Earth Sciences Data Lifecycle Study

December 2001

Original

Written By:

Richard McKinney EDC DAAC Operations Manager	Date	Timothy Smith Senior Data Information Analyst	Date

Approved By:

Thomas Kalvelage EDC DAAC Manager	Date	Mathew Schwaller ESDIS Science I/F Manager	Date

TABLE OF CONTENTS

1 INTRODUCTION 3
2 SCIENCE REQUIREMENTS FOR LTA 3
3 LONG-TERM ARCHIVE LEVELS OF SERVICE 5
4 REAL-WORLD EXAMPLES 6
5 EARTH SCIENCES ARCHIVE LIFE CYCLE 7
5.1 PRODUCT GENERATION 9
5.2 RECEIVE DATA INTO THE ACTIVE ARCHIVE 9
5.3 TRANSITION TO LONG-TERM ARCHIVE (LTA) 11
5.4 LONG-TERM ARCHIVE 11
5.4.1 Data Format(s) 12
5.4.2 Documentation 12
5.4.3 System Requirements/Performance 12
5.4.4 Data Volumes 12
5.4.5 User Services 12
5.4.6 Science Requirements 13
5.4.7 Archive Media 13
5.4.8 Ancillary Data 13
5.4.9 Algorithms 13
5.4.10 Metadata 13
5.4.11 Output Products 14
5.4.12 Data Management 14
5.4.13 Security 14
6 LONG-TERM ARCHIVE REQUIREMENTS 14
6.1 REVIEW AND RE-USE OF ACTIVE ARCHIVE DETAILS 14
6.2 LONG-TERM ARCHIVE APPROVAL PROCESS 14
6.2.1 NSLRSDA AAC 14
6.2.2 PRB 15
6.2.3 DART 15
6.2.4 Data transcription/media migration 16
7 STUDY RECOMMENDATIONS 16
APPENDIX A 20
APPENDIX B 23
APPENDIX C 27
APPENDIX D 28
APPENDIX E 32
ABBREVIATIONS AND ACRONYMS 35

1 Introduction

Funding for the development of this report was provided under the sponsorship of NASA's NewDISS formulation activities. As such, the primary purpose of this report is to address the specific needs of NewDISS to describe and understand the life cycle for NASA's Earth science data products. The ultimate goal of such understanding is to ensure the safe handling of NewDISS-era data products as they migrate from data providers to active archive and long-term archive, even as numerous individuals and institutions take responsibility for the product along the way.

With respect to the land sciences, the U.S. Geological Survey is the federal agency with the statutory responsibility for long-term archive (LTA) of the nation's cartographic and Earth science data. In a 1993 Memorandum of Understanding (MOU, see **Appendix A**), NASA and USGS agreed that the USGS would assume responsibility for land sciences and related data collected under NASA's Earth Observing System (EOS) and allied missions. The MOU specifically calls for the USGS to take on the following activities.

> "…USGS will fund long-term archive functions. USGS will fund archive and distribution functions, including operations and maintenance costs for EOS and related data more than 3 years old…USGS will develop with NASA an EOS Archive Data Management Plan to ensure transition of EOS and other appropriate land remotely sensed data from the short-term archive to the long-term archive.

Under the direction provided by the MOU cited above, this document assumes that NASA's land sciences data will ultimately transition to the USGS National Satellite Land Remote Sensing Data Archive (NSLRSDA, see **Appendix B**). NSLRSDA is an initiative operated by United States Geological Survey's (USGS) Earth Resources Observation Systems (EROS) Data Center (EDC).

2 Science Requirements for LTA

The rationale behind, and the need for, a national infrastructure for LTA of Earth observation data have been the topics of study by various scientific and policy-making groups. Most significantly, the U.S. Global Change Research Program (USGCRP) evaluated the purpose of a LTA program for Earth observation data and derived products, lessons learned from current and past experience, and the vision, guiding principles, and essential functions necessary for the success of such a program. The USGCRP concluded (USGCRP 1999) that a national LTA program must:

- Enable and facilitate the best possible science and highest-quality assessment for making policy and business decisions
- Document the Earth system variability and change on global, regional and local scales, building and maintaining a high quality base of data and information and establishing the best possible historical perspective critical to effective analysis and prediction
- Ensure archive holdings, facilities, and services are actively promoted and made readily available to the maximum number of users
- Enable and facilitate future research.

A recent report from the National Academy of Sciences (NAS) Committee on Earth Sciences (CES 2000) has also contributed to the definition of requirements for essential data system services need ensure a long-term satellite data record in support of climate research. The CES report defined following principles to help ensure the preservation of the climate record:

- Accessible and policy-relevant environmental information must be a well-maintained part of our national scientific infrastructure
- The federal government should 1) provide long-term data stewardship, 2) certify open, flexible standards, and 3) ensure open access to data
- Because the analysis of long-term data sets must be supported in an environment of changing technical capability and user requirements, any data system should focus on simplicity and endurance
- Adaptability and flexibility are essential for any information system if it is to survive in a world of rapidly changing technical capabilities and science requirements.
- Experience with actual data and actual users can be acquired by starting to build small end-to-end systems early in the process
- Multiple sources of data and services are needed to support development of climate data records
- Science involvement is essential at all stages of development and implementation.

Based on review and reporting from the USGCRP and the CES, the scientific guidance for LTA can be divided into three broad categories. Overall, the scientific community has recommended that the LTA address the issues of **setting priorities** for the LTA, identifying the **data content** of the LTA (including documentation required for long term use of the data), and specifying the **stewardship and data services** to be provided by the LTA.

This document focuses primarily on the Earth science data lifecycle and LTA infrastructure, that is, on stewardship and data services issues. Brief treatment is given to the policies and procedures related to LTA data content and priorities (what products transition to the LTA and when). It is assumed that the issue of what NASA products transition to the LTA (and when this takes place) will need

to worked as an on-going collaboration between NASA, USGS and the Earth science community (see **Appendix C**).

3 Long-Term Archive Levels of Service

As stated above, and consistent with Agency and Federal policies, NASA's Earth observation data will ultimately transition to the long-term stewardship of the Nation's operational agencies. The stewardship of these data can be described as a set of LTA Levels of Service (LOS). The paragraphs below describe the LOS provided by the USGS for data products within National Satellite Land Remote Sensing Data Archive (NSLRSDA). Subsequent sections of this document describe in detail the life cycle steps leading to the LTA, and the specific requirements on data products and data providers that enable the successful archive of NASA's ESE.

For land remote sensing data, it is the USGS Archive Program that manages the Nation's land remote sensing and cartographic treasures. NSLRSDA, described in detail in Appendix B, has the mission to preserve national scientific and intellectual resources for future generations by maintaining a common set of basic spatial data. A key goal of NSLRSDA is to maintain these archival resources for the future study of water, energy, and mineral resources, to permit understand the effects of natural disasters, to protect the quality of the environment, and to contribute to the Nation's economic and physical development. Central to this mission is the reliable collection, maintenance, and distribution of a record of the Earth's surface; a record that is comprehensive, historical, permanent, and impartial.

USGS EDC has a 25-year history of maintenance, preservation and exploitation of satellite land remote sensing data. The USGS have established a set of LOS for data and documentation within the EDC archives as described below.

Level 1 Deep Archive Services. A deep archive level of service involves only offsite back up storage of a data set. Data products are only generated for the primary archive facility on an as needed basis. All data documentation requirements are retained by the primary archive. This service is provided to data providers to satisfy NARA (National Archive & Records Administration) requirements. The data provider retains archive responsibilities for the data. Level 1 archiving could be the initial phase for further archive developments.

Level 2 Collection/Directory Level Archive Services. This service involves collection level metadata that could be hosted on a web accessible server that may include Clearinghouse services. The collection level metadata may point to online or near line data files delivered by the data provider for distribution. Data products are "copy only" data files and no format or value added processing is performed on the data. This service is appropriate for homogeneous data sets that involve common data processing routines and algorithms.

Level 3 Data Set/Inventory Level Archive Services. This service involves the delivery of a complete data product suite for each data set. The data sets are delivered to the archive where they are ingested by data type either online or near-line for data distribution. Data access is provided through database services or web services that assure user access and product distribution options on hard media or through network access.

Level 4 Data System Level Archive Services. This service involves the delivery of the raw data and the system to process on demand/user specific products tailored to the data requestor. The service is comparable to the Landsat 7 National Land Archive Products (NLAPS) system in operations at the USGS/EROS Data Center. FGDC Metadata is searchable through a supporting database that is used to manage the archive and produce the data products. Product specific FGDC metadata are also generated on the fly for each product created for the user. Other support data files are also made available to the user via hyper links and ftp services (algorithms, ephemeris, and ancillary data files supplied by the data provider). Users provide specific data parameters for their products and receive the data are hard media or through network services.

4 Real-World Examples

USGS EDC has had 25 years of experience in maintenance, archive and distribution of land remote sensing data. A few examples of applying archive LOS to satellite data products are provided below.

Level 1 Archive service provided for the National Climatic Data Center (NCDC). EDC serves as a deep/backup archive for microfilm records. Data access is limited to NCDC and their approved agents/representatives only.

Level 2 Archive service provided for the Global Land Cover Characteristic (GLCC) data set. GLCC data will be hosted on a USGS web site with access to static files. Metadata Information is supported on static web pages only.

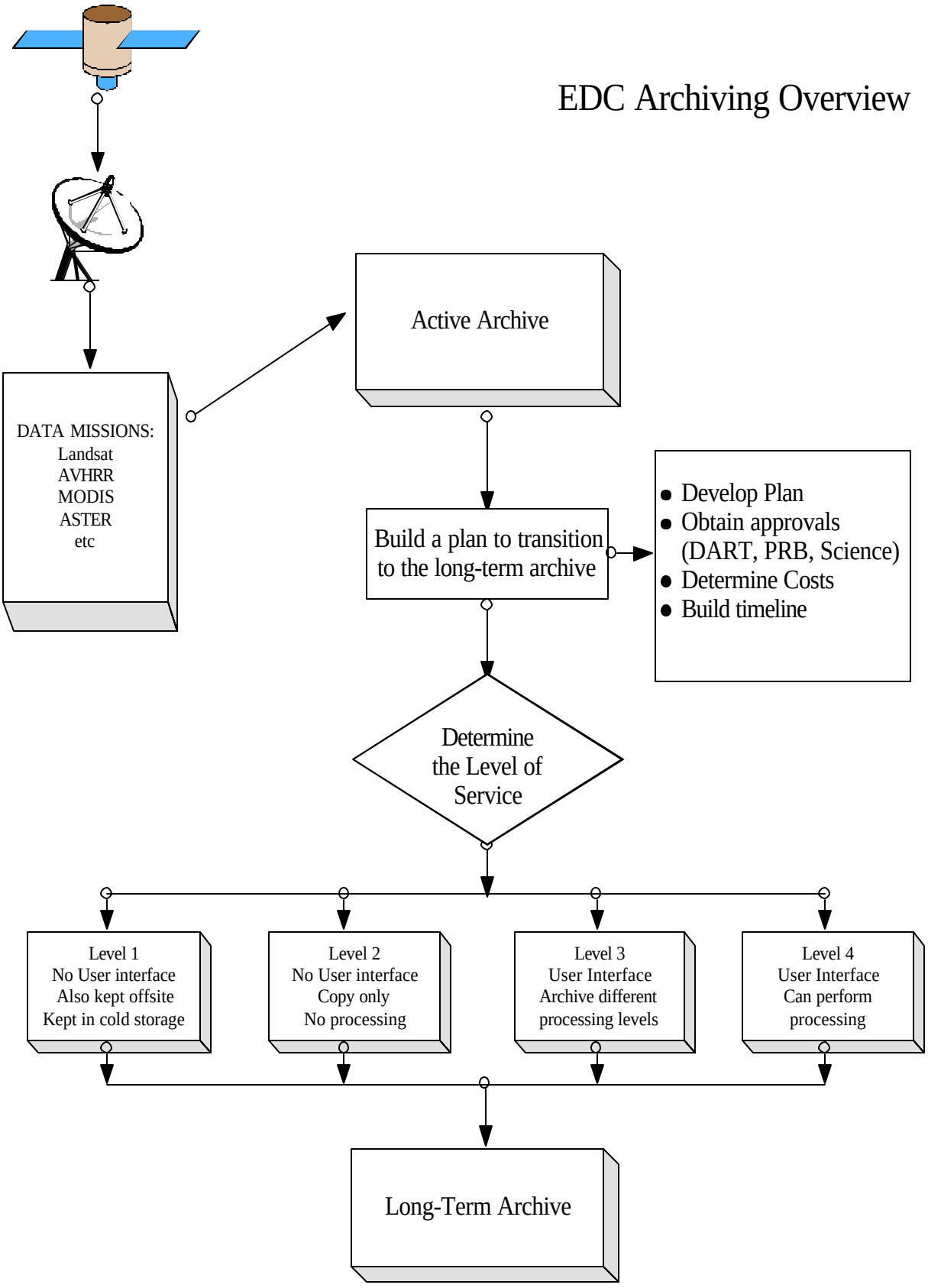
Level 3 Archive service provided for the Corona (declassified imagery) data collection. The imagery is supported by researchable information systems (Global Land Information System and Earth Explorer) and order options are

available for a limited set of products. FGDC Metadata is available at the collection level and at the entity level if appropriate.

Level 4 Archive service provided for the Landsat 7 system on the USGS side of EDC. Full service capabilities provided by Earth Explorer with custom product options, FGDC metadata generation on the fly and available for customer products. This service is tightly coupled with a product generation system that can be tailored to the unique needs of separate customers.

5 Earth Sciences Archive Life Cycle

EDC Archiving Overview



The Earth Sciences Archive Lifecycle is accomplished in the following segments: Product Generation; Receipt of data into the active archive; Transition to the LTA. Each of these is described in greater detail below.

5.1 Product Generation

Product Generation is typically the "Data Missions" component of a science mission and active archive. The requirements for this activity focus on the needs of the original customer and involve all of the ground processing steps from data capture to product delivery and population of the active archive. A primary goal of this study includes providing guidelines to current/future missions to smooth out the transfer of active archive data and supporting metadata into the LTA.

5.2 Receive data into the active archive

Generally the data will be inserted into an active archive for use for the life of the mission. However, in some cases the mission may be over, or for some other reason, the data provider wants the data to go directly into a long-term archive. Arrangements such as these are often outlined in memos of understandings (MOUs). **Appendix A** is the current MOU in effect between the USGS and NASA.

1. Ingest

The data provider shall generate physical media formatted as specified in the Interface Control Document (ICD) between the data provider and the EROS Data Center. It is expected that this will involve maintaining the original media containing the data products and cataloging to allow for internal access and backup in the case of catastrophic on-line failure.

The ICD shall be specific as to include such items as media, format, organization, and tape label format. The active archive system reads and verifies the contents of the data as it is ingested and creates a log of results. In addition, if there are browse associated these data files they will be copied to a browse server. The metadata is loaded into the proper inventory database location.

The EDC shall receive Directory Interchange Format (DIF) and Guide document background information from the provider, generate the DIF files and Guide documents and populate the necessary directories or inventories to allow user access to this information.

2. Archive

The EDC system shall store the data sets in off-, near-, or on-line storage.

3. Search and Order

The EDC client will be configured to allow the data product queries according to the following basic criteria: Parameters, Data set, Sensor, Data center, Spatial, Temporal, Source, and Processing level. The EDC client shall be configured to allow data product queries according to the extended search criteria to be defined by the Scientist. The EDC client shall be configured to allow integrated browsing of data products. The EDC client shall be configured to allow access to DIF and Guide information for the data collection.

The EDC system shall receive queries from and send responses to the server, according to protocol as specified in the EOS Data Gateway Messages and Development Data Dictionary; V0 and ASTER/ECS Message Passing Protocol Specification. The EDC system shall pass initial order information from the EDG to DORRAN. DORRAN shall send data product orders to PDS for production. DORRAN shall accept requests for product order status. DORRAN shall provide order-tracking functionality for orders.

4. Product Distribution

EDC shall distribute metadata information along with the data file(s). The format shall be documented in the detailed design phase. EDC shall distribute a Readme file along with data file(s). The format shall be documented in the detailed design phase. EDC shall distribute the datasets via the following methods/media: ftp, 8mm, CD-ROM, or DVD. The EDC system shall provide a tape label for the distribution media. The format shall be documented in the detailed design phase. The EDC system shall provide a shipping label for the distribution media. The format shall be documented in the detailed design phase.

5. Data and Metadata Management

Metadata for the data collection shall contain and be formatted as defined in the ICD. EDC shall check data and metadata database entries upon ingest. EDC shall provide a graphical user interface (GUI) to access and modify metadata for the data collection. The EDC shall provide an internal method of accessing data from the archive.

5.3 Transition to Long-Term Archive (LTA)

In order to transition a database from the active archive to the LTA a plan to transition the database needs to be generated. Within the plan there are several criteria, which must be included. An initial level of effort should be estimated. This will help determine an approximate cost estimate. In addition to the labor all other costs such as equipment and materials also needs to be itemized. An agreed upon schedule for the transition needs to be in the plan. With all of the above provided in detail the necessary funds and approvals are asked for.

After the level of effort, cost estimates, and schedules are determined the data provider, or representative, submits this information to the DART committee for their review and approval. If they deem the data collection to be a strong candidate they will recommend to the PRB that resources be given to transition the collection from the active archive to the long-term archive.

After the necessary approvals and interface requirements are given the program manager must decide which level-of-service is the best fit for the desired data. Along with the levels-of-service the Program Manager for the Long-Term Archive must prioritize the inclusion of data collections. Next the system needs to be developed within the Long-Term archive for the data. Ideally, the long-term archive database for this data collection and the active archive database should be identical. This would lessen the transition time and cost.

5.4 Long-Term Archive

1. Implement the system within the Long-Term archive
After the development of the system there is a transition process. The data collection itself needs to be either moved to a more permanent location or have the new LTA database point to the existing data. This may or may not be a long process depending upon the size of the collection and the differences between the active and long-term archives.

2. Provide ongoing support
As with all systems there is a certain amount of ongoing maintenance required to sustain a system. There will be software and hardware failures that need to be corrected as well as the normal backups and upgrades required. In addition, as explained

later in this document there will be a need to migrate the data onto newer technology media as well as to upgrade the operating systems and COTS software.

5.4.1 Data Format(s)

The data format of the active archive should consider using the Open Archival Information System (OAIS) to meet the following requirements:

- Contain the appropriate data and information required by the mission.
- The format should facilitate fast read and write times.
- Allow for maximum compression.
- Variable formats are allowed for different types of data.
- Non-proprietary formats are preferred or a migration plan should be included for transition to an open format.

5.4.2 Documentation

Thorough documentation is a must. It is easier to build this in at the beginning of a mission rather than putting it off to the end of the mission and trying to patch up what will be needed to meet FGDC requirements and properly document associated algorithms and products. Archive data sets transferred to the Long Term Archive must meet FGDC metadata standards. Other support documentation may be used to populate help files and create descriptive text in the information system used to access the mission data in the Long Term Archive.

5.4.3 System Requirements/Performance

The archive and information system should be tuned to retrieve data in a timely manner. The system should also be load tested to benchmark the level of service thresholds. Periodic routines to verify data qualities should also be built into the system.

5.4.4 Data Volumes

The Active Archive will need to communicate the data volume(s) and the nature of the required data products to the long-term archive. This information will be used to establish the archive system requirements (on line, near line, off line) and size the products for distribution on various media (CD, DLT, DVD, ftp, etc).

5.4.5 User Services

The long-term archive will provide aid to users and scientists who access the data. User Service support includes a variety of tasks that will be

negotiated with the Active Archive prior to data transfer. These efforts include the following:

- Implement data sales policies and procedures for the data set.
- Organize the Billing and Accounting system for the data set.
- Provide science support as needed/requested.
- Help develop an easy to use system from a customers' standpoint
- Provide necessary inquiry and sales reports.
- Promote the visibility of the data through conferences, brochures, and web references.

5.4.6 Science Requirements

As part of the preliminary report, a plan will be developed seeking input from the Earth science community in the development of the final long-term archive report.

5.4.7 Archive Media

The Active Archive needs to consider media risks early in its mission development to help avoid media obsolesce issues. EDC routinely plans media lifecycles to at least seven to ten years. Early discussions with prospective long term archives is recommended to identify new media conversion requirements and assure that no data are lost in data transcription activities. A joint plan to verify migration integrity is also highly recommended.

5.4.8 Ancillary Data

Active Archives may involve ancillary data that are not stored with the data collection. The Active Archive sponsors are responsible for providing access to the appropriate ancillary data to support required long term archive services for the proposed data collection.

5.4.9 Algorithms

All algorithms must be included along with the data collection documentation. Active Archives should also provide a mechanism to enable an easy approach to changing or adding new algorithms as needed.

5.4.10 Metadata

Metadata is provided at a collection level, granule level and product level. Necessary metadata to create higher-level products or to perform research deliverable capabilities are generated on the fly. The granule level metadata may also be linked to online browse capabilities where appropriate

5.4.11 Output Products

Provide various format and media options catered to the needs of the customers of the data collection. Services may include standing requests capabilities, bulk ordering, and/or volume discount capabilities negotiated with EDC. Product swaths, subsets, and/or sub-banding may also be recommended. Product options for a data collection may change as market demand for the data changes.

5.4.12 Data Management

EDC builds databases that are easily accessed, modified, and converted to new information systems and/or connected databases to service the unique requirements of each data collection. Development environments may include Oracl, Sybase, Informix, Access, or others.

5.4.13 Security

Security involves safety measures extended to protect the data from potential miss-use (ie; system hackers) as well as safe guarding the data environmentally (temperature, humidity, physical access, etc).

6 Long-Term Archive Requirements

6.1 Review and re-use of Active Archive Details

Ideally, the transition of data from the Active Archive to the Long-Term Archive would be most efficiently done if the database were identical as possible. This is only accomplished if the requirements needed in the above section were also requirements for the Long-Term Archive.

This will shorten the time to convert data, which in turn lessens the initial cost as well as the support costs. In addition, the search and order capabilities will automatically exist. The turnaround time to create products would vary depending upon the level or service provided.

6.2 Long-Term Archive Approval Process

6.2.1 NSLRSDA AAC

The Archive Advisory Committee (AAC) is an external group established to support and guide the National Satellite Land Remote Sensing Data Archive (NSLRSDA). NSLRSDA is a U.S. Government mandate (Public Law 102-555) that specifies that the U.S. Department of Interior "provide for long-term storage, maintenance, and upgrading of a basic, global, land remote sensing data set ...and shall follow reasonable archival practices to assure storage and preservation of the basic data set and timely access for parties requesting data." The Department of Interior has delegated the U.S. Geological Survey's EROS Data Center as the site for NSLRSDA. The AAC is composed of fifteen members with the following requirements:

- Two members from academia: one laboratory researcher and one classroom educator
- Four from government: one Federal, one State, one Local, and one science archivist
- Four from industry: one data management technologist, one licensed data provider, one value-added data provider, and one end user.
- Five others: one non-affiliated at large, one non-governmental organization, one international non-U.S. representative, and two at-large from any sector.

The AAC meets 1-2 times a year to debate relevant topics and issues, provide archive recommendations, and propose course of actions to the Secretary of the Department of Interior and the NSLRSDA program.

6.2.2 PRB

The Program Review Board (PRB) is the internal guiding council for EROS Data Center programs. The PRB deals with programmatic issues from a Center-wide perspective, such as program initiation, program reviews, budget allocations, program guidance, priority coordination, and resolves procedural and conflict issues between programs.

6.2.3 DART

The Digital Archive Review Team (DART) is an internal EROS Data Center (EDC) group comprised of representatives from the four major work units at EDC (Satellite Systems, Science and Applications, Computer Services, Data Services) and is chaired by the EDC Archivists. The DART is entrusted with the responsibility to "receive, evaluate, prioritize, and recommend to the appropriate EDC Manager or to the Program Review Board (PRB) actions regarding new mission data sets proposed for EDC and/or the transfer of data sets archived elsewhere." The DART process involves the evaluation of eight major aspects of a data collection. They include:

- Science Relevance
- Spatial Coverage
- Temporal Coverage
- Distribution Potential
- Political Relevance
- Supporting Metadata
- Risk of Loss
- Availability

Appendix D is the DART tool that is used to collect this information.

6.2.4 Data transcription/media migration

Transcription and media migration strategies are an ongoing activity conducted at the EROS Data Center. Active archives should touch base with EDC on a periodic basis prior to offering data collections to better understand what is being done in the long-term archive with the potential of easing the data transition process by staying compatible with the long-term archive or collaborating with EDC to develop a transfer strategy. EDC's long range plans for the archive is to refresh the archive media every 7-10 years or, as technology requires.

7 Study Recommendations

Based upon our experiences at EDC there are several recommendations, which will be documented in detail in the final report. In order to transition data from a temporary (short-term) archive to a long-term archive there are several issues and questions that need to be addressed. At the time of mission preparation, the LTA requirements should be addressed to provide a smooth transition from the active archive. Normally this is an ideal situation that is not always possible.

There are several details of the data collection that need to be clarified. The following bullets are comments pertaining to information that should be addressed when anticipating the data to be put into the LTA:

Some of the minimum requirements regardless of the level of service include:

- Provide the name and description of the data collection
- Ensure that the Science community is involved in the decision of what to retain of the data collection in the LTA and what other supporting information and services are required.
- Provide any current commitments or contracts that must be continued after the transfer to the LTA.

- Document distribution, security, and access restrictions that may exist for the data collection or with data in the collection.
- Provide the data in a documented and easily ingested format (i.e. FGDC and Open Archival Information System- OAIS)
  - http://www.fgdc.gov
  - http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html
- Provide a copy of all supporting documentation appropriate for the data transfer.
- Specific data details may include, but are not limited to:
  - General description of the mission.
  - Identify the sensors used and sensor specific details.
  - Identify the processing flow as these data were acquired, processed and added into the active archive.
  - List the data characteristics of the raw data and those of higher-level products.
  - Describe the various processing levels along with the techniques used to generate data products.
  - Provide descriptions of supporting ancillary as needed.
  - Provide a description of the active archive data access system (data searching, browse, and order functionality).
- Provide technical points of contact (i.e. science, data management, project management, operations, user services, etc).
- Provide the status of the data collection as to whether it is static, continues to grow, or is being reprocessed.
- Provide the technical information pertaining to the data:
  - Media that the data currently resides on.
  - File size (smallest, largest, and average file size).
  - Media format.
  - Initial data volume, monthly, and yearly as appropriate.
- Provide a description of the file naming convention.
- Provide a description of the metadata:
  - Field names, lengths, types descriptions, relationships to other fields, etc.
  - Geographic points (center points, corners, bounding coordinates).
  - Provide requirements for metadata additions or revisions.
  - Determine the FGDC compliancy of the metadata.
  - Describe any special needs for ingesting and maintaining metadata.
- Provide browse imagery and requirements associated with the data if it exists.
- Supply any special requirements needed prior to ingest and storing data in the LTA (i.e. reformatting, de-compression, hardware issues, etc).

The following list, although may not be entirely complete, depicts the types of requirements needed in addition to the above list. Each level of service, by definition, has different requirements needed.

Level 1 Deep Archive Services

- The LTA archive is to be maintained as an offsite backup with limited data access. Customer service is limited to manual efforts and a referral process back to the original data provider.

Level 2 Collection/Directory Level Archive Services

- Retain appropriate **ephemeris**, **algorithms**, and **ancillary** data used in the creation of the data collection.
- Provide **collection** level metadata that could be hosted on a web server.
- Customer service is available to assist users in downloading the data.

Level 3 Data Set/Inventory Level Archive Services

- Retain all **ephemeris** and **ancillary** data in addition to the **algorithms** used to create the data collection for later reference.
- Supply **metadata** for each data product.
- Customer service support includes a file by file accountability if what was done for each granule in the collection.

Level 4 Data System Level Archive Services

- Retain all **ephemeris** and **ancillary** data in addition to the **algorithms** used for each data granule for later reference.
- Customer service includes recommendations for enhanced products and full service support.
- Support and maintain required product generation software and hardware.
- Provide all **documentation** needed to generate higher-level products and metadata for each data granule.
    - Supply the technical information desired for data distribution (i.e. media, format, etc).
    - Provide a description of the customer products and user specified parameter options.
    - Provide technical training to LTA customer service personnel in order to properly address customer questions.
- Recommend a system to the LTA for users to easily access and order products from the data inventory.
- Ensure the Science community is involved in making the decision of retaining the data collection in the LTA and which levels of data is to be retained.

We will develop a plan to build the final report and share this document with the Science Community. A draft final report will be generated within two months. At that point we will solicit a representative from the Science Advisory Panel, the NSLRSDA Archive Advisory Committee, and additional slate of scientists from USGS and NASA. Contacts may be done through phone contacts or through a workshop. Advice from this process will be incorporated into the final report.

APPENDIX A

MEMORANDUM OF UNDERSTANDING

BETWEEN

THE U.S. GEOLOGICAL SURVEY

And the

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
FOR

PARTICIPATION IN

THE EARTH OBSERVING SYSTEM DATA AND INFORMATION SYSTEM

Purpose

This Memorandum of Understanding (MOU) addresses the relationship between the National Aeronautics and Space Administration (NASA) and the U.S. Geological Survey (USGS) required to plan for, implement and operate the Earth Observing System (EOS) Data and Information System (EOSDIS) Land Processes Distributed Active Archive Center (DAAC). DAAC operations include: product generation system functions, information management system functions, and archive and distribution system functions. This agreement elaborates upon the NASA/USGS MOU for Experimental Land Remotely Sensed Data Processing, Distribution, Archiving, and Related Science Support dated March 16, 1988, and supersedes the NASA/USGS MOU for EOSDIS dated 12/03/92. Based on this agreement, USGS will proceed with all necessary actions, including acquisition of required facilities, equipment and staff, to support EOS and related land data management responsibilities.

NASA's authority to enter into this interagency MOU is contained in the National Aeronautics and Space Act of 1958, as amended, 42 U.S.C. 2473(i)(5) and (c)(6). USGS's authority to enter into this interagency agreement is the Economy Act (31 U.S.C. 1535).

NASA Responsibilities

NASA will define requirements and implementation guidelines for systems required to manage EOS and related land data and to produce and distribute data products. These systems will be provided through the EOSDIS Core System (ECS) contract. NASA will deliver data from EOS and other appropriate sources from ground reception other locations to the Land Processes DAAC.

NASA will fund DAAC short-term archive functions. NASA will fund product generation system functions, information management system functions, and any needed ancillary systems and equipment for the life of the project. NASA will fund archive and distribution functions, including operations and maintenance costs for EOS and related data to acquisition date plus 3 years. NASA funding will be pursuant to written orders signed by a NASA Contracting Officer (see FAR subpart 17.5 and NASA FAR subparts 18.17.5 and 18.17.70). This MOU is not to be considered a funding document for the purposes of recording NASA obligations.

USGS Responsibilities

USGS will manage and operate the EOSDIS Land Processes DAAC, which will archive, process and distribute EOS and other appropriate land remotely sensed data. USGS will fund long-term archive functions. USGS will fund archive and distributions functions, including operations and maintenance costs for EOS and related data more than 3 years old. USGS will fund construction at the EROS Data Center to provide the basic facilities required to support EOSDIS Land Processes DAAC activities, beginning in early 1994 with occupancy planned for late 1995.

USGS will participate in the development of EOSDIS, with responsibility for liaison with the ECS contractor, participation in design reviews of ECS deliverables, and operation of ECS-deliverable systems. USGS will prototype early EOSDIS (Version 0) systems required to process and distribute existing land remotely sensed data sets as precursor data sources for EOS research. USGS will develop with NASA and EOS Archive Data Management Plan to ensure transition of EOS and other appropriate land remotely sensed data from the short-term archive to the long-term archive.

Implementation Approach

USGS will construct the basic facilities required, with facility maintenance expenses included in annual DAAC operation and maintenance cost over the life of the EOS Project, in accordance with NASA and USGS funding responsibilities defined earlier. Funding will be modified annually to account for distribution of funding responsibilities between short-term and long-term archives.

Annual plans will be submitted by USGS to NASA, based upon NASA guidance, with negotiation as required to develop an annual Statement of Work for the Land Processes DAAC.

Period of Agreement

This agreement is applicable for the life of the EOS Project, or until superseded by subsequent mutual agreement and is subject to availability of appropriations to support programmatic responsibilities.

Signatures of Approval

U.S. Geological Survey

Signature on file 9/27/93
Robert M. Hirsch Date
Acting Director
U.S. Geological Survey

National Aeronautics and Space Administration

Signature on file 9/30/93
Shelby G. Tilford Date
Acting Associate Administrator
For Mission to Planet Earth
National Aeronautics and Space Administration

APPENDIX B

NSLRSDA

Managing the National Satellite Land Remote Sensing Data Archive (NSLRSDA) along with the long term cartographic and earth science data archive is a large responsibility. Every year new as well as persistent challenges are addressed through sound record management practices and creative engineering efforts performed by the dedicated staff of the United States Geological Survey's (USGS) Earth Resources Observation Systems (EROS) Data Center (EDC). These challenges occur in digital and film records. They number in the hundreds of thousands of images, and directly threaten the treasures of NSLRSDA and the Long Term Archive (LTA) through the media they exist on and the format they exist in.

Preservation efforts consume most of the resources for EDC's archive service efforts with the balance directed towards data management. To address the preservation and data management thrusts, emphasis is placed upon project lifecycle and management techniques ensuring that resources, schedules, and communications are planed and utilized efficiently.

Preservation and data management efforts are coordinated through EDC's Archive Program. This program focuses on the Nation's land remote sensing and cartographic treasures as they relate to the USGS National Map Program (NMP) objectives, the specific EDC strategic goals, and overall USGS mission objectives. The Archive Program primarily supports the Cooperative Topographic Mapping and the Land Remote Sensing NMP themes.

The EDC performs the critical role of managing and preserving the day-to-day remote sensing, cartographic and Earth science archives through the use of database and collections management techniques and strong adherence to archival standards. Digital archival transcriptions systems like LACS, TMACS, WBVT, and SPOT/TM-R, as well as their replacement transcription systems, are designed, procured, deployed, and managed through the Archive Program area. Support for the Archive Advisory Committee and CEOS activities are included within this functional area.

The EDC Archivist who receives archive additions, purges, and overall guidance direction from the PMO leads the Archive Service area. The PMO acts upon data set recommendations by the Data and Archive Review Team (DART) that is chaired by the Archivist and supported by the Research and Applications, Satellite Systems Initiatives, Architecture and Technology, EOS Support, and the Production Services managers at EDC.

The Archive Program directly aligns with, and receives support from, the USGS Global Change Program, the NMP National Map initiative, and the National Satellite Data Management Initiative which contribute to meeting the following goals:

EDC goals:

- Is the preferred long-term **archive** for land remote sensing data from U.S. Government programs, commercial systems, and selective foreign systems.
- Is the principal **archive** and distribution facility for USGS geospatial data.
- The EDC strengthens its standing as a world leader in **data archiving** and **data management** through continued improvements in information technology and effective use of **advisory bodies**.
- The EDC leverages new **archive**, communication, and visualization technologies to pursue innovative research and analysis directions.

National Mapping Discipline goals:

- Ensure the production and availability of basic cartographic and geographic spatial data of the country.
- Coordinate national geospatial data policy and standards.
- Provide leadership for the **management of earth science data** and for information management.
- Acquire, process, **archive, manage**, and disseminate the **land remote sensing data** of the Earth.
- Improve the understanding and application of geospatial data and technology.

Department of the Interior goals:

- Continue to be a center of excellence for **archiving, preserving**, and processing land remote sensing and other geospatial data.
- Improve NMD's current land remote sensing and geospatial data and information archive capacity, including **management** of data from emerging technologies.
- Establish **archive partnership agreements** for management of USGS land remote sensing and geospatial data and information, including classified data.
- Expand NMD's role in the management and implementation of land remote sensing satellites, including operational activities.

- Provide Science for a Changing World and is composed of the following elements:
    - Resource managers make decisions based on accurate, reliable, and impartial scientific information.
    - The loss of life and property from natural disasters is minimized through the availability of timely scientific information.
    - Federal, State, and local governments and the private sector have access to shared **national databases of natural resources information**.
    - The public has easy access to earth science information.

The Archive Program directly supports the NMD Land Remote Sensing Program through the population and management of the National Satellite Land Remote Sensing Data Archive. Implementing LACS, preparing for SRTM and SIR-C acceptance, and working in partnership with NOAA National Climatic Data Center as reciprocal offsite archives are areas of strong alignment with this NMD Program. Providing archive services for the geospatial, cartographic data associated with Cooperative Topographic Mapping Program illustrates a strong linkage with that NMD Program, also. Preserving the documented changes in the land surface area will enable the Archive Program to serve the NMD Landscape Analysis Program, as well.

The National Map Draft for Public Comment document, dated April 26, 2001, includes the following references that imply the necessity of a strong Archive Program:

- "The National Academy of Public Administration (1998) identified twelve broad Federal functions, ranging from economic and community development to emergency management and defense to environmental protection, which require spatial data. The report summarized the Federal Government's role in spatial data functions as "one in which government is expected to help ensure public safety, manage the public lands for multiple uses, **preserve the nation's resources for future generations**, and help meet the basic needs of an expanding economy." [Page 5]
- "Federal leadership and commitment are needed to ensure that basic spatial data are available to support Federal agencies in accomplishing their missions. The USGS has the mission to lead the development and **maintenance of this common set of basic spatial data**…" [Page 7]

The USGS/EROS Data Center has traditionally been a user and a manager of satellite land remote sensing data. Nearly 25 years of information extraction from satellite remote sensing data have demonstrated the broad utility of these data. Information derived from Landsat data, for example, has permitted scientists and program managers to study more effectively problems related to our water, energy, and mineral resources, to understand

the effects of natural disasters, to protect the quality of the environment, and to contribute to the Nation's economic and physical development. Central to the application of these data is the reliable collection, maintenance, and distribution of a record of the Earth's surface; a record, moreover, that is comprehensive, historical, permanent, and impartial.

APPENDIX C

NewDISS Data Lifecycle Study Task
September 2001

The contractor shall conduct a 6-month study (with the option of a further 6-month extension) of the requirements and implementation options needed to manage Earth science data products throughout their lifecycle in the NewDISS era. NewDISS is a NASA initiative, which is currently in the formulation (study) stage. The overall goal of NewDISS is to establish a unifying framework of standards, core interfaces, and levels of services to facilitate access to data and information as provided by a distributed, heterogeneous network of data systems and service providers.

The preliminary report shall review and describe the life cycle for Earth science data products. The report shall also identify, in terms of operations concepts, several options for long-term archiving of Earth science data products including static tape archive, process on-demand archives, and any other options that the contractor may deem technically feasible, cost effective, and appropriate to the scientific community. In developing the preliminary report, the contractor shall consult existing documents on long-term archiving, such as those prepared by the USGCRP (U.S. Global Change Research Program) and the NAS (National Academy of Science). As part of the preliminary report, the contractor shall develop a plan for seeking input from the Earth science community in the development of the final long-term archive report.

The final report shall identify the critical points in the Earth science data product life cycle that may affect successful long-term archiving. Issues to be considered in this regard include responsibilities for funding, responsibilities for data management, data format, metadata content, documentation, and any others deemed necessary and appropriate by the contractor. The final report shall also address the functional and interface requirements of the long-term archive. In addition, the final report shall propose draft specifications that could be included in NASA Announcements of Opportunity or other solicitations that would help ensure data provider compliance with NewDISS lifecycle requirements.

The contractor shall also participate in the NASA-NOAA LTA Implementation Plan Development Team meetings, and may contribute to the NASA-NOAA implementation plan as necessary and as directed by NASA. As directed by NASA, the contractor shall also contribute to the NewDISS cost study on issues related to long-term archiving. The contractor shall participate in two NewDISS workshops (one on LTA and one on NewDISS costing) during the 6-month study period.

APPENDIX D

DART Decision Tool

Data Set: _____ Date: _____

Sponsor/POC: _____

Background Information:

Summary of Estimated Costs

Science	DAAC	Archive	Prod/Dist	Information	Telecommunications

SCIENCE RELEVANCE:

What is the current and future utility of this data set within the DOI / USGS /EDC mission?

What is the significance of this data to the Earth science and commercial user community?

How does this data set compliment or supplement EDC's current archival holdings?

Additional science justifications:

Overall Science Recommendation:
_____
_____

- Estimated costs: FTE: $___ Equip.: $___ Supplies: $___
  Annual O&M: $___

Signed by: _____

DAAC INVOLVEMENT:

- Is there DAAC involvement with this Data set? If yes, explain:
- If the DAAC is involved, what pieces will DAAC cover and for how long?
- Estimated costs: FTE: $___ Equip.: $___ Supplies: $___
  Annual O&M: $___

Overall DAAC Recommendation:
_____
_____

Signed by: _____

ARCHIVE ELEMENTS:

- Is this data collection to be archived at USGS/EDC forever or for a set period of time?
- Will a digital transcription or a film migration be required?
- Is this dataset dynamic or static? If dynamic, describe:
- What is the total expected data volume?
- Does this data collection have any security/access restrictions? If yes, describe:
- Who is the primary archive and the secondary archive?
- Estimated costs: FTE: $___ Equip.: $___ Supplies: $___
  Annual O&M: $___

Overall Archive Recommendation:

_____
_____

Signed by: _____

PRODUCTION / DISTRIBUTION:

- Is Hardware processing required? If so, will a system be provided for the processing?
- Is there a recommended product list?
- Is a COFUR / cost recovery analysis required?
- Is this data collection to be distributed solely from USGS/EDC? Yes / No
- Was this data previously distributed? If so, in what format?
- What is the recommended format?
- What, if any, ancillary files are to be distributed along with the data (tfw, fgdc, README, etc.)?
- Market- / customer analysis or product lifecycle information attached?
- What is the desired timeframe for implementation of the distribution capability (ASAP, 6 months, etc.)?
- Estimated costs: FTE: $___ Equip.: $___ Supplies: $___ Annual O&M: $___

Overall Production / Distribution Recommendation:

_____
_____

Signed by: _____

INFORMATION SYSTEMS:

- Describe the metadata available:
- Is there a requirement for digital preview images (browse)? If yes, describe:

- Estimated costs: FTE: $___ Equip.: $___ Supplies: $___ Annual O&M: $___

Overall Information Systems Recommendation:
_____
_____

Signed by: _____

TELECOMMUNICATIONS ELEMENTS:

- Are data to be delivered to the Archive?
  If Yes specify: Product Sizes: _____ Delivery Frequency: _____ Anticipated Customers:
  _____

- Are products to be delivered from the Archive?
  If Yes specify: Product Sizes: _____ Delivery Frequency: _____ Anticipated Customers:
  _____

- Will there be a high demand for information services? Yes /No
  If Yes, Explain: _____

- Estimated costs: FTE: $___ Equip.: $___ Supplies: $___ Annual O&M: $___

Overall Telecommunications Recommendation:
_____
_____

Signed by: _____

APPENDIX E

Long-Term Archive Infrastructure

The following is a brief summary of services provided by the EROS Data Center (EDC) that may be appropriate for the Earth Science Data Lifecycle Study.

The EDC's **Customer Service** is responsible for the information, research, ordering, accounting and billing associated with the remote sensing and other geospatial data products offered by the U.S Geological Survey. Customer Services is also responsible for the EDC DAAC User Services and the Federal Brokerage Program. EDC DAAC User Services provides customers with access to the information and data associated with NASA's Earth Observing System Program, and the Federal Brokerage Program provides commercial satellite data to authorized federal government agencies.

Customer Services provides a worldwide user community with information, research, ordering and accounting/billing for digital and photographic products, remote sensing and other geospatial data products available through EDC. These products include satellite, aerial, and digital cartographic data, and are available to customers in a variety of formats (that is, hard copy, digital tape, or the Internet. Customer Services also provides visiting scientists and walk-in customers with ordering and research assistance. A variety of research tools are available to query and select products (Earth Explorer, GLIS, EDG and other information used to represent and support data collections). Other services include EDC tour support, information system demonstrations, as well as providing a variety of literature on the products and services offered at EDC.

EDC's **Data Management** is responsible for the maintenance and integrity of the film archive, and monitoring the integrity of data inventories on EDC computer systems. Technical Area activities involve the receipt and inventory of new data, quality assessment, data base entry and update, and dissemination of data base information and reference aids (micrographics, digital catalogs, etc). The film archive consists of over 100,000 rolls of imagery containing in excess of 12 million frames. The data inventory contains nearly 1.2 million geographic accessions referencing EDC records and 3 million geographic accessions referencing foreign-held Landsat data.

Data Management maintains the archives in such a manner that the masters are protected from any damage and are rapidly retrievable for reproduction. Temperature and relative humidity are monitored to ANSI specifications for

film storage. Various topographic, navigational, and other specialized maps and charts are also maintained.

EDC's **product generation systems** support a variety of products. Product services range from archive copies to file copies to various options and format changes to custom image processing. Products are produced for delivery over the network or on hard media.

EDC's **Dissemination** area prepares and packages all data products including photographic materials, digital media, and Compact Disk (CD) products. This activity involves sleeving photographic film, cutting finished photographic products into standard and uniform sizes, correlating products to work orders, matching the image and digital tape or CD work to the shipping Document, packaging the products for shipment, and shipping products by the most cost-effective means. Automated shipping systems are maintained by Dissemination and are responsible for their accountability, security, and ensuring adequate funding is available in the USPS postage meter to operate on a daily basis.

Systems Architecture provides for the initiation and coordination of the development and maintenance of center-wide systems and information infrastructure strategies. These strategies reflect a 3-year forward-looking vision for the migration of EDC's system architectures. Architectural planning ensures the alignment of the requirements of the systems with the business process that support the enterprise mission, promotes interoperability, redundancy and security for the information systems and evolves application and maintenance of a collection of standards by which the enterprise evaluates and acquires new systems. The framework consists of hardware, software, networks, databases and procedures that support the sharing and processing of information center-wide

Technology Investigation performs trade studies, system benchmark testing, and system prototype testing for computer and information related technologies. This work involves simple component level evaluations up to complete system integration testing of hardware, software, and networking technologies to assure they meet vendor specification and more importantly EDC program requirements. Personnel design test plans, identify and report on candidate technologies for exploration, and propose technology options. This activity documents component technology life cycles and establishes timelines for strategic technology deployment.

Infrastructure Development provides planning, design, development and implementation of organizational-wide systems. Examples of the type of services supported include: Systems Engineering support for database upgrade; Integration of new messaging systems; Overall system integration that will benefit the entire enterprise. The Infrastructure Development Activity

will establish and demonstrate principles and practices for designing, integrating and evolving systems using previously built and commercially available components.

Abbreviations and Acronyms

ACC	Archive Advisory Committee
ANSI	American National Standards Institute
CEOS	Committee on Earth Observation Satellites
DAAC	Distributed Active Archive Center
DART	Digital Archive Review Team
DIF	Directory Interchange Format
DISS	Digital Image Sciences Section
DORRAN	Distributed Ordering Research Reporting & Accounting Network
ECS	EOSDIS Core System
EDC	EROS Data Center
EOS	Earth Observing System
EOSDIS	EOS Data and Information System
EROS	Earth Resources Observation Systems
FGDC	Federal Geographic Data Committee
ftp	File transfer protocol
ICD	Interface Control Document
LACS	Land Archive Conversion System
LTA	Long-Term Archive
MSS	Multi-Spectral Scanner
NAS	National Academy of Science
NASA	National Aeronautics and Space Administration
NewDISS	New Data and Information Systems and Services
NMD	National Mapping Discipline or National Mapping Division
NMP	National Map Program
NOAA	National Oceanic and Atmospheric Administration
NSLRSDA	National Satellite Land Remote Sensing Data Archive
PMO	Program Management Office
PRB	Program Review Board
SPOT	Systeme pour l'Observation de la Terre (France)
TM	Thematic Mapper
TMACS	TM and MSS Archive Conversion System
USGS	United States Geological Survey
USGCRP	U.S. Global Change Research Program
WBVT	Wide Band Video Tape