access, transfer, management, preprocessing, processing and analysis. The scientific goals of the program cannot be easily realized without the development of a formal computer-based structure to facilitate these requirements. Therefore, the development of PLDS is of crucial importance to the success of ISLSCP.

I.4 SCIENCE SCENARIO 4: HYDROLOGIC MODELING AND SOIL EROSION/PRODUCTIVITY MODELING

R. Ragan
University of Maryland

K. Langran
NASA/NSTL

I.4.1. PROJECT BACKGROUND

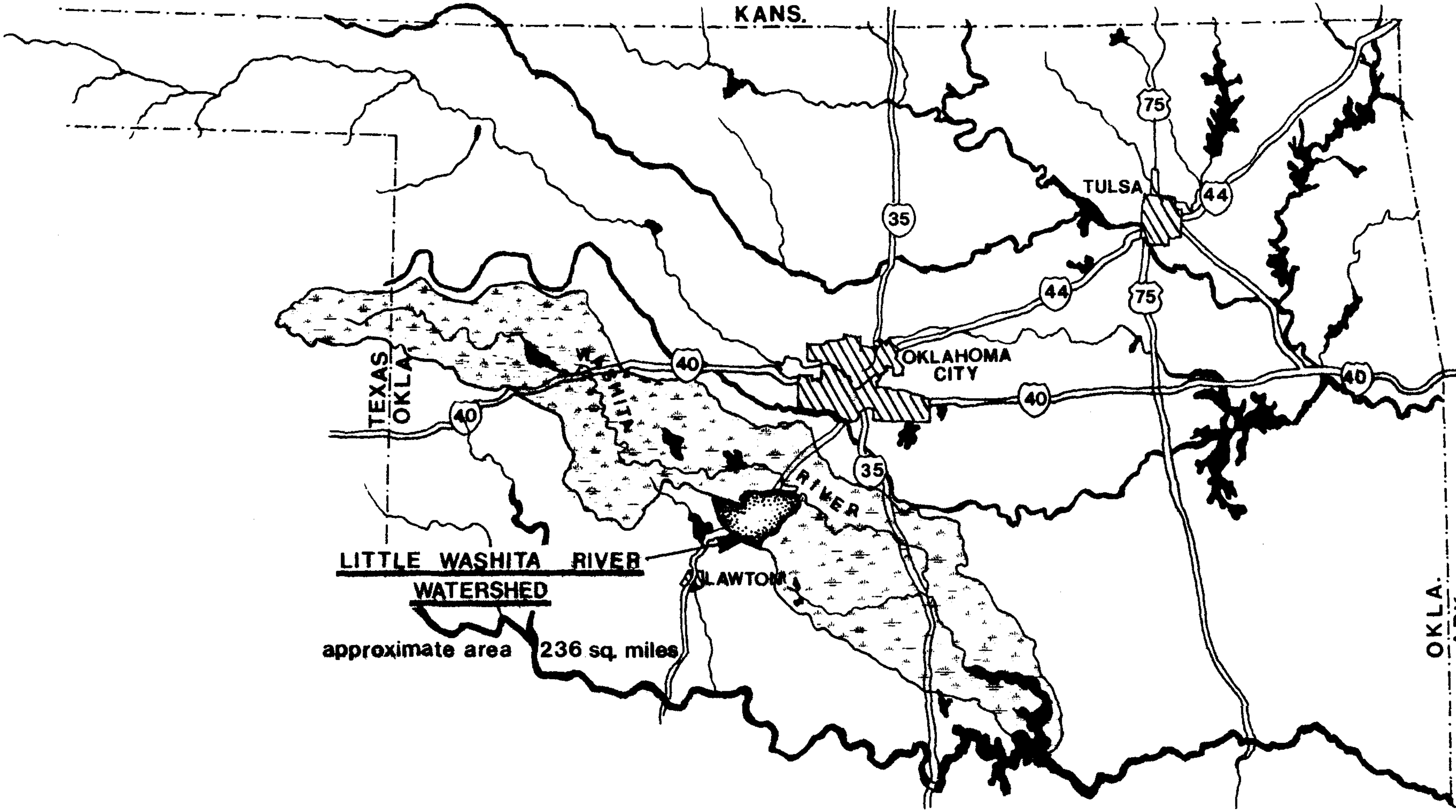
Serious gaps in scientific knowledge continue to limit the quality and efficiency of models which measure the relationships of hydrology, soil erosion and sediment yields, and the effects of erosion on soil productivity. Inaccuracies in the results of these modeling tasks frequently lead to incorrect policy decisions that produce significant personal and economic losses on an annual basis. Remote sensing has created new sources and types of data, and recent developments in computer and communications technologies provide opportunities for scientists to translate data from multiple sources into hydrologic and soil erosion/productivity information not previously available. This new information has led to the development of a new family of simulation models that, potentially, can better meet our forecasting objectives and provide the improved research capability necessary for a better understanding of the basic processes.

These models have been developed and tested with historic data. However, many questions concerning the utility of the models and the scientific validity of some of their formulations have not been examined because it has been impossible to obtain and interface critical data elements. The existence of a PLDS to interface an extensive data set with a distributed scientific community is the only mechanism that can allow a comprehensive evaluation of this new generation of remote sensing centered hydrologic and soil erosion/ productivity models.

There are a number of hydrologic and soils-related projects being conducted by NASA, USDA and universities in the Little Washita River Basin, Oklahoma (Fig. I.4.1). There is an excellent historical data base and a well-designed data collection system is in operation. Missing is an efficient means of distributing the data among all of the users, operational software to merge multiple data planes in order to derive critical information, a system that will allow the interfacing of NOAA data bases for the regions surrounding the Little Washita, and a mechanism to obtain, interpret and distribute digital format satellite imagery.

Under the AgRISTARS Land Resource Program, NASA-NSTL is presently conducting two research projects (Conservation Practices Inventory, Soil Erosion Modeling) in the Little Washita Basin. Various digital image processing techniques are being used with MSS, TMS, TM and SIR-A data in an attempt to highlight the selected conservation practices. Multi-temporal data are being evaluated to determine which physiographic and biological conditions (e.g., weather, dormant versus

FIGURE I.4.1 STUDY AREA FOR SCIENCE SCENARIO 4



WASHITA RIVER WATESHED, OKLAHOMA

NASA
ERL/NSTL

active vegetative growth, and in some cases, socioeconomic conditions), are significant in identifying conservation practices and needs. In the second project, the Soil Erosion Modeling task, Landsat MSS and TM/TMS data are combined with digitized soils and topographic data in the Universal Soil Loss Equation (USLE) framework to build a soil erosion data base. Statistical analysis techniques are used to determine the degree of correlation between remotely sensed and field generated data sets at selected training sites.

The data required to define the parameters used in modern hydrologic and soil erosion/productivity models are extremely difficult to collect in the field and then make available for access by the scientific community. The result is that scientific investigations in these disciplines require a disproportionate amount of time and effort to simply get and verify data with little time available for analysis. The scientific community will use the PLDS to address a series of extremely important science issues and, of equal importance, learn how to conduct research in an arena that provides real time access to both adequate data and powerful remote computer based systems.

I.4.2 OBJECTIVES

As stated above, there is the absence of an efficient means to obtain, interpret, and distribute remotely sensed and ancillary data, as well as a need to interface different data bases among users. The PLDS could greatly overcome these problems and allow the government and university scientists involved in the ongoing projects to expand their efforts to meet the following scientific objectives:

Hydrologic Modeling:

1. Test advanced remote sensing based hydrologic models to assess their utility and the scientific validity of their individual components as future tools for real time streamflow forecasting.

2. Improve the scientific community's understanding of the role of individual hydrologic processes on the overall behavior of large river systems.

3. Provide the scientific knowledge needed to allow the future use of space platform sensor systems in estimating evapotranspiration fluxes in terms of biomass, vegetative type, and sensible parameters.

Soil Erosion/Productivity Modeling:

1. Develop and test techniques for correlating model-simulated data (described below) with remotely sensed data. This includes developing a capability for combining remote sensing and <u>in situ</u> measurements for providing a realistic definition of the temporal and spatial distribution of precipitation, soil moisture, and rates of evapotranspiration as they relate to potential soil erosion and sediment yields.

2. Develop more comprehensive models that can measure erosion and sediment yields, and quantify the effects of erosion on soil productivity.

I.4.3 APPROACH

Hydrologic Modeling:

The existing computer-limited geographic system used on the Little Washita by USDA will be upgraded both in operational capability and by the inclusion of the critical data sets outlined in the subsequent sections. The NASA/USDA remote sensing based hydrologic models will then be applied to the Little Washita as a continuous streamflow generator for the 1972-79 period. This application of the models will provide for both calibration to the specific climatological and physiographical conditions of the Little Washita. Further, the model will allow experimentation to determine the sensitivities of the various components with respect to their role in the hydrologic processes, and to determine the accuracy requirements of future generations of sensors that will be used to support continental and global scale hydrologic analyses. The NASA/USDA models will then be operated on more recent and current data to explore the scientific communications constraints of real time forecasting operations.

One of the anticipated achievements of the Little Washita project, as supported by the PLDS, will be that scientists involved in the original development of the models and the needed computer systems at various locations around the United States will be brought together with an efficiency that has never been possible. Further, these scientists will not be encumbered by a lack of computer capabilities at their present locations.

Soil Erosion/Productivity Modeling:

The need to estimate soil erosion losses in conjunction with non-point pollution control and future soil productivity has become essential in agricultural decision-making and conservation planning. To understand the basic principles and processes of water erosion and sedimentation, and the effects of erosion on soil productivity under specified land use and management practices, new models have been developed over the last two decades which are applicable to a variety of watershed conditions. The Universal Soil Loss Equation (USLE), which has been the most widely applied erosion model since the late 1960's, is an example of a simple mathematical model designed to predict long-term soil losses from sheet and rill erosion on given field slopes under specified land use and management practices.

The SWRRB model (Simulator for Water Resources in Rural Basins), which has been implemented on the Little Washita River Basin by the USDA-ARS Southern Plains Watershed and Water Quality Laboratory, was designed to predict the effect of management decisions on water and sediment yields with reasonable accuracy for ungauged rural basins throughout the United States. The major processes included in the model are surface runoff, percolation, return flow, evapotranspiration, pond and reservoir storage, and sedimentation.

Soil productivity is the capacity of a soil, in its normal environment, to produce a particular plant or sequence of plants under a specified management system. Maintenance of soil productivity depends upon management practices as well as soil and site characteristics, the major ones being soil rooting depth, topsoil thickness, available water capacity, plant nutrient storage, surface runoff, soil tilth, and soil organic matter content. Soil erosion depletes soil productivity, and a major research effort is presently underway to quantify the relationships between plant growth and those soil attributes affected by soil erosion.

The EPIC (Erosion-Productivity Impact Calculator) model was developed recently by the USDA Agricultural Research Service to determine the relationship between soil erosion and soil productivity in the United States. The ability to combine remote sensing and in situ measurements for providing a realistic definition of the temporal and spatial distribution of the hydrologic components as they relate to potential soil erosion and sediment yields is one of the major research objectives of this scenario.

The application of remote sensing techniques to soil erosion studies, especially within the framework of the USLE, is well documented; but, almost no effort has been made to apply remote sensing techniques to the more comprehensive soil erosion/productivity models. In the initial study, or laboratory phase, the SWRRB model with archival data would simulate hydrologic and erosional conditions in the Little Washita Basin during selected time periods. Sensitivity analysis would then be performed to determine the degree of correlation between the model outputs and corresponding remotely sensed data. For example, remote sensing can estimate land cover related hydrologic parameters, or directly measure soil moisture, snow depth or water equivalent, sediment load in a water body, precipitation, and watershed characteristics. Similarly, indirect measures, or indicators, can be used to determine hydrologic parameter values (e.g., plant conditions), as an indicator of soil moisture levels.

Based on the results from the laboratory study, field studies will be implemented using assigned training sites, ground monitored hydrologic and erosion data, and remotely sensed data to determine the correlation between model outputs and corresponding remotely sensed data. An evaluation will be made of present sensor capabilities with possible recommendation for new devices for the overall monitoring of hydrologic and soil erosion/sedimentation model parameters.

I.4.4 CURRENT OPERATING PROCEDURE

The principal and potential investigators for this scenario are listed below. A primary investigator is a center which is directly involved in the initial research and/or has a data base that is part of the implementation of the research. A potential investigator is a center that has either a data base or could implement some aspect of the research plan in the latter phase of the scenario.

INSTITUTIONS:

Principal	Potential
NASA-NSTL	NASA-Goddard
USGS	USDA/SCS
USDA Agricultural Research Service	NASA-JSC
University of Maryland	U.S. Hydrology Laboratory
Oklahoma State University	National Weather Service
University of Missouri	National Parks Service
	Okla. State Climatological Survey
	Okla. State Geologic Survey
	Okla. State Archaeologic Survey
	University of Oklahoma
	Texas A&M
	University of Kansas

DATA SETS REQUIRED:

Historic data to test model:

- -TM two dates for study area
- -TMS two dates
- -MSS six (historical - 4 scenes/year; current 4 scenes/year)
- -TIMS two dates
- -SIR-A one date
- -SLAR one date
- -GOES dates to be selected for specific research experiments
- -NCIC digital terrain data, (DMA) 1:250,000 quad, (DEM) 1:24,000 quad

-Ground verification data, e.g., land use/land cover, two dates for thematic accuracy verification

-Historical meteorologic, hydrologic and soils data collected in the field or from ground stations, e.g., streamflow, water quality, temperature, humidity, wind

-Daily precipitation and temperature (tabular) data, evapotranspiration rate maps, and ground water distribution maps.

-Lithological, structural, geochemical and geophysical (tabular and cartographic) data.

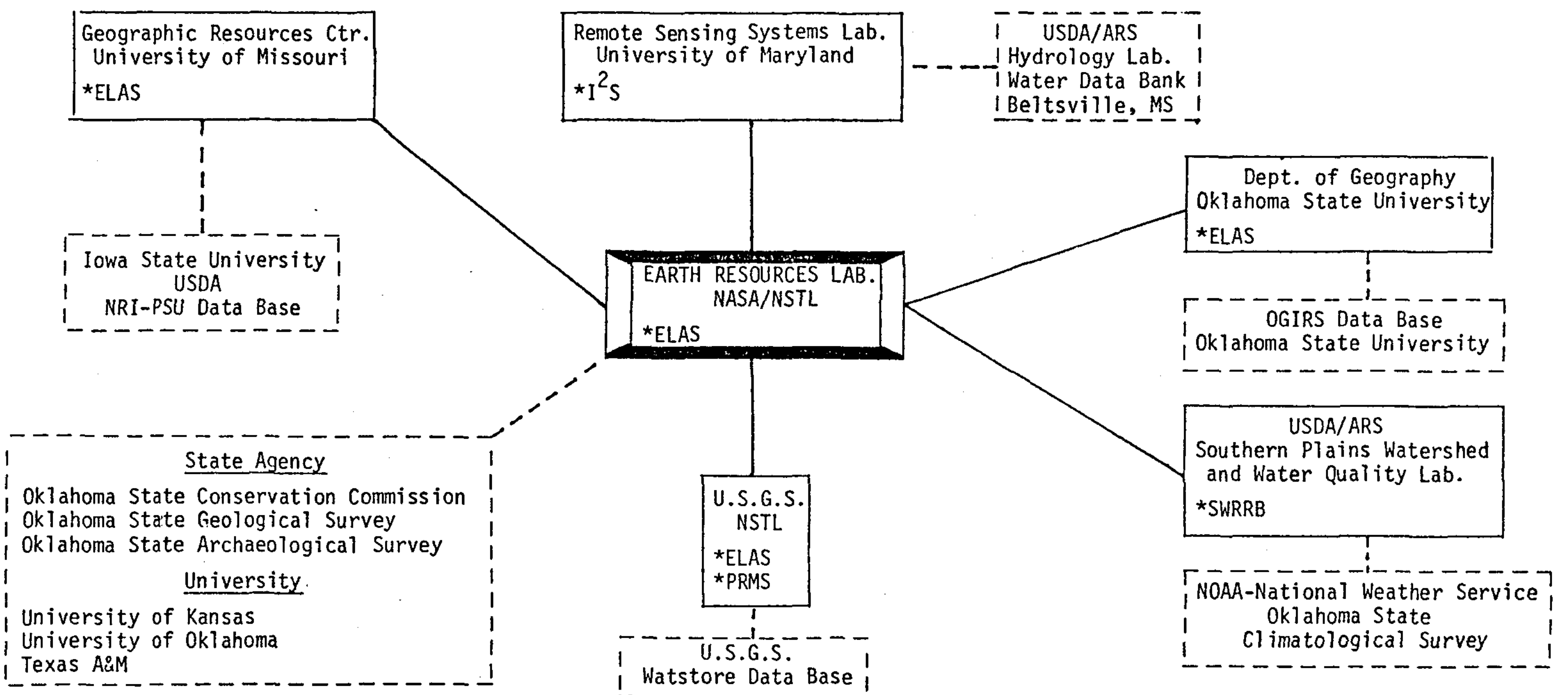
I.4.5 COMMUNICATION BOTTLENECKS

There is a need to establish workstations at each principal investigator (PI) center which will provide easy access to data base sources and the subsequent processing of different forms of data. There is also need to develop appropriate software interfacing between application software systems (e.g., ELAS and I^2S) and the numerous data bases in the PLDS network (See Fig. I.4.2).

Figure I.4.2 Data Network for Science Scenario 4

<u>Pilot Land Data System Scenario Number 4</u>

Soil Erosion/Productivity Modeling - NASA/NSTL
Test of Real Time Remote Sensing Based Hydrologic Modeling - University of Maryland



———— Principal Investigators
- - - - - Ancillary Data Sources
*Application Software System

ELAS - Earth Resources Laboratory Applications Software
PRMS - Precipitation, Runoff Modeling System
NRI-PSU - National Resource Inventory/Primary Sample Unit
SWRRB - Simulator for Water Resources in Rural Basins
OGIRS - Oklahoma Geographic Information Retrieval System

I.4.6 SCENARIO FOR PLDS SUPPORT OF RESEARCH

A. Initial Stage of PLDS Implementation:

1. The hydrology/soil erosion modeling science scenario includes a network of six principal investigators (PI) linked to nine data base sources. (See Fig. I.4.2)

2. Data encoding and formatting of nine data bases is required. This would include the developing of a common format and the linking of the ELAS and I^2S application software systems via that format.

3. Preprocessing - Most of the PI centers have preprocessing capabilities. However, in the calibrating of data sets, there is a need to measure the quality of data, and then provide subsequent enhancement procedures when necessary.

4. Communications Networks - Establish workstations at each PI center and implement a network among centers which will provide easy access to data sources and the subsequent processing of the data.

5. There is a need to access catalog data among centers. There is also a need to provide additional storage facilities and support activities at each PI center.

6. Processing - Most of the data processing will be done in a geographic information system (GIS) environment (overlaying of multiple data sets); therefore, there is a need for an efficient distribution network among PI centers and the various data bases in the PLDS network. Data registration capabilities would include: raster to polygon, polygon to polygon, and polygon to raster.

B. Final Phase of PLDS Implementation:

1. Complete system access to nine remaining source data bases.

2. Develop and implement a job task sharing system among PI centers, the purpose of which is to improve scientific productivity by having centers share sequential and/or parallel data processing tasks and analysis. For example, one center may first preprocess a remote sensing data set and then send it to another center for georeferencing, who in turn would send the output to a third center for classification and enhancement. Another possibility would be for centers to specialize in one phase of the research process, yet share in the final output.

I.4.7 SUMMARY

Accurate estimates of soil erosion and its effects on soil productivity are essential in agricultural decision making and planning from the river basin to the national level. The existence of a PLDS interfacing extensive data sets within a distributed scientific community is the only viable mechanism that will allow a comprehensive evaluation of a new generation of remote sensing centered models.

The breadth of the user community, the variety of the data sets and the distribution requirements make this science scenario an excellent case study for examining the type of problems that will be involved as NASA progresses toward global scale information systems capabilities. Solving the infrastructure, technical, and user need problems that will be encountered on this relatively small area will provide the experience base that is absolutely critical if NASA is to be successful with its global scale strategy.

I.5 SCIENCE SCENARIO 5 - MULTISPECTRAL ANALYSIS OF SEDIMENTARY BASINS

Tom Farr
Jet Propulsion Laboratory

Robert Singer
University of Hawaii

I.5.1 BACKGROUND

Instruments and techniques for analysis of remote sensing data have improved over the last few years, but there have been few concerted efforts to apply the variety of new techniques to a single geologic problem. The Multispectral Analysis of Sedimentary Basins project is an outgrowth of the GEOSAT project, in which a few test sites were studied with a variety of remote sensing systems and techniques to assess the potentials of geologic remote sensing. The basins project is designed to use new techniques for analysis of remotely sensed data obtained by a variety of sensors at many wavelengths for geologic analysis of a major sedimentary basin.

Sedimentary basins are large structures (>100X100 km) that occur throughout the world and that often contain economically significant amounts of oil, gas, coal, and other resources. In addition, sedimentary basins provide a record of the depositional and tectonic history of an area. The keys to efficient exploitation of the nonrenewable resources of a sedimentary basin are a knowledge of the distribution of geologic units both at the surface and within the basin, and an understanding of the evolution of the basin through time.

I.5.2 OBJECTIVES

The objectives of this project are to:

a) evaluate the utility of remote sensing data for mapping subtle variations in sedimentary lithology;

b) apply remotely sensed data to geologic mapping of a large sedimentary basin (Wind River Basin, Wyoming);

c) compare remote sensing data to conventional field mapping data;

d) combine remote sensing data of surface properties with geophysical data of subsurface properties to generate a three-dimensional representation of a basin; and

e) employ findings to constrain models of basin formation and evolution.

I.5.3 APPROACH

a) Acquire orbital and aircraft remote sensing data, geophysical field and seismic data, and field and laboratory spectral data from a large sedimentary basin (Wind River Basin).

b) Register the data to digital topographic maps and calibrate the data.

c) Obtain seismic, gravity, and magnetic data, and integrate them with the registered remote sensing and topographic data.

d) Define spectral and textural characteristics of sedimentary rocks in the test area.

e) Use all available information to map the surficial distribution of rock and soil types, and develop advanced techniques for efficiently analyzing this multidimensional data set.

f) Extend the mapped surficial distribution of rock units into the third dimension through correlation of surface and subsurface data.

g) Develop a scientific rationale for collection and analysis of remote sensing data of sedimentary basins.

Institutions and Funding

JPL: Mr. Ronald Blom, Dr. James Conel, Dr. Diane Evans, and Dr. Harold Lang

University of Hawaii: Dr. Robert Singer

Funding through NASA HQ: Nonrenewable Resources Program

Data Types

a) Regular array: images, digital topography, processed reflection seismic data. Image data are obtained by many different sensors with different radiometric and geometric characteristics, resolutions, and pixel sizes. Digital topographic data presently exist at two basic pixel sizes: 60 m and 30 m.

b) Polygon (vector): contoured data, digital geologic and cultural maps. Contoured data may include processed gravity and magnetic data.

c) Irregular array: gravity, magnetic, geochemical (e.g., NURE). Usually gravity and magnetic are obtained in an irregular sampling distribution. This creates unique problems for the correlation of these data with regular array data.

d) Line: aircraft scatterometer, Shuttle Multispectral Infrared Radiometer (SMIRR), seismic. These data are obtained by a regular sampling along a line. Usually photographs are obtained that allow registration of the data to a map base.

e) Tabular: laboratory and field reflectance spectra.

Laboratory data are obtained on pure reference samples, as well as samples returned from the field, and form a library of reference spectra. Field spectra are obtained at points that can be related back to a map base, and are used as ground truth for remote sensing data.

f) Maps: "analog" geologic maps.

Data Quantities

a) Regular array:

- 1 TM scene (7 frames X 5972X5972 pixels)
- 3 MSS scenes (4 frames X 3000X3000 pixels each)
- 2 Seasat scenes (6000X6000 pixels each)
- 2 Heat Capacity Mapping Mission (HCMM) scenes (1000x1000 pixels each
- 10 Topographic maps (2000X2000 pixels each)
- 15 Aircraft SAR, TIMS, AVHIR, etc. scenes
- ? Processed 3-D reflection seismic data

b) Irregular array:

- ? Gravity
- ? Magnetic

c) Line

- ? Seismic data
- 5 Aircraft Scatterometer tracks (4 polarizationsX10 anglesX500 points/track)

d) Tabular

>500 lab and field reflectance spectra ($\leq$2000 points/spectrum)

I.5.4 CURRENT OPERATING PROCEDURES

Existing data need to be located. At present, this involves contacting centers such as EROS, and ordering listings of available image data over the test site. Once located, the data need to be obtained. At present, photographic prints, etc. are ordered from the owner of the data.

Calibration of remotely sensed data is imperative for quantitative studies. At present, calibrated data are obtained directly from the source, or the investigator must calibrate the data using experimental techniques. Many projects, including this one, require coregistered data sets. At present, this is accomplished by the investigator by finding corresponding tie points in the images and reformatting one to match the other.

Once the above processes have been accomplished, the various investigators need to share the intermediate data sets. At present, this involves sending Computer Compatible Tape (CCT) copies by mail. It is also desirable to share data at various points of processing to facilitate discussions, solution of problems, and comparisons of processing algorithms.

I.5.5 TECHNICAL REQUIREMENTS

DATA LOCATION AND ACQUISITION - For a complex study such as this, many data types are required, from diverse sources. Access to a central directory is a key requirement.

DATA TRANSMITTAL - Tape data are now mailed, as are intermediate product images. Elimination of the resulting delays would speed the research and enable truly cooperative analyses among the physically remote researchers. A 1.5 Mbs link would be used as soon as available. Spectral data has been transmitted over 1200 baud links, as has text data, but this is awkward. These and catalog and directory conversations would benefit from standard protocols and interfaces, and an increase in data rate to 9.6 kbs.

PREPROCESSING - Data registration, overlays (particularly of disparate data sets), and geocoding occupy inordinate amounts of investigator time. As a minimum, the system must provide expeditious, streamlined techniques. This is a high priority.

FORMATTING - The various input data are not in compatible formats, and must be converted to common form for overlay, transmission, and comparisons. A system-wide data format structure is needed immediately to avoid proliferation of further incompatibilities.

CALIBRATIONS - These also occupy more investigator time than is desired. If these are not satisfactorily performed by the data sources (and to date they have not), a system calibration service will be most valuable.

IMAGE ANALYSIS - This is presently done at all project locations. This is satisfactory, although a means to streamline the processing (particularly the CPU-intensive processes) and to intercompare results are needed. At some time in the future access to increased computation power will become important.

NETWORK PROCESSING - Adequate data-interchange rates, common formats, and coordinated interfaces will enable processing steps to be carried out at various locations, as best suits a particular situation. This will be an important contribution to experiment efficiency and flexibility.

OUTPUT PRODUCTS - Data must be sent between locations for high quality film production, for example. If the University of Hawaii location cannot be supplied with suitable equipment, a system service for precision film recording would be valuable.

DATA STORAGE - Not a major problem for active working data. However, results will not be conveniently available to other experimenters without some archive repository. As this is not a part of the task, the PLDS would perform a valuable service if such an archive were available. In particular, a spectrum archive should be established to allow easy access to spectra as required during the task and afterwards. Because of the relatively lower data volume compared to images and resident expertise, this archive is appropriate for UH or USGS at Denver to curate.

I.5.6 RESPONSE OF THE PLDS TO THE REQUIREMENTS

DATA ACQUISITION - The PLDS could establish a catalog query and access structure and establish contact with other catalogs and data bases, such as the USGS Earth Science Information System.

DATA TRANSMITTAL - 1) The system will arrange for 9.6 kbs transmissions between the various investigator locations, notably at this time JPL and the University of Hawaii. The USGS Denver center will be included if the spectra library is located there. 2) NASA is planning a 1.5 mbs link (the Program Support Communications Network, PSCN) between Centers. The PLDS could extend this to the UH. 3) Intra-center connections to the network will be required. These will be arranged as the intra-center configurations are determined.

PREPROCESSING - The PLDS could develop production-quality data rectification, registration, mosaicking, and data subsetting techniques. These may be available in several of the NASA Centers, such as JPL and GSFC. The PLDS is considering establishing these as value-added services; if this is done, they will be available to this science task.

FORMATTING - The PLDS should establish recommended format(s) for use within the system and potentially across the different pilots. Any system-processed data would adhere to these formats.

WORKSTATIONS - The PLDS may develop at least a prototype workstation, with a standard executive and standard interfaces to the analysis modules. As these are adopted by the investigators, already available algorithms and new algorithms as developed by the investigators may be shared.

CALIBRATION AND SPECIAL PROCESSING - This scenario requires calibration, particularly for atmospheric removal and seismic processing and display. Additionally, new analysis techniques will be required and investigated for multisensor, multidimensional data sets. The PLDS may undertake the development of various advanced processing techniques, particularly for those things which are required across scenarios.

NETWORK PROCESSING AND ANALYSIS - The PLDS is considering an Intensive Computation Subsystem, which will make the large computers at several Centers available for use by non-residents. As the high-speed network is implemented, these services will be available to all PLDS members.

DATA STORAGE - A specific request of this science task is the establishment of a spectrum archive. The PLDS may assist in this establishment as requested.

I.5.7 SUMMARY

The PLDS will build on NASA-supplied services, such as the PSCN and capabilities available in the Centers to support this and other scenarios. All of the stated requirements of this scenario can be met, although the undertaking of the special processing should remain a science task until it can be shown to be accomplished on a production basis.

I.6 SCIENCE SCENARIO 6: MONITORING ENVIRONMENTAL CHANGE

Charles Robinov
U.S. Geological Survey

(*Note: The U.S. Geological Survey (USGS) submits this scenario as a suggested, needed area of scientific research which will benefit the entire land resources community and be well served by a program such as the PLDS. Although the USGS suggests this scenario, there is no commitment on its part to undertake such a program.)

I.6.1 BACKGROUND

This scenario describes research in using satellite data to monitor changes which will be significantly accelerated and made more productive by the PLDS. The PLDS will improve the productivity of scientific effort in this research in three ways: better data access, remote processing, and improved research communication.

Monitoring of environmental change is one of the most cost-effective uses of Earth satellites. The ability to view the same area repetitively at a uniform rate with uniformly calibrated sensors allows users to determine the rate, direction, and magnitude of change of various types of Earth features for land, water, and environmental management. Two general methods are used. The first involves classification and mapping of the desired features in an inventory mode and later reclassification and identification of the features to determine their change. The second involves simply the measurement of change in one or more of the parameters detected in satellite images (such as albedo) and then determining the type of feature and condition that has changed. Both the methods can be used for detection of rapid changes of state of features, but the second is better for the characterization of small rates of change of the condition of features.

Research needed in the field of change detection and monitoring should be concentrated on the investigation of the spectral bands and combinations that are most usable, the determination of the coarsest resolution that is effective, the repetition rate, and the methods by which change is displayed for the users benefit and understanding. It is not expected that elaborate research plans will be needed but that the specifications for operational use will be developed in a manner commensurate with the needs of users and tested against their applications.

I.6.2 OBJECTIVES

The major objective is to produce methods for the mapping and repetitive monitoring of environmental change in a cost-effective manner that can be routinely implemented by agencies and institutions that are responsible for environmental affairs and management. This objective can be achieved through evaluation of presently available spacecraft, aircraft, and ground monitoring methods; design of new sensors and techniques for new data processing, and evaluation of the effectiveness of these methods.

The Project outputs should be:

(1) Well tested methods of detecting and mapping change,

(2) An operational scenario for current and repetitive change mapping,

(3) User test and acceptance of methods, and

(4) Operation of a prototype monitoring system.

I.6.3 APPROACH

(1) Search literature on satellite image change detection and mapping methods;

(2) Control test methods in field situations;

(3) Decide on standard methods of change detection for specific applications;

(4) Recommend implementation of methods using present technology (such as Landsat and meteorological satellites) or develop modified cost-effective technology;

(5) Implement operational methods for routine worldwide change detection mapping; and

(6) Provide users with subscription service to original data and change detection maps.

I.6.4 CURRENT OPERATING PROCEDURES

Currently, investigators of environmental monitoring methods must search out satellite and other data from numerous sources, including NASA and NOAA. They must arrange for data processing and analysis within their own institutions or elsewhere. Their colleague communication is generally through technical publications, symposia, or a colleague network. Much of the time spent in a research project is devoted to the mechanics of procuring data (which may be quite slow) and in putting in place the necessary processing mechanisms. A major drawback can be the length of time required to obtain current data (especially where it is needed for monitoring of an environmental disaster), and nearly concurrent satellite and ground data needed for rapid analysis.

The data sets required are high and low-resolution multispectral repetitive images of the earth's surface in digital form, registered to available small and large scale maps. Data quantities required are (1) for a pilot test - 200 to 300 images of various types, and (2) for operational use - 10,000 to 20,000 images per year.

The institutions included (or those that should be involved) are the:

U.S. Geological Survey

National Aeronautics and Space Administration

U.S. Department of Agriculture

Environmental Protection Agency

Federal Emergency Management Agency

I.6.5 TECHNICAL REQUIREMENTS

The technical requirements involve the need to do the research in a complete and well organized fashion with increased access to data and facilities regardless of organizational affiliation.

The temporal requirements are for:

(1) Current data to be measured by the analysts 3-5 days after collection; and for operational results, data to be measured by users 1-2 weeks after data collection.

(2) The network and communication requirements are minimal during pilot tests but may become large during operations. They cannot be specified in detail now but await the results of the research.

(3) The processing will require digital image analysis computers (VAX 11/780 or equivalent) with change detection, image registration, and mapping programs and a capability to produce high-quality output products.

I.6.6 SCENARIO FOR PLDS SUPPORT OF RESEARCH

The PLDS should provide both better access to the minimum set of data bases required and convenient access to data bases that would be less extensively used without PLDS. Improved data access is provided by data base catalogs, simplified data retrieval, and wideband communication. The use of more and a wide-variety of data will produce more effective and better tested methods of deducing change. Mail delivery of unscreened data is slow and difficult, and reordering scenes significantly delays results.

The PLDS may also provide access to remote processing. The change detection research requires very accurate registration. Processing nodes on the PLDS may provide several different registration algorithms and supercomputers with large memory and high computational speeds. The several institutions involved should jointly conduct the research, and so the ability to use remote change algorithms will accelerate the search for effective change detection methods.

The third way the PLDS will improve productivity is by facilitating scientific communication. Electronic mail and computer conferencing over the PLDS will improve research management and scientific interaction. Results can be disseminated over the PLDS to the land science community. The PLDS may also evolve to provide specialized bibliographic searches for land science.

This research scenario, Monitoring Environmental Change, can effectively demonstrate the PLDS in its earlier stages, because the research will be facilitated by improved access to data and processing without requiring a large scale operational LDS to accomplish its objective.

I.6.7 SUMMARY

Coordinated research among many groups of investigators in the environmental monitoring field can be facilitated within the PLDS by the provision of current data from numerous sources, access to processing facilities, and improved communication among working groups. The PLDS could eventually become a prototype for operational monitoring systems to be jointly operated by responsible agencies.

II. TECHNICAL APPENDICES

II.0 INTRODUCTION TO TECHNICAL APPENDICES

There are several technical areas that need to be addressed in order to begin to implement the PLDS as outlined in Section 5 and to build towards the total system capability described in Section 4. The following sections address major technological areas of concern for PLDS development, and provide a concise summary of background information to build on in the next important steps of system design, and implementation planning.

The future Land Data System (see Section 4), would be composed of several subsystems connected by a communication network subsystem. Each of the following sections contains important information for the design and implementation of one or more of the conceptual subsystems in PLDS development.

The technical appendices are concerned with:

II.1 Data Management

II.2 Communication

II.3 Data Input and Output

II.4 User Interface

II.5 Data Manipulation

II.6 System Support Services

II.1 DATA MANAGEMENT

II.1.0 GENERAL DATA MANAGEMENT CONCEPTS AND CONSIDERATIONS

To achieve its objectives, the Pilot Land Data System must improve the way that land-related data are managed. Earth scientists must be able to easily find all data that are relevant to their research objectives, access these data rapidly, and build large multisource data bases that are useful to the general land science community. The PLDS will have access to a diverse collection of digital and analog data sets, at various locations, and these data will be in different formats. Some of these data will be archived by the PLDS. Users at different nodes of the system will want to access, as well as transfer PLDS-archived and other data. These data will need to be transmitted by tape, an alternate hard copy medium such as digital optical disc, and by electronic means. The PLDS data management subsystem must be capable of addressing these considerations efficiently.

Within NASA, commercial Data Base Management Systems (DBMSs) have not generally been used for scientific data bases containing both spacecraft and ground information for a variety of reasons. Most NASA DBMSs have been used to locate data files for a given mission, but in recent years DBMSs have been developed that cross-catalog data sets from different platforms and sensors. (See Table II.1.1 for a list of DBMSs currently used at various NASA centers.) Users have increasingly requested the capability for greater integration of data from multiple sources and access to these data with a high level, general purpose query language (Fujimoto 1981, Lohman 1981, Lohman et al. 1983, NASA 1981a & b).

One example of a NASA data base approach that might be a useful prototype is the data from the JASIN (Joint Air Sea Interaction) experiment of the Pilot Ocean Data System (PODS). This system uses the INGRIS relational DBMS, the QUEL query language, and the Entity-Key-Link-Attribute (EKLA) model (Ramey et al., 1983).

Another potential prototype is NASA's Pilot Climate Data System (PCDS) which manages multi-source data sets. The PCDS has been developed to serve as a focal point for managing and providing access to a large collection of actively used earth, ocean, and atmospheric data from sources such as NIMBUS, SEASAT, and NOAA missions. The PCDS provides data catalogs, inventories, and access methods for selected NASA and non-NASA data sets. Data manipulation capabilities have been developed to enable scientific users to analyze data using graphical and statistical methods. The PCDS is implemented on a VAX-11/780 computer and uses the Transportable Applications Executive (TAE) as a user interface. In addition, a commercial DBMS (ORACLE), a graphics package (TEMPLATE), and a statistical package (PROTRAN/IMSL) are available (Smith et al., 1983). The PCDS is available to users on the DECNET communications network.

The PLDS Data Management Subsystem (DMS) design will be based on the data management needs of the scientific users. As discussed in section 2, these users require a DMS that is simple to operate, but sufficiently powerful to meet the research needs through the end of the

TABLE II.1.1 USEFUL DBMS/DMS INFORMATION FOR NASA CENTERS

CENTER	DBMS/DMS	CONTACT	COMMENTS
ARC	DBMSs used for land resource data in ARPANET.	Sue Norman (415) 965-5912	CRAY Mainframe used with a VAX front end user-friendly DBMS software package. Networking between California, Boston, and Washington D.C. Transfer of very large data sets handled through magnetic tape transfer.
GSFC	Pilot Climate Data System (PCDS)	Paul Smith and Lloyd Trenish (301) 344-5826	Provides scientists with a user-friendly integrated system of data catalogues, inventories, access methods, manipulation capabilities, and display support. ORACLE is the commerical DMS used.
JPL	General Information on Data Base Management	Jose Urena (818) 577-9442 Barbara Anderson (818) 577-9484	
	Pilot Oceans Data System (PODS)	Chuck Klose (818) 354-5036	
	Pilot Planetary Data System (PPDS)	Chuck Acton Jerry Solomonson (Code EL) (818) 354-3869	Natural language user interface combined with a relational data base.
	Multimission Image Processing Lab (MIPL)	N. Sirrig (818) 577-5740 Ray Wall (818) 354-5016	A complete DMS/Image Processing System based on extensions to TAE.
	Planetary Science Analysis Support System (PSASS)	Jim Weiss (818) 354-4529	
JSC	Earth Resources Information Sciences Data System	Bob Musgrove (713) 483-5528	DBMS that is capable of acquiring, managing, controlling, and providing registration and image processing of the data. ADABAS, a commerical data base manager, is used to help link several nodes.
NSTL	ELAS-Image Analysis Software has limited Data Base management capabilities	Sid Whitley (601) 688-3586	Highly modular software system that is used at 50-60 locations on a dozen hardware systems. There is an active users group.

century. The subsystem must include capabilities that do not exist in present Data Base Management Systems - some of these capabilities translate into more powerful hardware, and DBMS software while others can only be described as value-added services. In particular, Artificial Intelligence (AI) technology may be an important component of the DMS. Since the development of AI is just now moving from the pure research to engineering applications, developing a smart DMS will be a long term goal of the project, but some progress should be achievable by the early 1990's.

One of the PLDS project's long-term development goals is to offer a "smart" data management service to the scientist. Based on this goal, the following types of support services will need to be available to some degree:

- Natural language communication between the user and the DMS
- Intelligent management and control of the DMS which optimizes system performance and user support
- Intelligent assistance in data management processes using expert systems
- Automatic data detection and classification using AI processes

Value-added services that an AI-based DMS can offer will guide and advise the user based on requirements and needs, and the degree of intelligence of the system. Considering the DMS design concept, the following example scenario should be possible:

> Express a data set requirement (hypothesis) using specific data sources that a scientist would like to use to support a research project. DMS evaluates the hypothesis and determines that other additional data will be needed to provide the specified data set. Based on an evaluation using an expert system, the DMS determines that some of the specified source data does not exist and other data are too noisy. Based on this finding, the system suggests that the requested data set cannot be provided, but that if the user will accept a data set that has less resolution, a different data source can be used and an alternative application data set created.

This is the way most humans would attack the data search problem: hypothesize a solution and test to see if it is achievable based on existing data. If not, look for an alternative that provides an answer similar to what is needed.

The DMS for the future LDS should be supported by three AI systems. The first of these will be a natural language front end processor to allow the user to interact with the DMS using natural English text, which can then be parsed into a data base command or query. The second AI system will be a knowledge-based expert system for DMS control, management and operations. The third system will be a knowledge-based expert system for performing complex data search and information detection, identification and cataloging.

II.1.1 FINDING DATA

Computer data base systems for maintaining catalogs and inventories of land-related data will be developed as part of the PLDS. These systems will be interconnected and improved incrementally until they are expert systems capable of finding all data (within practical limits) relevant to a scientist's research objectives.

Data Browse Development

The capability of providing the data for browsing is complex, since it requires not only accessing the data but having the data available on-line. This could require extensive disc farms, or jukeboxes. The data sets that users request to browse will also vary in size from small to large complex polygon files and images. Incorporating this service may require relatively longer connect time and as user demand grows, queueing may become a problem.

The cost to browse image data via the PLDS can be minimized by using commercial equipment. A possible mechanism to accommodate a browse request would be to generate RETMA (commercial TV) format signals when requested. These transmitted data sets would then be recorded at the user node for display and manipulation. This function could be accomplished via digital or analog transmission with continuous data transmission possible if each frame has an address that permits each user to grab a requested frame. This concept can be accomplished with present telemetry technology.

Images presently on microfiche (like Landsat) may be digitized in RETMA format directly. Digital image files may have to be subsampled to reduce the amount of data for the TV type images, or the data can probably be compressed by a factor of 3 to 10 with little loss of quality. Digital video discs would be used for the browse subsystem allowing continued data set incorporation.

The above technology is available and can be incorporated into the PLDS. For example, the Planetary Pilot Data System (PPDS) has implemented a TV-format analog image storage and browse capability. A similar system using analog video disc recording for images and maps is being implemented for the Army Engineering Topographic Laboratory (Costanzo 1983). Both systems use digitally-addressed frames and commercial monitor display. Discs typically hold 54,000 frames per side, and can be readily duplicated. These discs would be suitable for a central browse, or for distribution of TV quality images to the distributed nodes.

II.1.2 ACCESSING DATA

The PLDS will also incorporate various mechanisms for timely access to data. These will include the adaptation of general data-handling principles, data compression schemes, and artifical intelligence techniques for optimizing network data flow.

II.1.2.1 General Data Access Considerations

Proprietary Data

Open data, available for all users to access, and proprietary data with restricted access, may be in the system. A principal investigator might desire restrictions on data because of current research activities. A mechanism must be established so that each of these data types can be recognized and handled appropriately by the data management system, but remain transparent to the general system user.

Data Timeliness

Most investigations require data to be delivered to the investigator in a timely fashion, to avoid costly project delays. The PLDS has no control over the time involved in placing data in the archives, but can work with the archive to minimize retrieval delays. If the user community requires access to data in a rapid and direct mode, then electronic transmission (see Appendix II.2) will be necessary. Whether or not this service for any large data sets will be an integral part of PLDS is as yet to be determined. Such a service could made available on the PLDS for an additional cost.

As mentioned earlier, network connections to various archives will be required and these interface modules must be properly designed. Interfaces may be for catalog queries or for data access. Interfacing to already existing data bases, such as the USGS data bases, the Pilot Ocean Data System (PODS), and the Pilot Climate Data System (PCDS), would serve as a useful early prototype activity to test these interfaces.

II.1.2.2 On-Line Directory and Catalog Design

Integral parts of a distributed system, which provides access to data in widespread locations, are the on-line directory and catalogs. The directory and catalogs will be used to identify the characteristics and locations of the data. The design will be two-tiered, consisting of one central directory (catalog of catalogs) and several specialized catalogs. The central directory will contain information which generally describes the data sets and will point to the specialized catalogs for more detailed data set descriptions.

The purpose of the PLDS catalog is to provide a central source of information, about a variety of data sets, in a standard format. This information should be sufficient to enable a user to determine whether to retrieve and use data from the data sets. This catalog should be available on-line through computer terminals at remote sites through the PLDS communication network. The catalog should contain summary information that can be queried by using keywords, initially, and eventually by the use of natural language.

Examples of information in the central directory include:

- Data type
- Data source
- Data set size
- Geographic coverage of data
- Time coverage of data
- Brief description of data
- Method and frequency of data collection
- Physical location of the data
- Availability and names of contacts for further information
- Pointers to specialized catalogs

The specialized catalogs include those that already exist at numerous agencies and institutions (i.e., USGS, the EDC INORAC retrieval system for Landsat data, etc.), and those that will be established in the near future. Curators of these specialized catalogs will provide summarizing information about their current data sets and providing this high level information to the Central Directory. Access to the local catalogs will be provided via the Central Directory and be transparent to the user.

Examples of information in the specialized catalog are:

- Specific data set identification
- Geographic location
- Cloud cover
- Data quality
- Acquisition dates
- Other acquisition parameters
- Time coverage
- Processing levels
- Ordering information
- Data set processing history

II.1.2.3 Data Storage and Compression

Data storage is important to the PLDS for two reasons. The first reason is data preservation. Data stored on magnetic tapes deteriorate over time unless actively maintained. Therefore, PLDS long-term plans should include consideration of storage media that will ensure data preservation. The second reason is on-line support of data transmission. This is not an immediate PLDS requirement; however, as the number of user requests for transmission of data increases, it may become necessary to store data in a form that permits on-line access to selected data sets with minimum network transmission time. Presently, the storage, manipulation and distribution of georeferenced data, including satellite data, are handled by a variety of approaches and systems. More often than not, these systems are not compatible in storage format, making interchange difficult.

The PLDS will be required to handle large volumes of land-related data for transmission, display and manipulations. For very large data sets, such as a full Thematic Mapper image (320 Mbytes), it is neither practical nor generally necessary to transmit a full resolution scene over the network nor maintain on-line many TM scenes. Therefore, it is necessary to develop summarization techniques that produce reduced volume data sets that are useful to many scientists. An example of a technique which achieves compression ratios of about 8-10/1 on MSS data with no adverse impact on the per pixel classification accuracy is the Cluster Coding algorithm developed by Hilbert (1977). Other techniques include the lossless compression to be used on Galileo (Rice 1979).

Further compression topics that need to be addressed when designing the DMS include:

- o Identification of techniques for significant data compression suitable for remotely sensed image data in the PLDS.
- o Recommendation of a system providing for browsing of image data, compressed data suitable for fast classification, and selected uncompressed data for limited use.
- o Algorithms for compression which take advantage of parallel and vector processing architectures.
- o Complementary decompression algorithms which work on relatively small computers (e.g., microcomputers) used in scientific workstations.

II.1.3 BUILDING VERSATILE DATA BASES

The PLDS will apply a system of data management principles, storage structures, and advanced hardware and software technologies to the large multisource data bases whose structure and content will improve information system performance and applications.

II.1.3.1 General Data Base Considerations

II.1.3.1.1 Geocoded Data Structures

A fundamental requirement of the PLDS will be the ability to store georeferenced maps or images that are in raster, vector, or hybrid structure. Raster ("cellular" or "grid") structures represent maps or images as two-dimensional arrays of numbers. Each number corresponds to a uniform-sized rectangular subdivision (called a "cell" or "pixel") of the original map or image. The principal advantage of raster structures is that it is easy to write manipulative software for almost any application when the data are stored in this manner. The major disadvantage of raster structures is that for spatially sparse data sets, these structures use computer storage space inefficiently. Also, raster structures may not be as accurate as vector structures for delineating region boundaries if the cells are large.

Vector ("linked" or "polygon") structures represent the three elemental spatial entities (points, lines, and regions) in an explicit manner. Points are described by their coordinates, lines are described by strings of points, and regions are described by strings of points which enclose an area. Each of those entities is usually preceded in computer storage by a "header" which identifies it as a point, line, or region and which contains associated line-map based cartographic modeling information. The variety of these applications and systems explains why vector structures are so much more diverse than raster structures. The advantages of vector structures are that they generally use computer storage space efficiently and can easily incorporate topological information. The basic disadvantage of vector structures is that it is relatively difficult to write computer programs for editing and manipulating data stored in this manner.

Hybrid structures combine the characteristics of both raster and vector structures. Examples of hybrid structures are "quadtree" structure (Hunter and Steiglitz 1979), "vaster" structure (Peuquet 1983), and "topological grid" structure (Goldberg 1984). Much research is being directed toward applying hybrid geocoded data structures to practical geographic data handling.

Maps and remotely-sensed images of the Earth are the major sources of land resources data. These maps and images have three characteristics which point up the importance of their storage structure. First, they are large, which impacts the storage capacity, transmission capabilities, and processing resources of present computers. Thus, using geocoded data structures which save computer storage space is important. Second, land data sets come from a wide variety of sources. Remotely-sensed data are usually represented in a raster structure, and ground-truth maps in a vector structure. Typically, data sets must be converted into a common structure (as well as a common projection and spatial resolution) before analysis can be carried out. Thus, effective data structure conversion and manual digitization procedures (which are often the most time-consuming, error-prone, and costly elements of geographic data processing) are of great importance to the PLDS. Third, maps and remotely-sensed images are subjected to complex computer analyses. The structure in which a data set is represented must be matched with the programming language and processing architecture of the computer.

II.1.3.1.2 Data Base Design and Maintenance

DBMSs today assume that all of the on-line data they manage are on magnetic disc. This may not be the case for much of the data involved in the PLDS. The PLDS DMS will have to interface with a variety of devices, such as magnetic or digital optical discs. It must provide access primitives for each device, reformatting as necessary for data transfer between devices, and tracking capability for the location of off line data. Data set maintenance, in an overall data management environment, requires controls for ensuring data currency, quality and validation. Specific guidelines will be established to provide for local and remote data backup, security of sensitive material, and testing procedures for the purpose of validation.

The PLDS Data Base Administrator (DBA) will have the ultimate responsibility for maintaining an audit trail for updates to the various data files residing throughout the network. The DBA's responsibility should also include the coordination of activities relating to data consistency throughout the project. Construction of a

Data Base Filebook for all configured data bases and files is mandatory. This filebook contains pertinent information about each of the data bases in addition to data definition, size, format, and coding convention for each element. In addition, access rights, in the form of keys or passwords, would be used to control sources and users' access to each data base and file.

Each participating center or institution would support the DBA through local Data Managers who would have responsibility for their local data bases. These managers would oversee the updates to their particular data base as required. Data currency would be the responsibility of the initiator, but the DBA would continually monitor individual data cells for frequency of update. The DBA would be responsible for coordinating large updates at the project level to ensure concurrency. The DBA would also ensure that pertinent documentation is disseminated to the various centers at the conclusion of major data modification. The quality must be the responsibility of the initiator of that particular data set. The DBA function will guarantee the availability and integrity of these data, including recovery from temporary loss due to hardware or software failures. In addition, the DBA should provide a mechanism that permits any user to validate data by providing all pertinent documentation related to the data source.

II.1.3.1.3 Data Validation

The aim of data validation is to ensure the integrity of the data sets; therefore, this responsibility has to reside with the data producers. In the case of derived data resulting from scientific research activities, the individual researchers will have to be responsible for validating their data sets before they are allowed to become part of the PLDS public archive. Because of the potentially large number of data sources, it is important that the PLDS have a standard validation procedure.

As was the case with data quality, the DBA staff would coordinate validation efforts, but the actual initiator must provide the expertise and assistance. The various data types required for support of this project would require different levels of verification. The extent of validation documentation would also vary between data sets. In addition, access routines would have to be validated to ensure that user queries are properly satisfied. This activity would be fully coordinated with the DBA.

II.1.3.2 Advanced Data Base Management System Development

In the past, human experts have supported users to assure proper system operation, control and management. However, these systems always suffered from three major flaws:

(1) they were unresponsive,

(2) data selection was limited by the cataloging or data dictionary which required previous analysis of the data,

(3) full use of the data base required extensive training and system experience.

The intelligent data base that is being proposed is expected to help the system users in several ways. First, it will provide a friendly interface between the user and the DBMS with a natural language front end that will convert English text into data base commands and queries. Second, it will help the user define the information required. In addition, the system should intelligently (and independently) search for, detect, analyze and catalog data sets for use in specific research or to expand the data base management system.

Following are the important highlights covering AI and the PLDS:

Natural language interfacing to a DBMS is now commercially available. These systems provide an English text interface between a DBMS and the user, such as the natural language front end being sold by Artificial Intelligent Systems, Inc. The system is promoted as being capable of parsing English text into commands and queries that the data base system can understand. The performance of this system has not yet been tested. However, IBM recently paid forty million dollars for a resaleable license for this relational data base management system which should give one some indication of the product's capabilities.

The second proposed AI system is an expert system for management and control of the DMS. Present engineering of such a system is feasible. An example of such a system is Rl, developed by D. McDermott for Digital Equipment Corp. for configuring VAX computer systems. In this system the solutions are viewed as a hierarchy of subsystems and each subsystem is treated as being related in a time independent manner. Although it is true that system control and management will not be time independent, the logical interferences will, which makes the solution set to the management and control problem very reasonable and the system development achievable.

The third proposed AI system of the DMS is an expert system for performing data search, detection, identification, and cataloging, as it is related to data base operation. The development of such a system is expected to be extremely difficult because of the complexity of the problem solving involved. The unique way in which individual scientists collect and describe data makes it almost impossible to impose standardization as is done in commercial applications. As a result, often the relationships between data sets is not exploited in a data base because there are differences in data description and the relationships linking observed values cannot be completely defined a priori.

Certain "hard data" about the data may be expected to be routinely supplied by the system, such as spacecraft and sensor, observation time and location (for space acquired data). Other, "soft data," may be available if they are derived, such as cloud locations, data quality, a priori conditions, data set processing history, principal investigator, or description of a mapped variable. The set of data to be available for search and the searchable relationships between them will be developed over time.

An expert system could be used to evaluate the information in a request and the types of actions that can be taken by the system and make the necessary translation to allow the request to be activated. In addition, the system may have the data that are being requested but not in the form that user wants it. For example, a user may request a Landsat image taken on June 15, 1984. The expert system would have the knowledge necessary to know that dates for Landsat images are in terms of days since launch rather than calendar date and would recognize the need to make a translation.

There is always the possibility that the system understands the user's query but does not have the requested data. As an example, suppose that a user requests satellite image data over a specified geographical area during a particular time period at a certain resolution. The system may have data with all of the attributes except for the resolution. In this case, the expert system could determine the data sets which satisfy these requirements to some degree and make recommendations for possible substitute data sets.

Another potentially useful application of the expert system is the analyzing and cataloging of the contents of data sets for specific attributes which can be added to the data base for use in future queries. The contents of a large majority of the data sets in a typical archive have never been reviewed for the purpose of identifying and extracting data content attributes. Therefore, research tends to be confined to a very small subset of data sets for which there is a considerable amount of information available. With an expert system, a knowledge base can be developed for recognizing predetermined features and also identifying "interesting" features that were not anticipated. During the computer's normally idle item, the expert system can be given the task of analyzing and cataloging features.

Such systems stress the limits of knowledge engineering because they require the building of a system where the data and knowledge may not be reliable, where the data are noisy and inconsistent, and where a large number of alternative solutions exist. However, the importance of such a system is that manual cataloging is limited by human cognitive abilities and the time available for performing such work. Given the increasing rate of data acquisition, an automated system is the only hope that researchers have for gaining access to the bulk of the information collected.

II.1.3.3 Advanced Hardware Technology

The future implementation of an intelligent data management system for the extensive LDS data will require storage resources exceeding present capabilities. Data storage requirements are expected to be in excess of 10^{12} bytes, including image data. The mass storage hardware needs for this amount of data will require new technology if the data management system is expected to perform in a near real-time manner.

Today, the hardware systems available for the storage of digital data include magnetic tape, hard and soft magnetic disc, and optical disc. The oldest of these storage media is the magnetic tape. Tapes will be used by PLDS in the near-term for archiving data and supporting off-line processing functions. As the need for quick data access to support on-line processing functions grows and the quantity of data being used increases, the usefulness of the magnetic tape as the PLDS storage medium will be limited by its capacity, access time and durability.

The second storage medium, magnetic discs, has direct application to the DMS mass storage requirements because of the large amount of data that can be randomly stored for rapid access on the large and medium sized computers. Presently, Winchester Technology is leading the way in magnetic disc storage with capabilities in excess of 300 megabytes (using thin film technology) for five and a quarter inch disc systems at a cost of less than five thousand dollars. Such systems are quite attractive for supporting the storage needs of the user workstation, but still fall short of offering the storage capabilities needed for the PLDS DMS.

Optical discs offer the highest data storage capacity of any technology and appear most suitable for fulfilling the DMS real-time storage requirements. Recently, systems have been demonstrated with disc capacities of 50 gigabits, at transfer rates of 300 megabits per second with a bit error rate of less than one in 100 million. Over the next 5 years, commercial systems will be available that offer multiple disc (jukebox) configurations that have storage capabilities as high as 100 trillion (10^{14}) bits.

Optical storage systems are projected to be of two types, write-once storage media and magneto-optic storage media which are erasable write and read hardware. The first type of system has a storage cost of between 1 and 5 cents per megabyte and is ideal for storing permanent data (such as Landsat) in the data base. It is also anticipated that a large number of write-once optical discs will be used. Such discs are not expected to be extremely expensive and should be available commercially in 1984 from Phillips, RCA, Storage Technology Corporation, and others. These systems are projected to cost $200,000 and will store on the order of 100 trillion (10^{14}) bits.

The second type of optical storage system is one in which the medium can be rewritten many times. This technology is immature compared to the write-once systems but expected to begin to be available commercially in 1984. The capacity of such a system is expected to be similar to the write-once system, but the cost will probably be 2 to 5 times greater. Presently, the performance of this hardware is limited by the high bit error rates that are being experienced and by the ability to manufacture the hardware at a reasonable cost.

This technology is currently being implemented at the Marshall Space Flight Center as part of the Data System Technology Program (DSTP). The DSTP will be centered around three VAX 11/780 computers with a jukebox storage system composed of 128 tellurium coated optical discs, each capable of storing 83 gigabytes of data.

II.2 COMMUNICATIONS

Communications and networking are at the very heart of the PLDS concept. The acquisition and control of all communications-related equipment and activity required for any NASA project or program, or for administrative use, is governed by strict policies and regulations.

II.2.1 INTRODUCTION

At the broad federal level, guidelines for the procurement and utilization of telecommunications are prescribed by the General Services Administration's (GSA) Federal Property Management Regulation, Subchapter F, pararaph 101-35.000. For NASA in particular, policies, responsibilities, and procedures for the acquisition, control, and management of telecommunications are delineated in the NASA Management Instruction 2520.1C (May 22, 1978). The following Section (II.2.2) reviews NASA policies that apply to the procurement, implementation, and use of communications services. The rest of this section will address specific communications technologies and their relevance to the PLDS.

II.2.2 STRUCTURE AND REGULATIONS REGARDING NASA COMMUNICATIONS

For purposes of definition, communication facilities and services within NASA are typically divided into two categories: 1) Operational Communications (NASCOM) and 2) Program Support Communications (PSC), previously called Administrative Communications. The first category, Operational Communications, consists of those circuits and facilities carrying space mission-related information to support NASA technical programs and projects. The second category, Program Support Communications, is a new concept that encompasses all administrative communication, including research data communications.

The NASA Operational Communication System (NASCOM) is a global system established and operated by NASA to provide long-haul operational communications support for all agency projects. NASCOM controls a large network of dedicated circuits, both land-line and satellite links, servicing the various space-related missions and their support facilities. These circuits and communications services are usually provided by leased arrangements from common carriers of terrestrial and satellite communications. Goddard Space Flight Center (GSFC) is responsible for planning, design, implementation, operation, and maintenance of NASCOM.

The Program Support Communications Network (PSCN) provides the following existing services:

- Voice teleconference
- High and low speed fax
- Telemail via NASANET

- Packet switched network for shuttle, legal, media, and Inspector General data.

Additional services to be added under PSCN, planned for implementation in FY'85, include:

- Supercomputer data links
- Video teleconferencing
- FTS service between NASA facilities and FTS support to GSA
- Other data communication requirements not satisfied by NASCOM

Marshall Space Flight Center (MSFC) is responsible for the planning, design, implementation, and operational management for the PSCN. All NASA field installations are responsible for on site operational and administrative communications of their respective installations. On site operational and program support communications which will be interconnected with NASCOM and PSCN must be coordinated and have concurrence of GSFC and MSFC, respectively.

For any project or program requiring communications, the following procedure must be followed.

(1) The project is responsible for establishing and documenting the long-haul communications requirements including, (a) a traffic model, (b) the network configuration, (c) recommendation for implementation, (d) interfaces to local communications, and (e) starting time and duration of service required.

(2) These requirements need to be validated and supported by the NASA Headquarters Program Office sponsoring the project or program. The Program Office then submits the validated requirements to the Office of Space Tracking and Data Systems (Code T). Code T transfers the requirements to the appropriate communications group, NASCOM or PSC, for implementation. The sponsoring Headquarters Program Office is responsible for ensuring that sufficient time is allowed for GSFC and MSFC to plan and implement the requirements.

For the PLDS, PSC would be responsible for implementing the long-haul communications physical links. However, design and implementation of protocols, local area networks, interconnect strategies, etc., would not be a PSC function but should be handled by PLDS personnel.

II.2.3 PLDS FUNCTIONS REQUIRING COMMUNICATIONS

The functional areas of the PLDS that have a requirement for communication services are:

- Cataloging query - Typically, the data traffic type generated is characterized by bidirectional, short, interactive messages.

- Remote resource sharing - It allows the user to access remote hardware and software resources.

- Data transfer - Involves the physical transport of data from producer node to user. The volume of data may vary from a small data file to very large data sets. The user requirements for data transfer range from 2 to 4 weeks for some research efforts to a few seconds for quick-look data (browsing).

- Electronic mail - Provides electronic personal communications among the PLDS user and producer community. It typically involves short letter-type messages and some medium-size data files.

II.2.4 CLASSES OF DATA TRAFFIC

There is a dichotomy between the requirements of two kinds of information transfer, interactive communications, and bulk data transfer.

- Interactive Communications - This type of traffic is centered around the cataloging and data request services and access to remote on line processes. It is characterized by interactive, unscheduled messages with relatively low volume and almost instantaneous response.

- Bulk data transfer - This type of traffic is associated with data transfer between a data archive and a user, between nodes of the network, or between users. It is characterized by a large volume, with lesser emphasis on response time.

II.2.5 COMMUNICATION TECHNOLOGIES FOR THE PLDS

Communications technologies of particular significance to PLDS requirements involve: data transport, communications protocols and control, and interconnections between Local Area Networks (LANs) and Long Distance Networks (LDNs).

II.2.5.1 Data Transport

The PLDS communication needs may be met by any of a variety of transportation links, including:

- Mail, courier delivery of computer media
- Dial-up
- Ground point-to-point
- Satellite point-to-point
- TDMA/Demand access
- Public/semi-public packet switched networks

- Local area networks

A data distribution network is defined as the aggregation of these methods to service the data distribution needs of the user community. Under this definition the network is assumed to constantly change and evolve, expanding and contracting according to demand.

The methodology for identifying and evaluating system transport alternatives is depicted in Figure II.2-1. The activity shown in the charts represents an initial iteration that will produce transport system concepts defined in terms of the general characteristics listed. Two inputs are crucial to the identification of alternatives: transport systems parameters defining the needs of the various scenarios, and knowledge of available data transport systems and technology.

Based on these inputs, "strawman" transport systems may be developed that satisfy discipline needs. These design concepts are defined to a level of detail sufficient to indicate particular choices with regard to the transport system issues listed in the center of the chart. The next step is an evaluation of alternatives, in terms of cost, flexibility, and the degree to which discipline requirements are likely to be satisfied. There are several choices for transport media which are discussed in the following sections.

Mail or courier:

Despite the extensive use of electronic communication methods, the mailing of magnetic or optical media may still be a very cost-effective method of data transport for very large data volumes when the response time requirements are of a few days or longer and not very critical. However, transfer costs are not trivial by the time material, copy, and shipping charges are totaled. In addition, total system throughput is typically unsatisfactorily low.

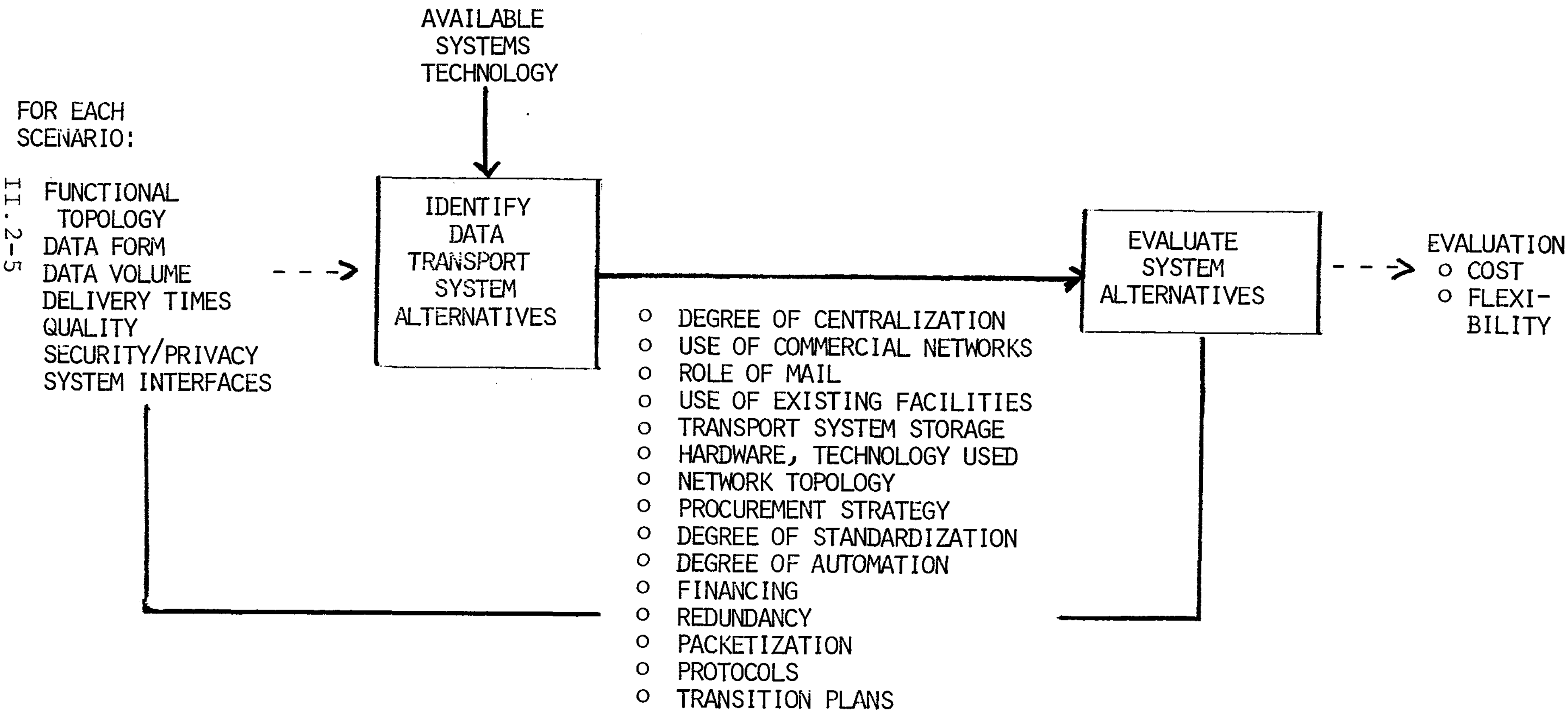
Dial-up:

Point-to-point dial-up communication service for voice and data can accommodate data speed ranging from 0 to 1200 bps, and 2400-9600 bps claimed with some newer and more expensive equipment. The costs are based on distance, time-of-day, day-of-week, and measured usage. Dial-up has the favorable properties of low fixed monthly costs, ubiquitous availability, and low-cost hardware. It has the problems of high noise, inconsistent line quality, relatively high connect costs, low transfer rate (although this is improving), and increasingly, difficulty in supporting a full-duplex mode (switched voice circuits or satellite delays form unsatisfactory dial-up connections).

As explained earlier, a class of traffic generated by some functions of the PLDS is typified by short, conversational, interactive messages. Some of these functions are the cataloging services, data request, and remote access to online processes. The costs of dial-up communication are usage-sensitive, making it a very reasonable alternative for such traffic.

FIGURE II.2.1

EVALUATION METHODOLOGY FOR TRANSPORT SYSTEM ALTERNATIVES



Point-to-point leased lines:

Point-to-point voice or data lines are leased from common carriers. Private voice-grade lines are generally used at up to 9600 bps. Beyond these transfer rates, leased lines become very expensive. Difficulties include a long lead time for installation (and a substantial installation charge), moderately expensive modem equipment, and fixed point-to-point routing. Advantages are a consistent line quality, and a usage-insensitive costing method.

AT&T remains the only common carrier to-date to provide a coverage of the whole territory of the United States. Other providers of leased lines are Western Union, Southern Pacific Communications Company, and ITT U.S. Transmission Systems.

Satellite point-to-point:

Satellite channels are offered for two-point service. Their limitations are their essentially point-to-point character (although the technology for multipoint service exists), their restriction in most cases to a small number of terminal cities, and their intrinsic, long propagation delay. The costing methods vary, but are never usage-sensitive. There is generally a cost component for the actual charges of the local telephone company for local distribution at both ends of the point-to-point channel service. Some satellite communication services vendors are:

- American Satellite Company
- CYLIX Communication Network
- RCA Satellite Service
- Southern Pacific Communications Company
- Western Union

Demand Access by Single Channel Per Carrier (SCPC) or Time Division Multiple Access (TDMA):

Most wideband communication channels, whether terrestrial or satellite, are dedicated to a single point-to-point route. The familiar narrow band telephone network, however, uses wide band trunk channels switched between users. Since dedicated wide band satellite channels are often temporarily unused, two different demand-access methods have been developed to share the satellite channels. These methods are called Single Channel Per Carrier (SCPC) and Time Division Multiple Access (TDMA). In SCPC, as the name implies, each voice channel has its own carrier. In TDMA, only one carrier at a time is present in the satellite transponder, but the carrier is rapidly switched between different ground stations.

Public/Semi-Public Packet Switched Networks:

There is a clear match between the interactive class of traffic of the PLDS and the capabilities of public packet switched networks -- orientation towards user terminals, interactive operation, and ready availability. The network is accessed by dialing the phone number of the nearest public network node. In most cases, this involves a local phone call.

An important feature of public packet switched networks is that the network provides a number of value-added services. These can include converting data from one character set to another, adjusting transmission speeds for maximum data flow without compromising the data's integrity, permitting a network device to talk to any other device on that network, routing data along the most efficient path, and reducing transmission costs by multiplexing data from multiple sources onto fewer telephone lines.

One of the most important value-added services is that the vendor takes full responsibility for the operation and maintenance of the network. For large customer-owned and operated networks, to provide for the monitoring, control and maintenance functions for all the components of the network can be not only a significant recurring cost, but can also involve an important initial capital investment. Although the costing method is different for all the networks, in every case it is usage dependent and distance independent.

Another value-added service that public packet switched networks typically offer is an electronic mail service. The electronic mail service can be used for the communication between the users of the network and for the transmission of small files of data. Other semi-public packet switched networks such as ARPANET, CSNET, and the proposed NASA Program Support Communications Network will have to be evaluated as they relate to the requirements of the PLDS.

Local Area Networks (LAN):

Local area networks provide high-speed communications among users and devices located within a relatively short distance from each other (from same building to a few miles away). The relevance of LANs is that they can permit a cost-effective sharing of resources on a local level. Some nodes of the PLDS will be part of local area networks (i.e., JPL's Institutional Local Area Network) and they will interface to the PLDS network by means of gateways (see Section II.2.5.3).

II.2.5.2 Communication Protocols and Control

Communications protocols are the extensive software procedures used to control digital data transmission among computers. Most currently operational protocols were developed by computer hardware vendors. Examples are IBM's System Network Architecture (SNA) and Digital Equipment Corporation's (DEC) DECNET. The best known multi-vendor protocol is used in the ARPANET network. The International Standard Organization (ISO) has developed a seven-layer protocol model called the Open Systems Interconnection (OSI) model. This is a model, not a specification. SNA, DECNET, and TCP/IP (used by ARPANET) also use five or six layers and can claim to approximate the OSI model.

The purpose of having several layers is to allow changes in hardware or functions with minimal impact. A single, multipurpose communications software program would require rewriting each time communications terminals were upgraded or links were reconfigured. The disadvantage of using a hierarchical protocol is that the protocol's robustness is achieved by an extensive software package, which requires considerable machine processing time in operation. This overhead reduces the effective transmission data rate significantly below the raw channel bit rate.

The CCITT has developed the X.25 standard for the three lower layers of the model, and it is expected that most vendors will eventually support this standard. Standards for the higher layers, though not currently available, will be developed in the future.

II.2.5.3 Interconnecting Local Area Networks (LANs) and Long Distance Networks (LDNs)

LANs are those that exist within one building or within a local group of buildings, while LDNs cross continents and oceans. LANs are inexpensive and provide wide bandwidths. LDNs usually have higher error rates than LANs and often have long (one-half second) delays due to satellite transmission. LDNs may use the switched telephone network, dedicated narrow and wide band circuits, or a combination of these. Because of their complex topology and high error rate, LDNs use layered communication protocols, similar to the OSI model, more frequently than LANs.

These fundamental differences between LANs and LDNs make it difficult to interconnect the two. Rather than making a simple hook-up, a computerized gateway that functions as a node on both networks must be designed. The gateway makes the LAN look like a single machine on the LDN. Gateways have been designed to convert protocols between some of the more common LANs and LDNs.

A better but more complex approach is to design the LANs and LDNs as a compatible, hierarchical communications network. For example, the lower layers of the protocol could be identical in the LAN and LDN. This would reduce the computational load of protocol conversion. A multi-vendor standard protocol, such as the CCITT X.25 may become, would greatly facilitate this approach.

II.2.6. IMPLEMENTATION APPROACHES FOR PLDS COMMUNICATIONS

Under NASCOM, a new capability being procured consists of a satellite-based TDMA system with a terminal node at each NASA Center. The initial capacity is divided into standard commercial telephone-compatible data rate channels. Also included in the system will be a circuit for a compressed full motion color two-way video conferencing capability for inter-Center conferencing use. The schedule for the TDMA service calls for implementation to begin in July 1984 and be completed at all centers by February 1985.

For PSC, a consolidation of existing functions and addition of new capabilities has been undertaken as noted in Section II.2.2. The new system is planned as a common user-integrated network consisting initially of gateways at fourteen locations (all NASA installations). Transmission channel capacity of at least 1.5 mbps capacity will be provided with a network control center located at MSFC. The current schedule for the system calls for vendor implementation twelve months after award of contract, the RFP having been issued in December 1983, with time to contract award following the normal NASA procurement cycle.

At this time, there may be several options for implementing long-haul PLDS communications depending on the structure of the proposed PLDS.

- In the first option, approval for 9.6 kbps inter-Center links could be obtained through NASCOM. The communication would be between VAX computers via the DECNET protocol. With this approach, the possibility exists of linking ARC, GSFC, and JPL at 9.6 kbps with compatible machines in 1984. Expansion of this concept to other Centers is a potentially easy undertaking.

- A second option utilizing the NASCOM capability is to design an experiment using the new TDMA system described above. This would permit participation by all NASA Centers with file communication rates that are compatible with transferring image data.

- A third option is to utilize the new PSCN when implemented. As currently structured, PSC would have the implementation responsibility for PLDS communications. However, as noted above, until PSCN is actually implemented, some experiments may be conducted with the NASCOM systems with switchover to PSCN in the future.

Depending on the magnitude of the communication service requested, Code T, GSFC and/or MSFC usually require greater than a year's lead time from submission of requirements until service is provided. Fortunately, initial requirements for 9.6 kbps inter-Center links were submitted from the Code E Information Systems Office (ISO) to Code T in July 1983 for implementation of services in 1984. These requirements need to be updated with specific detail as the PLDS refines its needs.

II.2.7 NETWORK CONTROL AND ADMINISTRATION

Network control is the application of real-time and near-real-time measures to control the operation of the network. Network management consists of system planning and engineering processes. These include the establishment of standards, practices, methods, and procedures for the performance and operation of the network, and analysis of the system performance to ensure proper operation and derive improvements.

The administration of an operational PLDS will include the detailed management of the system and the provision of support to subscribers. The major functions are:

- Operations management
- Accounting and billing
- Subscriber services

The management of a dispersed and heterogeneous system such as the PLDS will require establishment of detailed control procedures and close coordination with PLDS central management and users.

II.2.7.1 Operations Management

The operations management function provides configuration control, management and planning, coordination to detect and resolve problems, and general resource management. Operations management will be responsible for PLDS network control, and will perform appropriate monitoring, reporting, coordination, restoring, and systems maintenance functions. General systems engineering assistance would be provided to PLDS subscribers in installing and maintaining PLDS software and hardware. Planning of new system implementations, maintenance of a relevant standards data base, and oversight of standards adherence also would be performed.

II.2.7.2 Accounting and Billing

Accounting of resource utilization, both by consumers and producers of data and services, is vital to system planning and maintenance. Even in a non-commercial environment, knowledge is required of who is using what data and services. This information might be used to select data for on line or archival storage, planning system expansion, evaluating usefulness or data collected, identifying user-affinity groups, or allocating system costs.

The accounting and billing function also provides an activity profile of subscribers as well as charging and billing mechanisms for resource utilization. Regulating demand may be accomplished by charging for resources and services in such a way that demand does not overtax availability.

II.2.7.3 Subscriber Services

Substantial effort will be required to provide user support and training and to disseminate information about the operation of the network to current and prospective subscribers. This function will also provide for a formal feedback mechanism to allow subscribers to request changes in system services, to input additional requirements, and to report problems.

II.3 DATA INPUT/OUTPUT AND INTERFACE SUBSYSTEMS

II.3.1 INTRODUCTION

The overall goal of the Input/Output and Interface Subsystems is to convert external data to a form usable by the analysis nodes and the transport system. The majority of the land analysis capabilities developed by NASA have been implemented in a raster processing environment. Based on this, the Input/Output and Interface Subsystems must be organized to allow as much data as possible to be converted efficiently to the raster data environment. However, it must also address the efficient handling of vector data because of the role it plays as an input, its use in controlling data handling processes, and its potential in the underdeveloped realm of spatial analysis (see Section II.5.4).

The most efficient means of providing data and analysis capability to meet PLDS goals is to implement comprehensive data standards, without limiting flexibility. Sufficient knowledge and experience in dealing with raster and vector data are available to adopt data standards at least in these areas. Ideally, data formats and standards should be derived from both the analysis and communications requirements. However, the short range expectations for standardizing formats for analysis in NASA are dim. Therefore, a short range goal must be to develop a comprehensive methodology for data transport.

The PLDS must be able to handle whatever types of data, in whatever condition, the archives accessed by the system offer. However, the PLDS must be able to document the general condition of the data. This defines more system structure. Because data formats are specific to a given archive and the networking is best done with some commonality in data formats, the interface modules will probably be located at the archive. Initially, data reformatting software must be developed on a case by case basis. However, strategies must be developed to minimize the reformatting effort required.

II.3.2 STANDARD FORMATS FOR DIGITAL DATA

To make the digital interchange of the various types of data practical, a method must be devised for recognizably labeling them to allow software to properly translate them into formats internal to the nodes. This will facilitate the early transfer of data, and provide the path toward developing a standard transfer format family. This labeling can be done within the bounds of the new Landsat Format Family structure. It will require the documentation of the extant formats, the assigning of suitable recognition codes, and the design and development of translation modules. This will provide some degree of commonality between the data types.

A second step should be to devise a more common format and coding definitions to which the various source data can be translated. This format should also be within the Landsat Family. Its use may obviate much of the translation which would otherwise be required. In time, as more data are translated to the new format and coding scheme, this Format Family will evolve as a de facto standard. The third step should be to generate any new data in a format that is directly-compatible with the Family Interchange format structure.

An obvious standard for raster data is the Landsat Format Family. The raster format is in widespread use today and is flexible and efficient. What are needed are the extensions to cover all data types used in the PLDS, and the adoption of standard coding (e.g. for geographic position or for cadastral data).

Two national committees are developing standards: one convened under the USGS by to unify the data representations by the Federal Agencies (Federal Information Processing Standards -- FIPS) and the other from the National Bureau of Standards via the American Congress on Surveying and Mapping (ACSM), concerned with digital cartographic data standards. The ACSM committee is scheduled to produce a format structure recommendation by 1985, with eventual consideration for a FIPS standard. NASA is working with the Consultative Committee on Space Data Systems (CCSDS) to define guidelines for standardization of data at the institutional level. NASA should work to achieve compatibility among these activities.

II.3.2.2 Interchange Format Structure

The following structural argument is being considered by the ACSM committee as a general framework for developing the standards for interchange of digital cartographic data, including images. As these standards will have widespread applications, the PLDS should consider the same basic structure in developing its specific embodiment.

The interchange format is to be applicable at the ISO Open Network Level 6 or 7 (see Section II.2.5.2). That is, the transmission system will deliver a package identical to that delivered to it. The system may have applied protocols and formats within itself, but these are invisible to the data sender or user. The types of formatting discussed below are mission-independent formatting, mission-dependent formatting, and layered, onion-skin formatting.

II.3.2.3 Mission-independent Formatting

Data transmission may be at the undefined bit stream level, or the bits may be grouped into characters by the sender (this is the more usual case). The acceptable set of characters, such as ANSI X3.4-1977, must be mutually agreed upon by both the sender and receiver. As no information other than the characters themselves is coded, the meaning of the sequence must be prearranged.

The character sequence begins to acquire meaning as the characters are grouped into fields. Defining the lengths of the series of fields may be done through a directory entry sequence which gives these lengths, or by the use of defined field terminators to separate fields in the data stream. Again, the meanings of the fields and field sequence must be prearranged. At this stage, as these meanings are independent of the coding for field lengths, transmission is completely content-independent, although self-defining at the field level.

As more intelligence is added to the transmission, methods of coding some of the relationships between the fields and records may be added. This is the thrust of the ISO Data Descriptive File work for ANSI X3L5. In this approach, the hierarchical structure of the records may be defined as well as the field structures, although the meanings of the fields are user-dependent. This is as far as definitions should be invoked for general purpose data transmission.

II.3.2.4 Mission Dependent Formatting

The data definition allowed in the Data Description File allows specific meanings to be applied to the fields, and thus specialized format embodiments to be defined for different uses. This explicit data definition mode requires Data Definition Records (DDR), or equivalent File Descriptors, which describe each field in corresponding data records. The next step in specificity is the use of predefined keywords which, when encountered, implicitly define the structure of the data to follow. This method has the advantage that only necessary keywords are used, and may be used in any order. A keyword scanner would be used to recognize the format of all acceptable keywords; therefore, all keywords must be defined, together with their data structures.

As the definitions of the formats become more mission dependent, more implicit definitions (in separate documentation) may be employed where the keywords and the order of the fields are implicit, and the DDR provides only the sizes, structure, and (possibly) location in the data record. This can give more compact data sets, as the DDR and tags are not required, provided that all of the predefined fields are fully filled. In a completely implicit form the DDR only identifies the data records, and all structural information is given in the documentation. Thus, modifications of the format are difficult, requiring format document and reading software revision.

II.3.2.5 Layered, Onion-skin, Formats

Following the lead of the ISO in defining layers of protocol and related labels, a Volume Descriptor or Primary Label would be added to the data set. Various organizations have each defined equivalent, but incompatible, primary labels. Data set identification, data structure definition, and data elements are different, and should be defined at different stages in the format definition. Therefore, they should be in different records, with varying levels of control authority.

The Primary Label will be the outer layer of the interchange format. As its use will be global, Primary Label must identify the generating source, control authority, specific format and revision, specific data set identification, and identification of the format definition authority for the data. This label must also designate data set size, plus any other information which will be required to allow blind machine reading of the records that follow. To maximize generality, the primary label should embody a minimum of definition, which would be recognized by anyone. It would also have the benefit of allowing free area for local use. The Standard Format Data Unit (SFDU) of NASA and the CCSDS, and the Landsat Format Superstructure, serve this purpose.

At this point, any necessary universal record headers would be defined. The ISO Data Definition Records and the Landsat Format Family have found it advantageous to define record headers which include the record number, record type, record length, and various flags. These are to be used for all records in the interest of machine scanning.

Particularly in the transmission of images, multiple data sets (e.g., spectral planes or derived image-like data planes) will be required to describe a given geographical area. Thus, there will be several sets of data following a Volume Descriptor. The Landsat program has used File Pointers as secondary labels to identify these data planes and provide only sufficient information about each to allow reading of the corresponding data definition information.

Global recognition implies that the Primary Label be implicitly defined to provide a starting point for reading. Each Secondary Label may be flagged as implicitly defined or as explicitly including its own Data Descriptor information.

II.3.2.6 Standard Format Data Unit

NASA is funding a study to define an SFDU. The SFDU is a conceptual data object that is passed between users. The SFDU community is the international group of users of "space data." The SFDU consists basically of a formatted and labeled data set, and thus defines an interchange format. It will consist of a primary label which serves as a global identifier, a set of secondary labels which carry information about the data, and the data set itself. It also provides a nesting feature which allows an SFDU to consist of a set of SFDUs. The structure is exactly parallel to the Landsat Format Family with the exception of the nesting, and the various types of labels serve the same purposes.

Because of the parallel purposes and structure, because the Landsat Family is several years ahead of the SFDU, it is to be hoped that the SFDU structure can be defined to incorporate the Landsat structure. This will provide a common format family between the two groups of users. Continued contact with the SFDU implementers will be maintained to assure that maximum compatibility between the two systems is obtained.

II.3.3 STANDARD FORMATS FOR OTHER DATA TYPES

II.3.3.1 Vector Data

Vector data refers to point, line, and polygon data. The types of vector data used determines how they are handled in analysis, but poses no real dilemma in terms of format standards for data transportability. Since NASA owns so little of the U.S.'s vector data the most logical approach is to adopt the vector data standards that the USGS is developing. It is expected that this standard vector coding would be enclosed in an outer layer as described in Section II.3.2. This would ensure future compatibility with a vast and valuable data source, eliminate the need for NASA to fund such development, and serve as a step toward developing the type of overall capability within the government that the PLDS is trying to develop within NASA.

II.3.3.2 Catalog Data

Catalog data for the PLDS is not as critical as other types in terms of volume or in its role in analytical processes. However, it is essential in developing compatibility with highly diverse information management systems. This must be developed in conjunction with a system-wide protocol for requesting directory information. This combination of catalog data standards and protocol would allow rather simple functional interfaces to be generated between existing directories and the network. A committee consisting of knowledgeable representatives for relevant catalogs can develop the required standards and protocols very quickly.

II.3.3.3 Tabular Data

This data type consists of two general categories, data that are primarily human readable (character data), and data that are not covered under any of the previously defined data types. This would be the media by which reports, statistical findings, and various forms of ancillary data would be transported. The only requirements for handling this data type are to establish record lengths, number of data records, and an indicator of the character type (ASCII, EBCDIC, or binary). Standards for this data type could easily be established by the committee recommended for finalizing catalog data standards. These data would also be enclosed in the standard outer layer.

II.3.4 DATA INPUT TYPES AND REFORMATTING

The specific data inputs required by the PLDS are addressed within the science scenarios and will not be covered here. This section will address input of different categories of data.

II.3.4.1 Existing Digital Data

Existing digital data must be reformatted according to the standards applicable for its data category. Raster data must be reformatted to the Landsat Family Formats, and vector data will be reformatted to the adopted standard for vector data.

II.3.4.2 Map Data

Map data normally must enter the system by being converted to vector data. Maps and photographs could be scanned by an automatic digitizer and fed directly into the system as raster data; however, at present most maps are not amenable to automatic scanning. The current method for entering map data is manual digitization. The exact technique used will vary, but the data must be converted to the standard vector format to be used by analytical functions throughout the system.

II.3.4.3 Field and Sample Data

Their varied nature and use preclude establishment of rigid format standards for field and sample data. The data may be entered in several ways but must be converted to standard vector format. Much of the data can easily be entered by use of manual editing functions for vector data. The mass of the remaining data may be entered as tabular data.

II.3.5 EDITING AND QUALITY CHECK

Once data have been entered into the system, it is desirable to screen them before analysis begins. Once high reliability has been established, systematic quality checking can be discontinued; however, such capabilities can be maintained for troubleshooting.

II.3.5.1 Raster Data

The usual form of data checking is observing the data as displayed on a workstation. Manual capabilities exist for editing header and cellular data. This can be a method for correcting data values or control parameters, and as a means of manual input of some data. Sufficient capability exists for supporting the PLDS in this area, but again the format varies.

II.3.5.2 Vector Data

Editing vector data is more involved than editing raster data. Normal editing functions for printing, modifying, and manual entry of both control and point-related data are required. Some form of data display (image, graphics, or plotter) is required. Special purpose editing functions are needed where map data are digitized into line segments and automatically linked into polygons. These functions must be able to interactively assist the digitizer operator in obtaining error-free line segments that are linkable. Once a standard vector format is adopted, these functions must be available for use throughout the system.

II.3.5.3 Vector-to-Raster Conversion

To minimize the quantity of data to be transmitted, it is desirable to transmit it in polygon format if the user can use the data in this form. Otherwise, the data must be converted to image form before transmission. This implies that the PLDS must perform the polygon-image conversion as a value-added service. Any polygon geolocation and quality verification done by the PLDS should be handled at this time. Other possible value-added services are data cleanup, re-registration, scaling, and conversion to new coordinates.

There are two ways of converting vector input to raster output, direct gridding and interpolative gridding. In direct gridding the geometric information from both the vector data and the raster data must be used to develop a transformation into lines and columns of the raster data file.

Interpolative gridding is required when a data set exists as samples data and is to be used in multivariate analysis with raster data on a cell-by-cell basis. Such processes estimate the given parameter for each cell based on the sample data. Probably the most acceptable technique for generating these estimations uses cubic-spline interpolation to generate a minimum curved surface through all the sample data. Existing capabilities are probably sufficient for the needs of the PLDS but would be computationally slow for large volumes of data.

II.4 USER INTERFACE

II.4.1 INTRODUCTION

The PLDS will be effective only to the extent that it facilitates the sharing of data and resources among NASA scientists and research institutions. The NASA research community exists at a wide range of computational and technological levels. Some locations are well supported with supercomputers, staff, and a plethora of software; others have very little. System design must contend with the multi-technological user interface by first recognizing the existing levels of technology and then developing the system to support them.

The rapidly expanding capabilities of smaller, less-expensive computers, coupled with advances in computer networking, provides the basis for the all-important link to the PLDS. These smaller machines or "workstations," are powerful single-user computer systems which allow a wide range of options for many different levels of capability, quality and cost. The workstation approach provides a natural expansion of local system capacity, either by replicating individual workstations or by providing a port whereby remote computing resources can be accessed. Low in cost, microcomputer-based systems have been taking over many applications that previously were performed by large minicomputers, including data collection and data transmission. The workstation will become the personal library, communications device, document preparation tool and analytical engine to the user. (In this context, the user is the scientist doing research.)

A distributed system architecture, based on efficient communications and powerful workstations, decouples the burden of both I/O and smaller processes from the large central computer, leaving it to those processes for which it is most suited. In addition, the system capacity is limited primarily by the communications subsystem. Thus, large numbers of users can be supported concurrently with little performance degradation.

Low cost workstations can be the tool for accessing the resources provided by the PLDS network. These workstations can provide the hardware and software tools to process and display data in a timely and cost-effective manner. Communications form an important part of these systems, with the need to share data and information, access data catalogs and archives and share word processing. To provide these capabilities, data bases and processing systems must be linked by computer networks.

Implementation of the workstation concept will require research in all areas of computer technology. Communications, data sharing, distributed processing and user interface are just a few of the areas that will need better definition and development. An important first step will be to utilize the existing technology within NASA and NASA-funded research laboratories (e.g., universities) to support the PLDS and, subsequently, to determine the technological developments necessary to create a workable system. This activity must rely heavily on existing NASA research, first to define overall system needs, and second to focus existing system research.

The remaining discussion in this section deals with specific subsystem designs, and the hardware and software required to support that design. Alternative workstation configurations are outlined and communication requirements identified for the support of NASA activities in land research.

II.4.2 DESIGN CONSIDERATIONS

Successful PLDS operation using the workstation concept will depend upon having resources available for those tasks that can not be run at the workstation. The other major subsystems offer resources to the user via the workstation and communications subsystem for performing a broad range of functions including data base management, and computationally intensive processes.

Supported by powerful workstations, each user will be able to tailor the local capabilities to best suit the intended research, while having other resources available via high speed digital communication. Based on such an approach the user will not only be able to interact with the system to obtain the data necessary for research, but will be able to store, manipulate and analyze these data locally.

The concept of the User Interface Subsystem (UIS) is that processing will be performed locally except for those services where a common data base must be accessed or where the computing power of the local hardware or workstation is not capable of providing adequately responsive support. There will be a wide range of workstations available depending on the users' needs and the funds available for workstation acquisition. The benefits to be gained from this approach are a reduction in operational cost, an increase in overall system performance, the ability for modular growth and systems modifications with minimal impact on the user, improved user-to-center communications, and data sharing.

II.4.3 HARDWARE REQUIREMENTS, GUIDELINES, AND STANDARDS

Certain basic capabilities are expected of any type of workstation, including input and output, text creation and transmission, data storage, and processing and display. While it is not in the interest of participants in the PLDS to impose rigid standards of any kind, these requirements do have some rather specific implications regarding hardware.

With respect to the Central Processing Unit (CPU), it is clear that 16 or 32-bit architecture should be employed in the interest of processing speed. Similarly, sufficient memory space and mass storage, from disc and tape, is required to handle large amounts of data. There should be hardware floating point capability. Emphasis should be placed on choosing processor families which will offer upward compatibility to newer, faster, and better processes in the same family lineage. The M68000 or NS16032 families are a good example, offering upward capability with 32-bit architecture.

Modular design employing bus-based architecture offers the benefits of expandability and upgrading, tailoring the system to particular users' needs and budgets. Furthermore, a wide range of alternatives are possible in speed, storage, and display by using different products compatible with bus design.

Perhaps the most vital ingredient in the workstation concept is communication. While networking has been handled as a separate topic of discussion elsewhere (Appendix II.2), important options include:

- off line media data transfer, providing low cost, high-volume, slow data transfer (using tape or floppy disc),

- telephone links via modem, providing low cost, long-distance, low-volume, slow-to medium-speed transfer,

- local area networking, providing medium cost, short distance, high-volume, medium- to high-speed transfer,

- long-distance networking, providing high-cost, medium- to long-distance, high-volume, medium- to high-speed transfer, and

- satellite communication, providing high-cost, long-distance, high-volume, high-speed transfer.

II.4.4 SOFTWARE REQUIREMENTS, GUIDELINES, AND STANDARDS

Software support is a major and expensive component of the UIS. This software must be user-friendly, portable, compatible with other hardware and software, and responsive to a variety of needs. Such software support includes the operating system, higher level software development languages, and ancillary application utilities described below.

There are incompatibilities between operating systems that can lead to serious problems. These relate to difficulties associated with file transfer and distributed processing. With respect to file transfers, incompatibilities can result from the use of different representations for the character set (i.e., ASCII, EBCDIC, etc.), file formatting procedures, word sizes, representations for floating point numbers, etc. The primary problem in distributed processing is that systems employ different models for representing an executing task. These differences include process creation, control, termination, intercommunication, privileges, and security. Consequently, developing procedures to permit compatibility between operating systems is a necessary but highly complex task.

Operating systems must provide multiple vendor support, a flexible and efficient software development environment, and a friendly user environment. Tailoring of some user environments by systems programming is highly desirable, particularly if it can provide a network of distributed operating systems which is transparent to the user. Some viable options for operating systems are MS-DOS for

8086/8088-based 16-bit systems and UNIX for 16 and 32-bit systems based on processors from a number of manufacturers.

For efficient analysis, a wide range of applications software must be supported by the PLDS, such as FORTRAN and PASCAL along with more recent languages such as C, LISP, and ADA. In principle, the workstation would include some of the following capabilities depending on the type of workstation required:

- an operating system that permits some minimum level of compatibility with other PLDS computer systems
- high level languages that are supported by other PLDS subsystems
- a powerful editor for supporting software development and document preparation
- powerful analytical and support tools for assisting the user, including:
    - image processing
    - text editing and document preparation
    - graphical representation
    - data base management
    - intersystem communications and control

A large number of existing image processing packages are useful to the PLDS, including ELAS, CIE, Portable EDITOR, and TAE/VICAR2/IBIS. Similarly, statistical packages of interest include BMDP, SPSS, and SAS, while mathematics packages such as IMSL should be included. Text editing capability might be comparable to WORDSTAR for CP/M and MS-DOS systems and/or VI/NROFF for UNIX-based systems.

The availability of easy-to-use, yet powerful, software responsive to a variety of user needs is essential to the PLDS. A judicious choice of software will allow these requirements to be met.

II.4.5 USER LEVELS AND SYSTEM CONFIGURATIONS

The PLDS must be designed to serve a variety of users whose needs, perspectives and capabilities differ widely. For illustrative purposes, three typical levels of users can be designated in terms of their locally available resources:

- Level I

  High level of resources, large mainframes, extensive capability for data analysis and many data bases. Level I users are candidates for local nodes on the network -- most probably NASA Centers and universities.

- o Level II

  Moderate level of resources; mini or microcomputers, some data bases and moderate capability for data analysis.

- o Level III

  Low level of resources; little in-house computer capability or data bases.

Typical configuration for the three levels of users are provided in Figures II.4-1 and II.4-2a and 2b. Figure II.4.3 shows ways in which the three levels of users might interact and communicate. The same type of approach, specifying varying "levels," can be applied to the analysis of alternative workstation configurations. A number of workstation configuration options, analogous but not specifically related to the three levels of users, are described in this section. These systems vary significantly with regards to their computing power. However, each performs effectively and efficiently for the particular set of processing functions it was designed to handle. Preliminary functional descriptions for different levels of workstations are presented in more detail in Section II.4.7.

Low-cost computer systems have little to moderate processing power, fair to good graphics capability, and limited memory and storage. Low speed communications could include floppy disc, modem, and some tape and network capability. These systems can also be expanded with additional hardware. The capabilities of these 8-bit systems include 64-512K memory and 64K to 5M storage capability, and cost varies between $5,000 and $20,000. Examples of this configuration level include 8-bit CP/M systems such as RIPS (Remote Image Processing System) developed at EROS Data Center, IBM personal computers (PCs) and IBM-compatible systems.

More advanced scientific workstations offer an extremely wide range of price and performance options with bus-based designs. The workstation processor is expected to be at least a 16-bit microprocessor. The system will support virtual memory and memory management and have sufficient system interfacing so that hardware such as image processors, color graphics and array processors can be readily interfaced. The workstation's operating system must be compatible with the other major subsystems at the file, command, and process levels. In addition, the workstation must have the ability to store large amounts of data that can be accessed readily. Presently such a system may not be commercially available. It is extremely important for NASA to be previewing these systems, so as to better understand what will be available and, therefore, what can be used in the system implementation.

The range of capabilities for these scientific workstations include 1/2M to 4M memory and 80M-1G storage capability, and cost between $20,000 and $100,000. Examples of early generations of this genre at the low end of the cost scale include MIDAS developed by NASA/Ames Research Center and the SUN Microsystems (Stanford University

Figure II.4.1 Typical Level I User Configuration

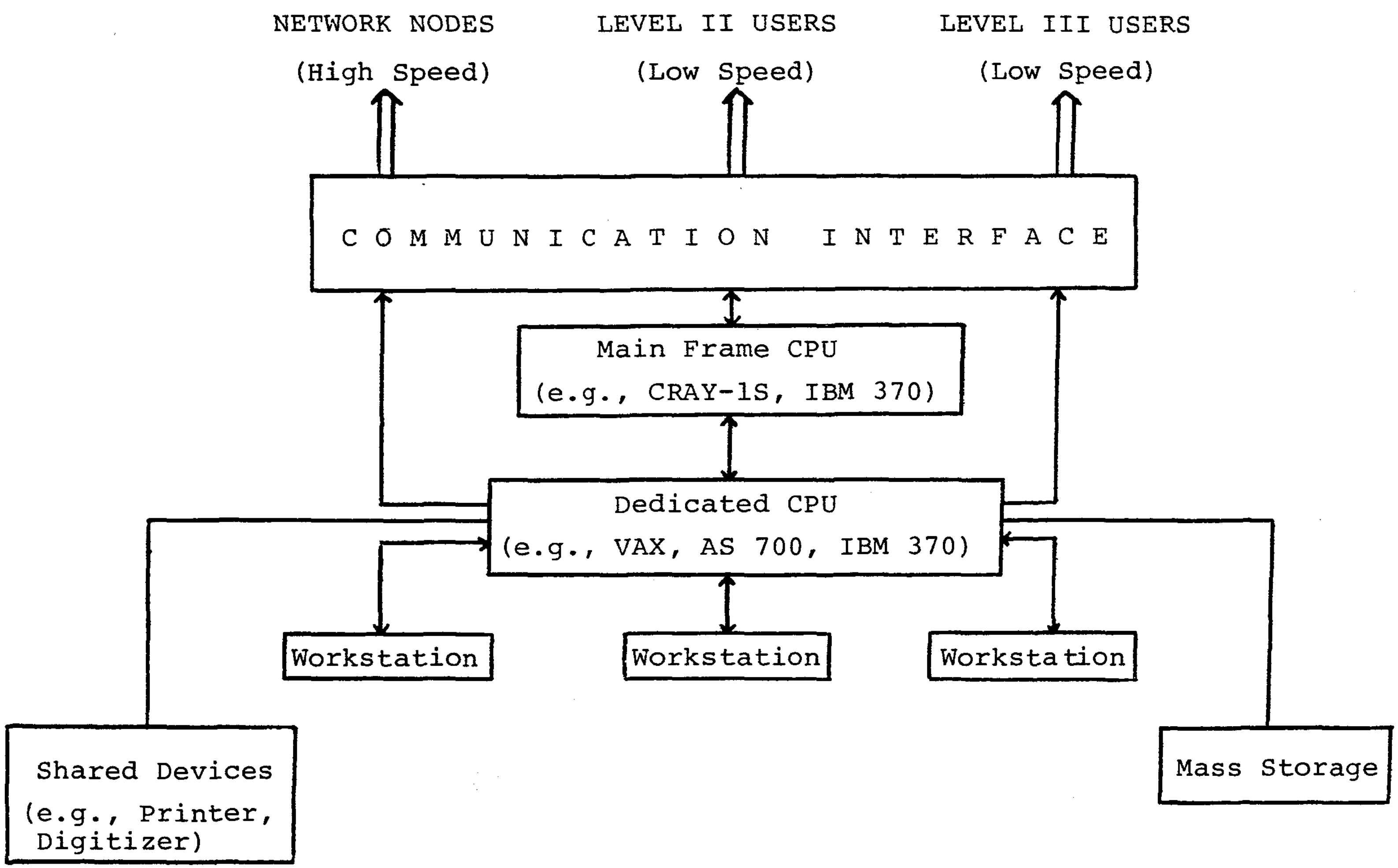


Figure II.4.2a Typical Level II User Configuration

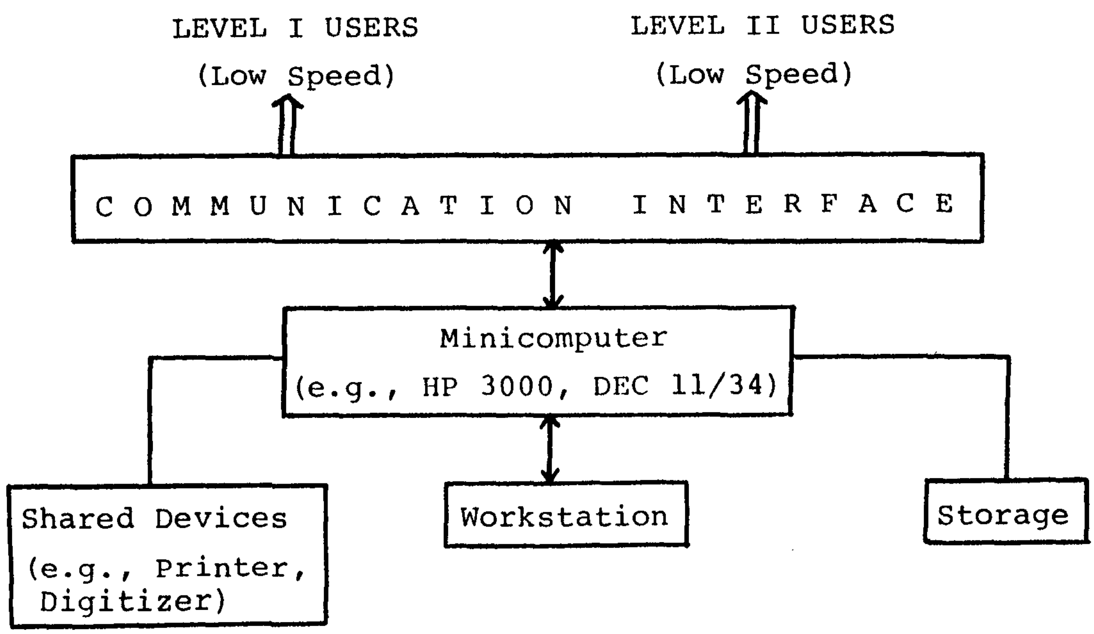


Figure II.4.2b Typical Level III User Configuration

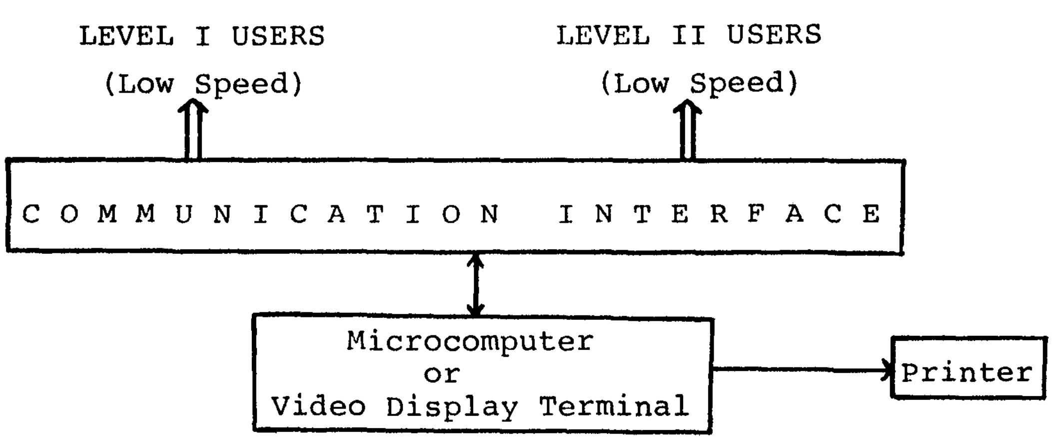
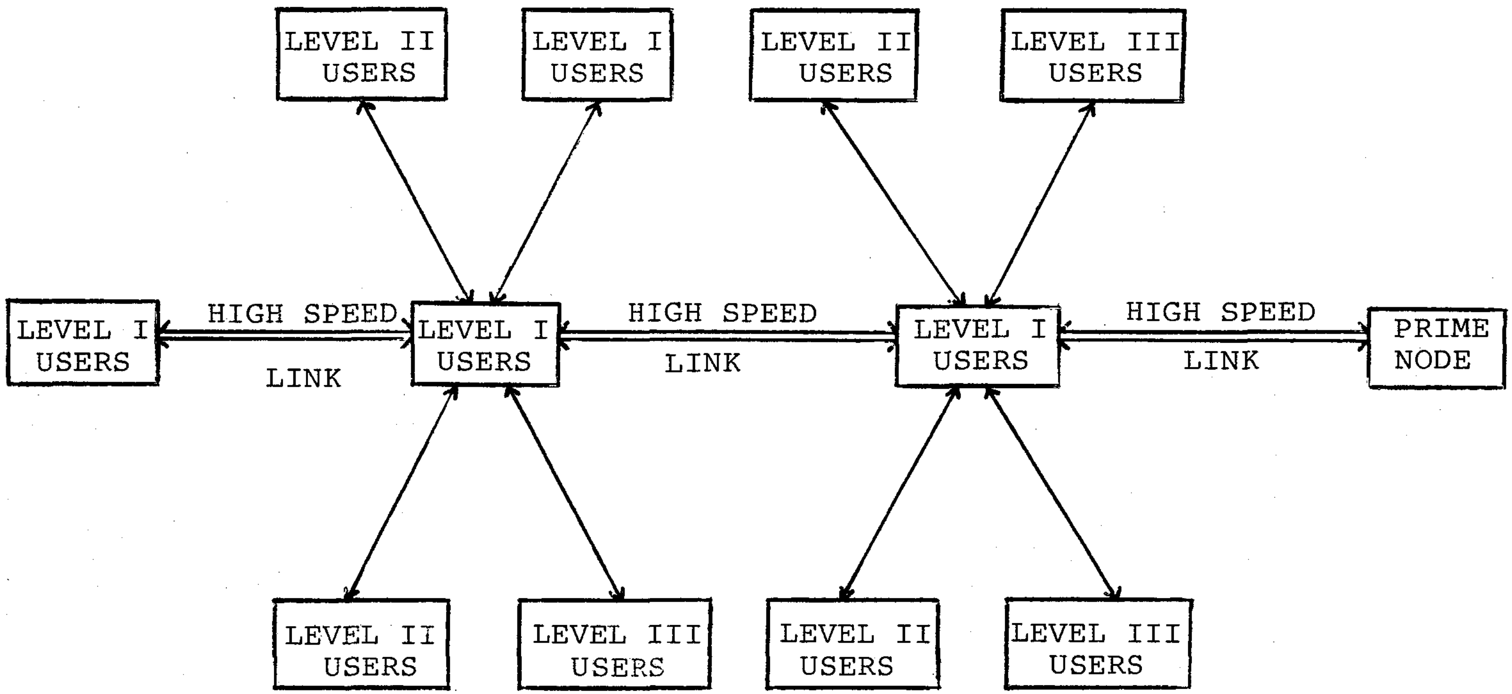


Figure II.4.3 Network Concept



Level I -- High Level of Computer Resources

Level II -- Moderate Level of Computer Resources

Level III -- Low Level of Computer Resources

Network) type of system. While these scientific workstations make effective stand-alone systems, they can also be networked or included as subsystems in larger, more powerful computing environments.

An important advance in computer technology is the development of artificial intelligence workstations -- high-cost, high-capability workstations tailored specifically for artificial intelligence applications. Artifical intelligence has been applied to problems involving image processing, natural language parsing and interpretation, geological exploration, and biological systems. Examples of these new systems include Xerox STAR, Diablo, Dorado, and Dandylion, the LISP machine, and Symbolics 3600.

II.4.6 TECHNOLOGICAL DEVELOPMENTS REQUIRED

The User Interface Subsystem will be the element used by every researcher involved in the PLDS, and therefore it warrants extensive study to ensure optimal design. Technological development is essential in a number of areas to achieve the full potential of the workstation concept. Specific areas of required research and the types of technological "breakthrough" needed are discussed below.

A major technical bottleneck for the PLDS concerns data volume. Current and future sensors (such as the Thematic Mapper) will involve the analysis of data sets many magnitudes larger than in the past. From the PLDS perspective, large-scale, on-line data storage (in the mega- and gigabyte range) might appear to be the most pressing problem that this implies. Actually, technological breakthroughs in the realm of optical discs hold the promise of providing adequate solutions.

The major problem for workstations is not on line storage of large data sets, but rather the mechanism, speed and cost of data transfer. Data transfer over a network is an essential consideration because large data set analysis will probably be performed on geographically-dispersed, larger machines. Clearly, the primary concern is speed. Faster methods of data transfer are essential to reduce costs and maximize efficient use of workstations. Efficient handling within a single workstation requires rapid data transfer from disc to memory (and back), and between processor and display modules.

Another major bottleneck relates to the availability, quality, appropriateness and transportability of software. The speed and efficient utilization of hardware is totally dependent upon the available software. However, the time and labor costs of software development and maintenance have become much greater than those of hardware. A uniform approach to software design that incorporates planning for future development is needed. In addition, emphasis on modular code design, proper documentation, and improved maintenance is required.

Other technological bottlenecks which must be addressed include:

- development and testing of communications protocols for distributed networks with workstations, large mainframes and supercomputers,

- o methods for incorporating new technological developments within existing workstations (modular in design), and,

- o optimum methods for interfacing with existing system hardware and software.

II.4.7 IMPLEMENTATION STRATEGY

The UIS is that subsystem of the PLDS that supports the interface between the user and other components of the system, such as the Data Management Subsystem and the Intensive Computational Processing Subsystem. In addition, the UIS provides a locally based, robust and powerful computing environment for performing a broad range of functions and operations. The UIS for the PLDS is envisioned as being a set of powerful microcomputer based workstations with a performance capability similar to the VAX 11/780 minicomputer (which is about 1 MIPS - millions of instructions per second). The design would be modular and not limited by CPU resources. The UIS would consist of large number of workstations ranging from a stand-alone image processing workstation to a small text editing workstation. In addition, the PLDS will offer limited support for intelligent and dumb terminals. Some specific recommendations for UIS implementation follow.

The ability to handle Bell Laboratories' UNIX operating system is a feature that most 16 and 32 bit microprocessors will share and is therefore a likely candidate for the UIS workstation. This software is already running on many mini's and mainframes, and provides a common link. An example of this high level inter-operability between mini and microcomputers has been demonstrated on a NSC 16032 based workstation running UNIX where over 80% of all VAX software was compiled and run without modification. Such inter-operability is one mechanism to achieve the desired performance of the PLDS.

A final consideration in the development of the workstation is the conversion of application software. Generally speaking, if the software is available on a minicomputer like the VAX 11/780, it should be easy to perform the conversion. However, there are issues that need to be addressed even with the high level compatibility of UNIX. These issues fall into two categories: technical and legal. The technical category involves addressing such things as using the co-processor for performing floating point operations, or the need for larger amounts of real memory or mass storage than are available at the workstation. The second category involves the acquisition of expensive software that presently runs on mini's and will need to be implemented on the functioning workstation. Examples of such software include ORACLE (a DBMS which cost $40,000 for the VAX), GIPSY (a $10,000 image processing software package), IMSL (a Fortran subroutine library which cost about $2,000 per year for a VAX), and DISPLA. There will be a real tradeoff between developing new software and licensing expensive software.

The following Table (II.4-1) provides a preliminary functional description of four types of workstations. These are intended only to provide interim guidance for development of the workstation until functional specifications are completed, early in the engineering phase.

Table II.4.1

Suggested Specifications for User Workstations

Levels

IPWS - Image Processing Workstation. <$50K, 32b
ASWS - Advanced Scientific Workstation <$25K, 32b
EWS - Engineering Workstation <$12K, 32 or 16b
GWS - General Workstation <$6K, 16b or 8b

		IPWS	ASWS	EWS	GWS
HARDWARE:					
o	a bus to support several hardware options	X	X	X	X
o	virtual memory	X	X	X	X
o	memory management	X	X	X	X
o	floating point hardware	X	X	(coprocessor)	
o	megabytes of RAM memory	4	4	2	1
o	megabytes of RAM storage	500	300	100	50
o	optional capability for using discs, and array processor	X			
o	high speed/high density tape drive	X	X		
o	floppy disc drives	X	X	X	
o	image processing system, 1024 x 1024 pixels, 4 planes, 8 bits deep	X			
o	graphics, 400 x 400 pixels, minimum resolution		color	b&w	
o	optional array processor to improve system performance		X		

Table II.4.1, continued

	IPWS	ASWS	EWS	GWS
SOFTWARE:				
o the UNIX operating system	X	X	X	X
o LAN high level communica- software	X	X	X	X
o a rule based system to support interactive operations with the various subsystems	X	X	X	X
o ORACLE data base management system	X	X	X	X
o a general purpose image processing software package	X			
o a programmers workbench for software development	X	X	X	X
o a writers workbench for development of technical papers and reports (includes speller and grammar checkers, formaters and typesetters)	X	X	X	X
o high level languages such as: FORTRAN, PASCAL, C, ADA, PROLOG, LISP	X	X	X	X
o graphics software	X	X	X	

II.5 DATA ANALYSIS AND MANIPULATION

II.5.1 INTRODUCTION

The data analysis and manipulation capability is a fundamental part of the PLDS. This section discusses several areas where the PLDS will perform data manipulation as a system service, in the form of data processing and data management. Methods by which the PLDS will interface with scientific data analysis, through the distribution of processing and sharing of algorithms on the network, are also described.

The PLDS should include a registration and rectification capability, a GIS which communicates with the DBMS and the image processing system, and an expert system capability which simplifies user interaction with the system. Each of these topics is discussed in the following sections.

II.5.2 RECTIFICATION AND REGISTRATION

Networking, data management and intensive computational systems are the highly visible elements of the PLDS. They each require that the data are in a readily useable form, i.e., digitized, rectified, registered and prepared for spatial analysis and modeling. Since multi-source data are typically not preprocessed, the PLDS will be required to provide that function.

The terms georeferencing, geocoding, rectification and registration have been used in the literature with various meanings. To avoid confusion, the terms will be defined below:

Spatial data are those for which the spatial relations between the data items are important. Georeferenced data have parameters provided to allow the data to be located with respect to a coordinate reference system. Geocoded image data have elements uniquely and systematically aligned along the axes of a coordinate reference system, with known position and scaling.

Rectification is the process of removing the instrument geometric signature. Geometric rectification is the operation which establishes the appropriate correspondence between a digital image and the segment of the earth's surface characterized by the image, that is, an "undistorted image".

Registration is the operation by which an image is mapped onto another image representing the same segment of the earth's surface. Consequently, digital registration of raster data requires pixel scaling in the two (or more) registered images to be the same, and pixels corresponding to the same ground area to be precisely superimposed.

II.5.2.1 Geocoding Considerations

Conversion of raw data to a geocoded form typically has two steps: georeferencing, using control point data from maps or prior imagery, and registration, to produce a new data set on the proper coordinate axes. Both steps are time consuming. Because much of the data in the archives may never be recalled by the system for digital use (for Landsat, the digital retrieval is estimated at 10% of the total data), optimum use of digital resources will be to achieve georeferencing during archiving, and geocoding during retrieval. Georeferenced data is required to permit geographical queries to the catalog. It also allows georeferenced imagery (such as the Landsat P-tapes) to be produced for users not requiring the geocoded product.

Geocoding of already archived data for non-interactive use would be provided by the PLDS as a value-added service. This is predicated on the concept that georeferencing and geocoding are extensive operations requiring some degree of specialized expertise to be accomplished efficiently. One current deterrent to the use of Landsat-type data is the effort and expense of registration, often not practical at a user facility. Thus, geocoding can be a valuable service of the PLDS, coupled with mosaicking to allow construction of analysis areas not covered by any one archived image (Zobrist 1983).

Providing on line interactive data requires that the data have been previously geocoded. It is unrealistic to require that currently archived data be routinely geocoded in anticipation of some possible future call. Rather, one practical scenario is that in which the data sets are geocoded the first time they are called, and then retained in that form. In this case, the PLDS establishes supplementary archives. This sets the structure of the system, and provides a method of control against unnecessary data requests (the more data requested to be geocoded, the longer it will take to deliver, and the more it will cost).

In the future, a new mission or sensor information system would be responsible for preprocessing the data, ensuring that the relevant ancillary data are present, and storing the data in the mission archive in a self-documented format. For Landsat data, steps are already being taken in this direction (Landsat Technical Working Group 1978, 1979). The PLDS will access the data from these more complete archives, translating and reformatting as necessary.

II.5.2.2 Performance Requirements

Specific performance requirements have yet to be defined for the PLDS. A general level of required performance can be defined which strikes a balance between useability, cost, and feasibility of achievement. Such a set has been derived (Ramapriyan 1980) from the discussions in the NASA Workshop on Registration and Rectification (NASA 1982) and Simonett et al. (1978):

- It is sufficient if the "system" performs as well as the users themselves do.

- Many Landsat users are satisfied with fitting the data to standard maps at 1:250,000 or 1:500,000 scale, implying errors less than 127 or 254 meters at more than 90% of the locations.

- Errors of less than 0.5 pixel for 90% of the locations for temporal registration are satisfactory for many applications (although some require $\leq$ 0.1 pixel error). The Landsat 4 specification is for 0.3 pixel temporal registration and 0.5 pixel image to map registration.

- Images should be rotated to the north - that is, pixel lines should be along a recognized earth-based coordinate system.

- Pixels sizes should preferably be multiples (and submultiples) of 50 meters.

Because most of the archived data are not georeferenced, geocoded, or registered, and each experimenter may require customized registration, the system should be designed to provide this service. The reports mentioned, and others, will serve as a basis for defining detailed system rectification and registration capabilities.

II.5.2.3 The Registration Process

The required process is described in considerable detail in Chapter 17 of the Second edition of the Manual of Remote Sensing (ASP 1983). Once an image has been preprocessed and enhanced, and an output grid selected, the precise registration displacements of a selected set of control point areas are determined. A warping model is then used to estimate the warping for every pixel in the output image, and the output data array is developed. For mosaicking of images, an additional step is required: intensity adjustments to produce a composite image without intensity seams at the joints (Zobrist 1983).

From the brief outline above, the required capabilities of the system may be identified:

- Image display

- Determining precise location of map control points

- Image processing

- Cross correlation of the control point areas

- Image recording capability to produce hard copy of images and graphics

- Terminal equipment for receipt and transmission of images, when such a capability is justified

Potential core capabilities of such a system exist in several places. What must be accomplished by the PLDS is to upgrade one or more of these systems and integrate it into the rest of the PLDS. Potentially suitable software is at a number of NASA Centers as well as the EROS Data Center. This needs to be reviewed.

II.5.2.4 Development Issues

Acceptance of the products from a central rectification facility will depend upon satisfactory performance by that facility. This requires that the product be as good as would have been done by an individual experimenter for a particular job. Accordingly, the facility must further the development of the rectification process and provide verifiable quality control. Some of the issues which have been identified are:

- The effects of the various interpolation algorithms and of multiple interpolation
- Optimum strategies for control point processing
- Application-specific methods for interpolating warping displacements from the sparse set of control points
- Methods of estimating and reporting errors in registration and the interaction of the correlation accuracy with the frequency and distribution of control points
- Large area mosaicking

II.5.3 DATA MANIPULATION AND SOFTWARE SHARING

Over the past 15 years, the hardware and software technologies for manipulation of image and related data has undergone a very significant evolution. There are several software packages and/or turnkey systems for "end-to-end" analysis available at or through various NASA Centers, universities and private companies. The term end-to-end implies starting from remotely sensed image data and other correlative data, producing interpreted output products (such as land-cover classification maps which are suitably georeferenced or rectified), and "geographic information" products (such as maps of suitability for development, proximity, erosion potential, and so forth). While the details of algorithms and capabilities for producing output products (film, display, printouts) vary, there is a large amount of commonality among the several software packages.

A survey of the circa 1980 "end-to-end" analysis software packages is found in Ramapriyan (1980). In the context of the PLDS, it is appropriate to update this survey, identify any other items of information, include the software packages which are most likely to be used by the participating institutions, and maintain a data base of lists of functional capabilities (applications program names and short descriptions) which are accessible on line to the PLDS users.

A number of new analysis functions will be required as the technology of remote sensing analysis progresses. These will be facilitated by the PLDS and are candidates for investigation. Some of these are:

- Analysis of multidimensional data from advanced instruments and multisensor data combinations;
- Classification of mixed pixels into their component parts, exploiting the high spatial and spectral resolution of TM and other data;
- Use of textural and contextual features;
- Automatic identification of "regions of interest" from the perspective of various applications;
- Generation of goal-oriented intermediate feature maps which will result in reduced effort at a user's workstation;
- Automated shape identification and size measurement (e.g., lengths of suburban streets, areas of lakes);
- Structuring the data from different sensors (or a given high resolution sensor) at different resolutions (e.g., a hierarchical structure) so that the data can be used at the resolution appropriate to the application; and
- Develop land analysis using the expert system concepts.

II.5.3.1 Software Exchange and Non-local Use Considerations

To facilitate software exchange, it is highly desirable to develop software under a common executive. This executive will provide the general data management functions of the user interface, input/output data set management, and other system services. Definitions of these interfaces will allow software to be interchanged more readily, and will allow the use of executives of various capabilities.

One evolving standard to be considered is the Transportable Applications Executive (TAE) as developed at GSFC, and the TAE/VICAR2/IBIS extension developed at JPL for image and other analysis. Both are available on the VAX 11/780. The PLDS should establish mutually agreed upon standards for future software developments and establish criteria for user friendliness to permit easy use by the infrequent user and potential remote use of the applications modules. Modules written and documented in conformance with these procedures would be eligible for the incorporation within the PLDS.

II.5.3.2 Allocation to Various Processors in the Network

Depending on the complexity of algorithms, image sizes, and data volumes handled, processing can occur on various machines in the network. These processors could be micros at Scientific Workstations (SWS), minicomputers, attached array processors or supercomputers such as the CRAY or CYBER. Flexible and expandable SWSs could be developed to maintain as much versatility as possible in local processing. The SWS could perform many functions for images of limited size (512 x 512 x 4 bands).

A larger repertoire of functions and the ability to handle larger size images (e.g., 2048 x 2048 x 8 or larger) would be available on minicomputer/attached array processor systems or mainframe computers A reasonably small subset of image/data manipulation functions could be performed on a supercomputer, such as the CRAY or CYBER. Use of such supercomputers in the network should permit both an off-loading of heavy computations and research into new computationally-intensive interpretive techniques to exploit the increased information extraction potential of high-resolution remotely sensed data (TM, SPOT, MLA, AVIRIS). Intensive computational processing on advanced high speed processors is discussed below in Section II.5.5.

Given an analysis scenario, optimal strategies will need to be worked out for allocating tasks among the various processors in the network, depending on data volumes, software availability, and computational complexity.

II.5.4 GEOGRAPHIC INFORMATION SYSTEMS (GIS)

A GIS will be defined here as a collection of procedures, computer programs, human resources, and hardware devices that supports the acquisition, storage, manipulation, and display of geographically-referenced data. Thus, the functions of data management and retrieval and generalized image analysis, which are sometimes included in the definition of a GIS, are considered here to be functions of the Data Base Management Systems (DBMSs) and Image Processing System (IPS), respectively. The reason that the more limited definition has been chosen is that it seems to match better the capabilities of the majority of existing GISs, the latter being primarily used as tools for spatial analysis and modeling. Generally, commercial GISs need to be improved in two ways: 1) integrated more effectively with DBMSs, IPSs, and statistical packages; and 2) increased in performance and flexibility of (as components in integrated systems).

In order to increase the performance and flexibility of a GIS as a component of an integrated system, the following improvements are recommended:

- A "comprehensive" set of generic GIS capabilities for land data analysis and modeling needs to be defined and implemented.

- GIS algorithms should be implemented, as appropriate, on advanced high-speed processors.

- Digital images and maps should be retrievable through DBMS queries.

- Spatial information from maps and images (such as labels of geographic features) should be incorporated into DBMS schemes so that intelligent image-based queries are possible.

II.5.5 INTENSIVE COMPUTATION SYSTEMS

A research effort based on the manipulation of a large-scale land resource data base will require the storage and processing of tremendous amounts of data. Illustrative of this is the fact that a typical Landsat Thematic Mapper (TM) scene contains almost 300 million bytes of data, compared to the 28 to 40 million bytes of data in a Multispectral Scanner (MSS) scene -- an approximate ten-fold increase. For many of the CPU-intensive operations, data volumes of this increased magnitude can only be processed in a timely fashion by using large supercomputers. Support for the PLDS program will involve widely-dispersed research facilities and will require interactive access to computationally intensive systems at various institutions. This need for shared resources, coupled with a dynamic technology, will provide the stimulus for dramatic changes in the computing environment.

With the increased throughput of new processors and with declining computing costs, there is already a trend to move some of the scientific computing workload from mainframes to personal computers or workstations. This trend will be accelerated with the availability of components which will reduce computation time of some of the lengthy functions. The availability of high-speed data transfer networks and the complementary nature of these systems (microcomputers, minicomputers, and supercomputers) create a situation in which the most efficient system can be used for each specific processing task. Effective networking of these different-scale machines is essential to the attainment of this goal.

Thus, a distributed processing system must effectively link computer systems of different capabilities with user workstations. A critical problem is the establishment of an efficient network that will transmit information, data, and processing jobs among computer systems so that the processing efficiencies of the large machines are not negated by slow data transmission time.

What follows is an introduction to existing large-scale hardware array processors, computer architectures, and supercomputers, including a discussion of some of the networking software and other issues.

II.5.5.1 Attached Array Processors

One of the most important advances in hardware is the development of architectures tailored towards image manipulation, including pipeline, array or parallel processors. For convenience, we shall simply refer to them as Array Processors (APs). In many instances, such systems provide the computational power of mainframe computers (albeit for a smaller number of users), are less expensive and yield greater accessibility and better overall turnaround time. A survey of APs, focusing on those of interest to the remote sensing community is given in Ramapriyan and Strong (1983).

This section will cover two of these array processors - the Floating Point Systems (FPS) AP and the Massively Parallel Processor (MPP), the former due to its extensive use by various organizations during the past few years and the latter due to its novelty and extensive capability for handling two-dimensional arrays.

The FPS APs

A description of the design philosophy and architecture of the FPS array processors can be found in Charlesworth (1981). The FPS-APs consist of several functional units which can operate in parallel: Program Memory, Main Data Memory, Auxiliary (Table) Memory, Address Calculation Unit, X-Registers, Y-Registers, Adder and Multiplier. Processing of vectors can proceed at a potential maximum rate of 12 million floating point operation per second (MFLOPS). The key to obtaining maximum speed is to program in such a way as to keep as many of the processing elements busy simultaneously as possible.

Among the organizations using and/or developing remote sensing image analysis systems with the FPS-APs are: the Goddard Space Flight Center (GSFC), the Jet Propulsion Laboratory (JPL), the Lawrence Livermore National Laboratory (LLNL), the Lockheed Palo Alto Research Laboratory (LPARL), The Analytic Sciences Corporation (TASC), TRW, and the U.S. Department of Agriculture (USDA).

The GSFC's LAS facility (VAX11/780-AP180V) is used to support three image analysis terminals. It has several application functions available (or being implemented) using the AP180V. Among these are: radiometric and geometric correction of Landsat (MSS and TM) and other images, general arithmetic operations on images, gradient and median filtering, FFTs, edge correlation for matching image pairs, maximum likelihood classification, clustering, principal components, and canonical transformations.

JPL uses a SEL32/55-AP120B System for producing images from the Seasat Synthetic Aperture Radar (SSAR) data. These images will be available to the PLDS. The Multi-mission Image Processing Laboratory (MIPL) also uses an AP with the VAX 11/780, where it is used for various image analysis functions as appropriate.

The Massively Parallel Processor (MPP)

The Massively Parallel Processor (MPP) (Batcher 1980; Schaefer 1982) was delivered to GSFC by the Goodyear Aerospace Corporation in May 1983. It is a 128 x 128 array of Processing Elements (PEs) in an SIMD (Single Instruction Multiple Data) architecture. It was designed primarily for image processing and pattern recognition algorithms involving local neighborhood operations on the image data. Hence, each PE is connected only to its four nearest neighbors. Aside from the large number of PEs, one of the most novel features of the MPP's architecture is a staging buffer with the I/O interface to the array. This buffer can store image data and can also reformat the data under program control, thus making it very easy to read out multi-dimensional arrays into the PEs in the form of bit-planes at 160M bytes/second.

In its present configuration, the I/O between the mass storage device and the MPP arrays is via the host computer (a VAX 11/780) and the staging buffer. Thus I/O rates are limited by the bandwidth of the host interface which is only 1 Mbyte/sec. The staging buffer is capable of handling up to 40 Mbytes/sec at its input and, between it and the array, up to 16 Mbytes/sec. Present plans are to augment the MPP with a high speed disk drive system with a 40 Mbyte/sec transfer rate connected directly to the staging buffer.

The following is an example of the MPP's processing speed. An ISODATA clustering program implemented on the MPP performs 100 iterations on a 512 x 512 x 4 image with 16 clusters within 18 seconds. This would require about 225 seconds using an AP180V. A comparison of estimated times on an FPS AP180V and the MPP for various image analysis steps is given in Ramapryian and Strong (1983).

II.5.5.2 Computer Architecture

Image processing operations can be characterized at two levels: 1) Low level, in which the same processing function is applied to all pixels, such as filtering, noise removal, and geometric correction. These are suited to a SIMD structure such as is provided in the array processors discussed above; 2) Image analysis, in which the image cannot simply be considered as a large matrix of pixels, such as occurs in feature extraction or other pattern recognition processes. These may involve many operations on a common data base, and are well suited to a parallel Multiple-Instruction-Multiple-Data (MIMD) structure. Some recent designs can be reconfigured to run in either mode. A review of current thinking is given in Reeves (1984),and in IEEE-CS 498 (1983).

Geographic information systems and operations on spatial data will be characterized by processes involving large neighborhoods, feature extraction, and other inferential processes. Therefore, the PLDS must expect to utilize, and potentially develop, MIMD systems aimed toward the geocoded data processing problems.

A joint research effort between JPL and the California Institute of Technology involves investigating the hypercube arrangement of local processors. The hypercube is widely recognized as a very efficient arrangement of microprocessors for a wide variety of problems (Kushner and Rosenfeld, 1983). It consists of a collection of 2^p independent computers running asynchronously, with each node (machine) connected to p others; the connection is topologically is equivalent to that of the corners of a (hyper)cube in a p-dimensional space. The primary goal of this work is to solve "real" problems on "real" hardware, thereby proving the concept of hypercube architecture for concurrent processing. A 64-node prototype machine is now in operation, and continued investigation and implementation of applications are underway. The PLDS should be in a position to take advantage of this architecture at the appropriate time.

A nearer-term development are VLSI processors which can be incorporated within, or as an adjunct to, micro and minicomputers. These are being designed to minimize number-crunching operations for defined processes. As one example, a multiple-summation chip is nearing completion which will serve for filtering, cross correlations, and interpolations, and will be implemented on a VAX 11/780 (Nathan, 1983). This will expedite the bulk of these typically time-consuming processes by a factor of 100 or more. These will be available to the PLDS, and will make these operations practical on the larger workstations.

II.5.5.3 Supercomputers

Two supercomputer systems that exist at NASA facilities and could be made available through the PLDS are the Numerical Aerodynamic Simulator Processing System Network (NASPN) at NASA/Ames, and the NASA High Speed Computing Facility (NHSCF) at NASA/Goddard. They are described below.

The ongoing Numerical Aerodynamic Simulator (NAS) Program has as its technical objective the provision of the greatest feasible computational capability for CPU-intensive research at ARC. NAS will provide a significant increase in computational capacity that will be accessible to both local and remote users.

As part of this same program, ARC is developing a Numerical Aerodynamic Simulator Processing System Network (NASPN). The objective of the NASPN is to provide an integrated network of state-of-the-art computer systems and software designed to provide the full range of functional and performance capabilities to support the many facets of numerical simulation, including remote sensing analysis needs. When operational, the NASPN will support large-scale scientific processing, code development, graphics display, data storage and management, and associated ancillary processing.

The NASA High Speed Computing Facility (NHSCF) at GSFC is a major computing facility for supporting NASA Earth Science and Applications research programs at NASA Centers and universities. It provides high-volume data storage, cataloging and access capabilities needed to support computational models of physical processes, and a broad range of computing services to users, whether they are at GSFC or other institutions. The NHSCF consists of a Cyber 205 vector processor with an Amdahl 470 V/6 front end, an Amdahl 470 V/7B general processor, 110 gbytes of on line mass storage, high and low speed remote terminals, graphics display system, microfiche and the I/O units.

The NHSCF is working toward simplifying the complexities of supercomputer use while maximizing scientific productivity and will be establishing network links with other NASA and NASA-funded computing resources to provide complementary support to the scientific community.

These two programs are excellent examples of the kinds of efforts which the PLDS must encourage, providing valuable and unusual capabilities to a large user community.

II.5.6 ARTIFICIAL INTELLIGENCE AND THE PLDS SYSTEM DESIGN

A problem with existing and projected conventional data analysis and management systems is that they are time-consuming and require a high level of user expertise. This reduces the time available for the primary research and management tasks at hand. While many areas of artificial intelligence (AI) research are still struggling to show results, two specific topics within AI have now done so: "expert systems" that simulate human knowledge understanding and decision processes, and natural language (English) processing. Software packages for implementing these two AI techniques are currently becoming available from vendors.

AI techniques have the potential to reduce the level of user interaction required to use the PLDS. Specific areas where improved performance may be possible within the PLDS are in resource allocation management within a distributed environment, data management (archiving, distributed data concurrency control, etc.), and data analysis. Knowledge based expert systems can provide intelligence to the PLDS both in supporting independent operations of the system, and in providing to the user automated data analysis capabilities using knowledge and heuristic rules. Natural language processors can carry out automated translations between English sentence inputs and the command syntax required to run DBMS and other software packages.

Natural language processing is now achievable. Presently, there are several natural language software packages on the market that are intended to provide English interaction between data base management systems and users. Of these, the best known are the IBM compatible Intellect (Artificial Intelligence Corporation) and the DEC-compatible Themis (Frey Associates) systems.

Expert systems have also been successfully created. System development tools are now available from many vendors, including Symbolics, Intelligenetics, Technolege, and Xerox, and enjoying widespread use as time-saving aids in the construction of expert systems. However, unlike natural language processing, they require a major effort on the part of the user to construct a base of procedural knowledge describing how to do analysis tasks. Effective formalization of the knowledge domain for various science and information disciplines remains technically challenging, time consuming, and risky.

For PLDS, AI modules are proposed that would reside in and support the operation of two major subsystems; the Data Management Subsystem (DMS), and the Intensive Computational Processing Subsystem (ICPS). It is anticipated that their presence would offer significant enhancements over current system operational capabilities in the areas of system and subsystem operational management, natural language data base queries, and data analysis. AI concepts for the DMS are covered in Section II.3.2.

The ICPS is that component of the PLDS where all computationally intensive processing will be performed. A knowledge-based expert system can be developed to manage and control the distributed resources within the ICPS. The distributed processing power of this subsystem, when aggregated, is expected to be on the order of 100 to 1000 million operations per second (MIPS). Due to its distributed nature, there will be unique problems regarding effective utilization of its resources.

An ICPS expert system could control the utilization of the available resources through the dynamic selection and pairing of individual computer processors with active processing tasks. Factors to be considered would be the size of processor required, the location and availability of appropriate software, the existing loads at the candidate processors, and the minimization of communications delays. Whether or not an expert system is developed, this is a long-term problem that must be eventually resolved if the overall system is to be effectively utilized.

II.6 SYSTEM SUPPLIED SUPPORT SERVICES AND ADMINISTRATION

II.6.1 OVERVIEW

The PLDS network will interface with diverse computing environments at installations throughout the country. It is essential to the orderly operation of the system that a complete and accurate accounting and control system be implemented. This is necessary to insure that data deliveries to users are timely, that the host or "node" computers are not overloaded, and that adequate security and reliability are maintained. Each of the node computers must maintain an accounting system to keep track of user access, data distribution, and any other features such as value-added services. Regardless of who pays the bills, accurate records of usage and costs for each element of the system and of the overall system cost must be maintained. This is essential to planning for future expansion of the network, for supporting budget requests, for evaluating system utilization patterns, and so forth.

The overall responsibility for system accounting should reside at the 'prime node' computer that is responsible for overall system control (see Fig. II.4.3, Section II.4).

II.6.2 GENERAL ACCOUNTING

The 'prime node' is the computer system that maintains the master catalog system for the PLDS. This is the logical place to maintain overall accounting responsibility. The central accounting facility will be responsible for assigning user access privileges for the PLDS. It will also gather accounting information from the various Level I nodes and prepare an integrated usage report, both to 'fine tune' the network design and to determine billing and budgeting.

The various Level I nodes will maintain local accounting systems. Most mainframe systems already support some type of user-accounting facility which should be able to be adapted to the PLDS needs. This data will be transfered to the prime node via the network on a regular basis for inclusion in the overall accounting report. The Level II and III nodes are the system 'users' and need not have local accounting inputs to the overall PLDS. They will be the users that are accessing the Level I machines.

II.6.3 SECURITY

Security will be important to the PLDS for several reasons. First, it is essential that only authorized users have access to the system to ensure that critical resources are used only for authorized purposes. Second, portions of the data and software may be proprietary or copywrited. These rights must be protected.

The implementation of security is difficult problem. The first level will be the access security already used by most systems (e.g., user account assignments, passwords, etc.). Assignment of these will be the responsibility of the central accounting facility. This should be the only means of gaining access to the system.

Security of programs or data that are not public domain must be the responsibility of the owning facility. In any case, the status of all data or programs must be clearly stated in the catalogs. There must also be provisions in the data request system to identify data or copywrited software that cannot be distributed without special permission or license.

II.6.4 ACCESS CHARGES

A basic decision that must be made is how, or if, the user is to be charged for using the system. For example, a university doing research on a contract for one of the sponsoring agencies may not be charged for access. A researcher who is not supported by a PLDS sponsoring agency, but is authorized for access to the system, may be charged for the access.

Access charges are distinguished from charges for data or services. Charges for data or services should be determined by the archival agency for that particular data or service (e.g., EROS Data Center). Access charges should be uniform throughout the PLDS. Some charges, such as CPU time, may vary according to the particular type of machine and the costs associated with its operation (Cray versus VAX for example). In all cases, the accounting system should be capable of providing an estimate of charges for a particular work request in advance.

Accounting for access charges will be done by the Level I node being accessed, and transmitted to the prime node for inclusion in the overall report. Any billing that may be required should come from the prime node accounting system. Thus, a user will receive one statement of system use regardless of how many different nodes are accessed. The user should receive a summary statement of PLDS system usage whether there are any charges involved or not. It is also highly desirable that there be an on-line query system for accounting purposes. This would allow an individual user or user organization to query for resource uses, expenditures, etc. without waiting for a monthly or quarterly statement.

II.6.5 DATA CHARGES

Much of the data that is to be accessed through the PLDS must be purchased. It is essential that the catalog system reflect the price of data, whether it is a minimum fee to cover reproduction, media (tape, diskette, etc.), license fees, or other costs. These charges will be reported to the prime node for inclusion in the overall accounting report and for billing purposes. It is essential that the ordering system for data and services give the user real-time feedback about the cost of an order.

II.6.5.1 Value-Added Charges

Value-added charges may take two forms. The first is where the archival agency provides different forms of data (e.g., raw and rectified) that require an increase in production cost. The exact nature of the process applied to the data should be specified, as well as the charges involved.

The second type of value added-charge would derive from services rendered by a particular institution (e.g., registration of multi-sensor data, digitization of products). In this case, there should be a special catalog section for the service, telling what the service is, how it works and what it costs. The ordering system should allow accounting for use of these services, even though the service itself may be performed off line from the network. Again, the accounting records for the service will be transferred to the prime node for inclusion in the overall report.

II.6.6 ELECTRONIC MAIL

An electronic mail system is a very useful administrative tool in a network environment. It can be used to send monthly system usage statements, notices of system enhancements or changes, and for general communications between the various users of the network. NASA Centers and contractors can gain access to the NASA Telemail system, a service purchased by NASA from GTE Telenet. It should be practical and cost-effective to use this system for mail on the PLDS system. This will allow users immediate access to mail features, document transmission, etc. As the PLDS matures, it may be desirable to add mail features that are internal to the network. For the PLDS, Telemail provides a ready-to-use system complete with accounting and security features already in place.

II.6.7 SUMMARY

The PLDS should have an integrated accounting system with centralized accounting responsibility. This central authority will be responsible for authorizing system access, coordinating accounting procedures at nodes, providing cost estimates in response to user requests, and preparing periodic reports for overall system usage. It will also be responsible for coordinating system security.

The individual Level I nodes will be responsible for local accounting software which must be compatible with the guidelines established by the coordinating authority. The Level I nodes will also be responsible for establishing charges for local value-added services and data. System access charges should be established by the system accounting authority. Charges for CPU time and other local resources should be coordinated with the system accounting authority.

REFERENCES

Cited Papers:
(Number at end of citation in parentheses indicates section reference was cited in.)

ASP (American Society of Photogrammetry), 1983. Manual of Remote Sensing, 2nd ed. pp. 338-753. (II.5)

Batcher, K.E., 1980. Design of a Massively Parallel Processor, IEEE Trans. on Computers, September 1980. (II.5)

CODMAC (Committee on Data Management and Computation) 1982. Space Science Board, National Academy of Sciences, Data Management and Computation, Vol 1: Issues and Recommendations, National Academy Press. (1)

Charlesworth, A.E., 1981. An Approach to Scientific Array Processing: The Architectural Design of the AP-120B/FPS-164 Family, Computer, IEEE Computer Society, September 1981, pp. 18-27. (II.5)

Costanzo, D.J., 1983. Mapping Applications of Video Disc Technology, National Computer Graphics Association (NCGA) Annual Conference, June 1983, Chicago, Ill. (II.1)

Fujimoto, B., 1981. User Requirements for NASA Data Base Management Systems, Part One: Oceanographic Discipline. JPL Publication 81-50. (II.1)

Hunter, G.M., and K. Steiglitz, 1979. Operations on Images Using Quadtrees. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-1 145-153. (II.1)

Goldberg, M., 1984. Topological Grid Structure: A Data Structure for Earth Science Modeling, to appear in the Proceedings of the International Symposium on Spatial Data Handling. (II.1)

Gwynne, M.D., 1982. The Global Environment Monitoring System (GEMS) of UNEP. Environmental Conservation 9(1):35-41. (2)

Hilbert, E.E., 1977. Cluster Compression Algorithm. A Joint Clustering/Data Compression Concept. JPL Publication 77-43. (II.1)

IEEE-CS 498, 1983. Proceedings of a Workshop on Computer Architecture for Pattern Analysis and Image Database Management, Pasadena, CA, October 1983, IEEE Computer Society Order No. 498. (II.5)

Kushner, T.R., and A. Rosenfeld, 1983. Interprocessor Communication Requirements for Parallel Image Processing, in IEEE Computer Society Order No. 498, pp 177-183. (II.5)

Landsat Technical Working Group, 1978. User Compatible Tape (CCT) Format Family Requirements, CCB-CCT-0001. (II.5)

Landsat Technical Working Group, 1979. The Standard CCT Family of Tape Formats, CCB-CCT-0002. (II.5)

Lohman, G.M., 1981. Generic Functional Requirements for a NASA General Purpose Data Base Management System. JPL Publication 82-6. (II.1)

Lohman, G.M., J.C. Stolzfus, A.N. Benson, M.D. Martin, and A.F. Cardenas, 1983. Remotely-Sensed Geophysical Data Bases: Experience and Implications for Generalized DBMS. ACM Sigmod Record, 13(4):146-160. (II.1)

NASA (National Aeronautics and Space Administration), 1983a. Land-Related Global Habitability Science Issues. Land-Related Global Habitability Sciences Working Group. NASA Technical Memorandum 85841. (2)

NASA (National Aeronautics and Space Administration), 1983b. Global Biology Research Program; Program Plan. NASA Technical Memorandum 85629. (2)

NASA (National Aeronautics and Space Administration), 1982. Proceedings of the NASA Workshop on Registration and Rectification, N. Bryant, Ed., JPL Publication 82-23. (II.5)

NASA (National Aeronautics and Space Administration) 1981a. Data Base Management System Analysis and Performance Testing with Respect to NASA Requirements, Goddard Space Flight Center, Technical Memorandum 83942. E.A. Martin et al. (2)

NASA (National Aeronautics and Space Administration) 1981b. Data Base and Data Base Systems Related to NASA's Aerospace Program. Goddard Space Flight Center, Technical Memorandum SP-7045. (2)

Nathan, R., 1983. Large Array VLSI Filter, in IEEE Computer Society Order No. 498, pp 15-21. (II.5)

Peuquet, D.J., 1983. A Hybrid Structure for the Storage and Manipulation of Very Large Spatial Data Sets. Computer Vision, Graphics, and Image Processing. (II.1)

Ramapriyan, H.K., 1980. Survey of Image Analysis Software. Eastern Regional Remote Sensing Applications Center, NASA Goddard Space Flight Center. (II.5)

Ramapriyan, H.K. and J.P. Strong, 1983. Applications of Array Processors in the Analysis of Remote Sensing Images, Seventeenth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, May 9-13. (II.5)

Ramey, T.L., R.R. Brown, M.A. Melkanoff, and G. Rodriquez-Ortiz, 1983. The EKLA Information Model, Information Systems. (II.1)

Reeves, A.P., 1984. Parallel Computer Architectures for Image Processing, Computer Vision, Graphics, and Image Processing, 25(1):68-88. (II.5)

Rice, R.F., 1979. Some Practical Universal Noiseless Coding Technique, JPL Publication 79-22. (II.1)

Schaefer, D.H., et al., 1982. The Massively Parallel Processor, AIAA Journal of Guidance, Control and Dynamics, May-June 1982, pp. 313-315. (II.5)

Simonett, D.A., T.R. Smith, W. Tobler, D.G. Marks, J.E. Frew, and J.C. Dozier, 1978. Geobase Information System Impacts on Space Image Formats, Remote Sensing Unit, University of California at Santa Barbara, SBRSU Technical Report 3. (II.5)

Smith, P.H., L.A. Trenish, and L.V. Novak, 1983. Pilot Climate Data Systems: A State of the Art Capability in Scientific Data Management. (II.1)

Wheelock, C.B., 1983. Data Collection, Processing, and Management in Support of the Biospheric Research Program Boreal Forest Experiment. Information Sciences Branch. Earth Sciences and Applications Division. Johnson Space Center. (I.1)

Zobrist, A.L., 1983. Technology for Large Digital Mosaics of Landsat Data, Photogramm Eng. and Remote Sensing, 49(9):1325-1335. (II.5)

Background References

Aeronutronic/Ford, 1976. User Data Dissemination Concepts for Earth Resources, prepared for NASA/ARC, R. Davies et al., NASA Contractor Report CR-137904.

Bernhard, R., 1981. Giants in Small Packages, IEEE Spectrum, February, pp. 39-44. (II.5)

Blaser, A.(Ed.), 1980. Data Base Techniques for Pictorial Applications. Springer-Verlag Lecture Notes in Computer Science #81.

Computer Science Corporation, 1983. CSC/TM-83/6091. Very Large Data Base Management System (VLDBMS) Current Technology Review. G.B. Page et al.

Computer Sciences Corporation, 1983. CSC/TM-83/6154, An Evaluation of Data Base Management Systems for Use at the Science and Applications Computing Center.

Computer Science Corporation, 1979. Applications Data Service Generic Requirements, for GSFC, Contract NAS5-24300, Task 138.

Computer (Special Issue on Array Processor Architecture), IEEE Computer Society, September 1981, pp. 8-57. (II.5)

Computer (Special Issue on Computer Architectures for Image Processing), IEEE Computer Society, January 1983, pp. 10-80. (II.5)

Computer Technology Associates, Inc., 1983. Data Base System Architecture Study.

Freeman, H., 1974. Computer Processing of Line Drawing Images, Computing Surveys 6(1):57-97.

Goldberg, M., 1982. Geographic Information Systems at the Goddard Space Flight Center, NASA Technical Memorandum 83941.

LARS, 1977. Requirements of a Geographic Information System for Corn Production and Distribution, M. Baumgardner et al., Contract Report 112977.

Marble, D.F. (Ed.), 1980. Computer Software for Spatial Data Handling, Vols, 1, 2, and 3, IGU Commission on Geographical Data Sensing and Processing, Ottowa, Ont.

Nagy, G. and S.G. Wagle, 1979. Geographic Data Processing. Computing Surveys 11(2):139-181.

NASA, 1981. Applications Data Service Candidate Standards and Guidelines, T. Burns et al., Mitre Corp for GSFC, WP-81W615-1981.

NASA, 1979. Data Base Management Systems Panel Workshop: Executive Summary, N. Bryant, Ed., JPL Publication 79-70.

NASA, 1980. Data Base Management Systems Panel, Second Workshop Summary, G. Lohman, Ed., JPL Publication 80-50.

NASA, 1981. Data Base Management Systems Panel, Third Workshop Summary, U. Urena, Ed., JPL Report 81-52.

NASA, 1980. NASA End-to-End Data System Data Base Management System Functional Requirements.

NASA, 1979. Landsat-D User CCT Tape Format, Landsat LTWG, FOR-LSD-0001A.

NASA, 1982. The Multispectral Imaging Science Working Group Final Report, S. Cox, Ed., NASA Conference Publication 2260.

NASA, 1980. OSTA Data Systems Planning Workshop Report, R. DesJardin, Ed., ADS Study Office, GSFC (Draft 3).

NASA, 1981 Survey of Standards Applicable to a Database Management System, J. Urena, Ed., JPL Publication 81-88.

Ocean Data Systems, Inc., 1980. Ground Processing Concepts for the Operational Land Observing Systems, for GSFC, Contract NAS5-26025.

Peuker, T.K., and N. Chrisman, 1975. Cartographic Data Structures. The American Cartographer 2(1):55-69.

Ramapriyan, H.K., 1982. Registration Workshop Report, in Multispectral Imaging Science Working Group Report, Vol. 3, NASA Conference Publication 2260.

Rhind, D.W., 1977. Computer Aided Cartography. Transactions of the Institute of British Geography New Series 2(1):71-96.

Shapiro, L.G., 1979. Data Structures for Picture Processing: A Survey, Computer Graphics and Image Processing, 11(1):162-184.

USFWS (United States Fish and Wildlife Service), 1977. Guide to Land Cover and Use Classification Systems Employed by Western Government Agencies, FWS/OBS-77/05.

USFWS, 1977. User Needs Assessment for an Operational Geographic Information System, FWS/OBS-77/21.

Washington State University, 1978. A Case Study of the Operational Aspects of the Natural Resource Information System (NRIS) used for Natural Resource Planning, D.Steingraber, Computing Service Center, Wash State Univ, prepared for DOI/BIA.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM-86250		
4. Title and Subtitle The Pilot Land Data System: Report of the Program Planning Workshops		5. Report Date July 1984 6. Performing Organization Code EI-5
7. Author(s) Pilot Land Data System Working Group		8. Performing Organization Report No.
9. Performing Organization Name and Address Information Systems Office Office of Space Science and Applications National Aeronautics and Space Administration Washington, DC 20546		10. Work Unit No. 11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546		13. Type of Report and Period Covered Technical Memorandum 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Under the sponsorship of the National Aeronautics and Space Administration's (NASA) Office of Space Science and Applications' (OSSA) Information Systems Office (ISO), the Universities Space Research Association (USRA) assembled a Working Group and coordinated a series of planning workshops during fall 1983 and winter 1984 to draft an advisory report which NASA will use in developing a program plan for a Pilot Land Data System (PLDS). The purpose of the PLDS is to improve the ability of NASA and NASA-sponsored researchers to conduct land-related research. The goal of the planning workshops was to provide and coordinate planning and concept development between the land-related science and computer science disciplines, to discuss the architecture of the PLDS, requirements for information science technology, and system evaluation. This report presents the findings and recommendations of the Working Group.

The goal of the pilot program is to establish a limited-scale distributed information system to explore scientific, technical, and management approaches to satisfying the needs of the Land Science community. The PLDS would pave the way for a Land Data System to improve data access, processing, transfer, and analysis, thus fostering an environment in which land sciences information synthesis can occur on a scale not previously permitted because of limits to data assembly and access.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
land remote sensing, information systems, communications network, geographic information system, data base management system, earth sciences, global habitability	Unclassified - Unlimited Subject Category 82

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	170	A08

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