Land Processes Distributed Active Archive Center Product Lifecycle Plan

Open-File Report 2014–1139

U.S. Department of the Interior
U.S. Geological Survey

Land Processes Distributed Active Archive Center Product Lifecycle Plan

By John C. Daucsavage and Stacie D. Bennett

Open-File Report 2014–1139

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2014

Suggested citation:
Daucsavage, J.C., and Bennett, S.D., 2014, Land processes distributed active archive center product lifecycle plan: U.S. Geological Survey Open-File Report 2014–1139, 20 p., *http://dx.doi.org/10.3133/ofr20141139*.

ISSN 2331-1258 (online)

Contents

Figures

Tables

Abbreviations

ASTER Japan	Japanese Advanced Spaceborne Thermal Emission and Reflectance Radiometer Science Team
ASTER L1T	Advanced Spaceborne Thermal Emission and Reflectance Radiometer Level 1 Terrain and Precision Corrected
ATBD	Algorithm Theoretical Basis Document
CCSDS	[International] Consultative Committee for Space Data Systems
DAAC	Distributed Active Archive Center
DLL	Dynamic Link Library
ECHO	EOS Clearing HOuse
ECS	EOSDIS Core System
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
EROS	Earth Resources Observation and Science
ESD	Earth Science Data
ESDIS	Earth Science Data and Information System Project
ESDR	Earth Science Data Record
ESDS	Earth Science Data System Program
ESDT	Earth Science Data Type
ESIP	Earth Science Information Partners
ESRIN	European Space Research Institute
FGDC	Federal Geographic Data Committee
GCMD	Global Change Master Directory
GDT	Global Digital Topography
GFCC	Global Forest Cover Change
ICD	Interface Control Documentation
ISO	International Organization for Standardization
LP	Land Processes
LTA	Long-Term Archive Project
MEaSUREs	Making Earth System Data Records for Use in Research Environments Program
MODIS	Moderate Resolution Imaging Spectroradiometer
MOU	Memoranda of Understanding
NAALSED	North American ASTER Land Surface Emissivity Database
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration

NSLRSDA	National Satellite Land Remote Sensing Data Archive
OA	Operations Agreement
ODL	Object Description Language
OGC	Open Geospatial Consortium
OPS	Operations
PCS	Preservation Content Specification
PI	Principal Investigators
PLP	Product Lifecycle Plan
RMA	Reliability, Maintainability, Availability
RMOAIS	Reference Model for Open Archival Information System
SIPS	Science Investigator-led Processing Systems
SLA	Service Level Agreements
SOO	ESDIS Science Operations Office
SRTM	Shuttle Radar Topography Mission
TOMS	Total Ozone Mapping Spectrometer
USGS	U.S. Geological Survey
VIP	Vegetation Phenology and Vegetation Index Products
WBS	Work Breakdown Structure
WELD	Web-enabled Landsat data

Land Processes Distributed Active Archive Center Product Lifecycle Plan

By John C. Daucsavage[1] and Stacie D. Bennett[1]

Abstract

The U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center and the National Aeronautics and Space Administration (NASA) Earth Science Data System Program worked together to establish, develop, and operate the Land Processes (LP) Distributed Active Archive Center (DAAC) to provide stewardship for NASA's land processes science data. These data are critical science assets that serve the land processes science community with potential value beyond any immediate research use, and therefore need to be accounted for and properly managed throughout their lifecycle. A fundamental LP DAAC objective is to enable permanent preservation of these data and information products. The LP DAAC accomplishes this by bridging data producers and permanent archival resources while providing intermediate archive services for data and information products.

Introduction

At the front end of the product lifecycle, the Inception Phase, data and information are initially produced by Principal Investigators (Producers) for validation by a select subset of the land processes community. During this phase, a proper foundation is laid for best practices in anticipation of long-term data preservation. As products mature they transition to the Land Processes (LP) Distributed Active Archive Center (DAAC) in the Active Archive Phase of the lifecycle. The LP DAAC promotes the accessibility of data and information products to the at-large land processes community and provides highly reliable interim preservation with higher levels of user service. Ultimately, the LP DAAC facilitates the migration of the data and information products to the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center in the Long-Term Archive Phase, which provides permanent preservation in the final phase of the lifecycle.

This document, which serves as a template for project planning, is written from the point of view of the LP DAAC, which is chartered to provide leadership in the capture of valuable land process science data and ensure these data are preserved. The LP DAAC serves as an advocate for the products as they move through the lifecycle from Inception to Active Archive and from Active Archive to Long-Term Archive, ultimately resulting in permanent preservation. Product capture is the first step in providing community-wide access to data and information. In its advocacy role, the LP DAAC ensures that product attributions align with land processes standards so that data and information are easily discoverable and usable by the community. The LP DAAC also advocates that products adhere to interoperability standards so that products new to the land processes community can be integrated with existing products.

For this plan to be effective, the LP DAAC must be involved with product producers during the Inception Phase to help the Producer select content styles and data formats that efficiently integrate with existing LP science community standards. A primary goal for all science and system stakeholders is to minimize the need for product transformations to support community-wide interoperability and to cost-effectively migrate data and information products to the archive phases. The LP DAAC also supports the Producer by providing a detailed understanding of USGS preservation policies and processes; therefore, the LP DAAC assists the Producer in documenting content styles and data formats in the Producer's proposed and final Data Management Plan. This LP DAAC Product Lifecycle Plan summarizes each of the three Product Lifecycle Phases and provides a framework used by LP DAAC Data Scientists to plan data activities leading to lifecycle migrations.

Background

Ramapriyan and Moses (2013) provide the justification for the mission critical status afforded to the requirement of the National Aeronautics and Space Administration (NASA) for the preservation of science observations. One of NASA's strategic objectives is to study Earth from space to advance scientific understanding and meet societal needs. The Earth

[1]Stinger Ghaffarian Technologies, Inc., contractor to the U.S. Geological Survey Earth Resources Observation and Science (EROS) Center.

Science Data System (ESDS) Program resides within NASA's Earth Science Division and supports the above strategic objective by providing end-to-end capabilities to deliver data and information products to users. The data resulting from NASA's missions are a valuable resource that needs to be preserved for the benefit of future generations.[2] These observations are the primary record of the Earth's environment and therefore are the key to understanding how conditions in the future will compare to current conditions.

In the near-term, as long as the data from the missions are being used actively for scientific research, it is important to provide easy access to the data and services commensurate with current information technology. For the longer term, when the focus of the research community shifts toward new missions and observations, it is essential to preserve the previous mission data and associated information, which will enable a new user to understand how the data were used for deriving information, knowledge, and policy recommendations; to "repeat the experiment" to ascertain the validity and possible limitations of conclusions reached in the past; and to provide confidence in long-term trends that depend on data from multiple missions.

Unlike some Federal agencies, NASA is not legislatively mandated to preserve data permanently, however it is essential for NASA to preserve all the data and associated content beyond the lives of NASA's missions to meet the near-term objective of providing access to data and services for active scientific research. Also, NASA has to ensure that the data and associated content are preserved for transition to permanent archival agencies.

Overview of Product Lifecycle Plan

This Product Lifecycle Plan (PLP) provides a vision of roles, responsibilities, and activities led by LP DAAC Data Scientists in each phase of the lifecycle to transition data and information products from their inception to their final preservation state.

For the purposes of this report, product preservation for Earth science data and information is executed within a three-phased lifecycle: (1) Inception, (2) Active Archive, and (3) Long-Term Archive. Phase 1 provides short-term preservation, Phase 2 provides interim preservation, and Phase 3 provides long-term preservation, with activities that lead to the selection of one of two permanent preservation options. Phase 2 includes the delivery of services that provide an operational bridge between the other two phases. In this lifecycle model, NASA sponsors both the Inception and Active Archive phases while cooperating with USGS to provide the Long-Term Archive phase.

The NASA Earth Science Data and Information System (ESDIS) Project, a part of NASA's ESDS Program, sponsors several Distributed Active Archive Centers (DAACs), each of which is aligned with specific Earth science disciplines providing interim preservation. This report is concerned with data and information related to the Land Processes (LP) Earth science discipline assigned to the LP DAAC operated by the USGS. The USGS has a mandate from the Department of Interior (Land Remote Sensing Policy Act of 1992; Public Law 102-555, 15 U.S. Congress 5601) to establish a public domain archive of satellite data of the Earth's land surface. EROS and the LP DAAC synergistically share the objective to provide information and services to the land processes Earth science user community including data and information product preservation.

This LP DAAC PLP is written from the point of view of DAAC Data Scientists who have the responsibility to oversee the planning and operation for the migration of land processes data and information products from the Producer to EROS systems.

Scope

The scope requires that the LP DAAC facilitate operational relations between data and information product Producers, NASA ESDS, and EROS. At the highest level, the LP DAAC must gain authorization from NASA to establish interfaces with new Producers to permit the consumption of resources provided to the DAAC by NASA. Likewise, the LP DAAC must gain concurrence from EROS to establish new interfaces to consume resources.

To the extent possible and reasonable, the LP DAAC assumes a leadership role in planning data and information product migrations in a manner providing minimum cost and disruption to all operations servicing the lifecycle. To achieve this goal, the LP DAAC engages with the Producers as early as possible to both identify that which should migrate and advocate for the early adoption of standards compatible with the archival systems. Because the LP DAAC and other EROS systems are designed with compatible import/export capabilities, the bulk of the technical work for this activity is to establish interfaces with the Producer. In addition, substantial management level work is required to obtain the NASA andUSGS agreements necessary to permit new products to migrate to archival systems. The work required to obtain these agreements anticipates the archival appraisal process and lays the groundwork necessary to assure that scientific need and benefit is well documented and accepted by the LP community.

[2]Data processed from the Total Ozone Mapping Spectrometer (TOMS) instrument in the late 1970s did not indicate the dramatic loss of ozone over the Antarctic because the quality control algorithms used in processing the raw ozone data treated very low values of ozone as bad readings. When in situ data became available in the early 1980s indicating a hole, the legacy algorithms were modified and the raw data "from the archive" were reprocessed, providing clear evidence of ozone depletion. This well-known use case highlights the importance of being able to go back to the archive for data and information products. (Centre for Atmospheric Science, 1998; National Aeronautics and Space Administration and National Oceanic and Atmospheric Administration, 1999; Welch, 2014)

Key Standards Organizations

NASA and USGS both contribute to the development and rely on the application of data standards of consequence for data formats, data transport, metadata standards, and content standards for archive and preservation. Organizations that define standards used throughout the product lifecycle include the U.S. Federal Geographic Data Committee (FGDC), International Organization for Standardization (ISO), and the Open Geospatial Consortium (OGC). LP DAAC Data Scientists serve product lifecycle stakeholders by advocating adoption and use of agreed-upon standards and protocols for maximum benefit to the LP science community. Standards promote interoperability with existing products, ultimately fostering cost-effective approaches for research into new products as well as reuse of post-production systems for discovery, access, and preservation.

Preservation Roles and Responsibilities

Specific to this document, the Inception, Active Archive, and Long-Term Archive product lifecycle phases align respectively to the Producer, LP DAAC, and EROS organizations that operate and sustain each phase. The roles and responsibilities of each organization with respect to the "preservation characteristic" of each phase are summarized in table 1.

Product Lifecycle Plan

A three-phased data and information product lifecycle from the point of view of roles executed by each of the organizations operating the phase is summarized in figure 1.

The Inception Phase is the domain of the science Producer, who originally provides the products as well as short-term preservation; the Active Archive Phase is the domain of the LP DAAC Operations, which provides intermediate preservation; and the permanent Long-Term Archive Phase is the domain of EROS (fig. 1). USGS sustains a number of projects at EROS having permanent long-term archival responsibilities; the LP DAAC is slated to work with the EROS Long-Term Archive (LTA) Project for long-term[3] archival services. Administratively, products held by the various EROS projects are destined for offsite storage in the National Archives and Records Administration once the science utility expires; however, products may be additionally certified for inclusion in the National Satellite Land Remote Sensing Data Archive (NSLRSDA) and stored permanently at EROS.

The overlapping of phases indicates a time period of dual "possession" of products while migration is occurring from one phase to the next. Following migration from one phase to another, the originating organization is released from formal preservation responsibilities but may retain the products to provide an additional level of disaster recovery. During the overlap, the originating organization may also continue to provide discovery and accessibility services. The phases with respect to activities performed are characterized in figure 1, whereas appendix 1 provides a more detailed point of view that contrasts functionality and level of service across the phases. Further discussion of the phase characteristics is beyond the scope of this document; this document focuses on only one characteristic—preservation.

This plan emphasizes the role of the LP DAAC Data Scientists with respect to product preservation relative to each of the lifecycle phases. Thus, to the extent possible, the LP DAAC strives to become embedded in the activities providing product preservation for both the Inception and Long-Term Archive phases; therefore, the LP DAAC provides continuity of policy and tests operational concepts while serving as an advocate for interoperability across each of the phases.

Inception Phase

This section provides the framework necessary for the LP DAAC Data Scientists to plan Inception activities relative to enabling the most cost-effective method to ensure preservation of new Producer data and information products. This phase is initiated when a NASA-funded program assigns a project or mission to the LP DAAC, or when a community or legacy product has been recommended for inclusion in the LP DAAC archive.

Data Producer Characterization

Unlike LP DAAC and EROS, which are well-defined organizations, the term "Producer" has been left generic in this document. As indicated in table 2, Producers can be classified into two groups relative to NASA's programmatic mandate to preserve data and information products.

As shown in table 2, the generic term "Producer" allows for considerable variance as to what data and information products are available and require preservation through LP DAAC and EROS resources. For example, new Producer products may utilize existing sensors that have artifacts preserved in other efforts, reducing instrument-related preservation to simple acknowledgement of archival products servicing the original Producer. In other cases, legacy products may not adhere to Earth Observing Systems Data and Information

[3]How long is long-term? Long-term is a period of time long enough for there to be concern about the effects of changing technologies, including the support of new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the immediate future and beyond the point when the data producers are available for consultation. EROS provides two permanent preservation paths for products: those EROS holdings ultimately slated for the National Archives and Records Administration and those certified for inclusion in the National Satellite Land Remote Sensing Data Archive (NSLRSDA).

Table 1. Organizational preservation roles by lifecycle phase.

[LP, Land Processes; DAAC, Distributed Active Archive Center; USGS, U.S. Geological Survey; EROS, Earth Resources and Observation Center]

Phase	Producer	LP DAAC	USGS EROS
Inception	Identify and produce data and information products that are needed and benefit the land processes science community Preserve products while they are being exploited for initial research and validation Select formats for products that are as compatible as possible with those existing within the land processes science community Work with LP DAAC data scientists to identify candidate products to be preserved in downstream archive systems	Identify product need and benefit Support producer in developing draft and final data management plans Using a repeatable checklist process, identify products from new Producers that serve the at-large land process community Advocate for product content that integrates with observations from other land process producers Advocate for product formats that are compatible with other land process community formats and LP DAAC systems	Support preliminary need and benefit appraisal. Support compatibility analysis of format and content for new products with respect to the land process community atlarge. Support requirement analysis to identify algorithms, data, and information products necessary for regeneration.
Active Archive	Notify and coordinate with LP DAAC when new versions of products will be released and advise of any requirements to remove all or parts of old versions Notify LP DAAC of specific products that have been used in publications that should be retained regardless of versioning Advise research community when responsibility for dissemination of products has been transitioned to archival systems	Plan migration activities for products from the Producer to LP DAAC systems Acquire and archive each product version as it is released by the producer providing preservation until products migrate to USGS EROS Provide data management for redundant versions Preserve key data in the version that supports any given publication Preserve either the means of regenerating the higher level products or the products themselves to ensure reproducibility and verifiability of scientific results	Support high level timeline planning to receive new products from the LP DAAC. Support preliminary capacity requirement analysis. Support pathfinder simulations to receive new products from the LP DAAC.
Long-Term Archive	Work with LP DAAC to identify essential subsets of data and information products to be preserved long term.	Oversee migration of new producer products to the USGS EROS. Support USGS EROS in producing an official Science Assessment to justify that new Producer products be permanently archived. Advise the land processes community when responsibility for dissemination of new Producer products has been transitioned to USGS EROS.	Provide official Science Assessment documentation. Support detail timeline planning to receive new products from the LP DAAC. Provide capacity to hold new products. Provide permanent archive of new products.

System (EOSDIS) standards for content and formatting, so preservation activities may necessitate substantial effort to create tools to transform content and formatting to ensure compatibility with EOSDIS and EROS systems.

The detailed Work Breakdown Structure (WBS) for any plan to migrate and preserve data and information products from a Producer to the LP DAAC will vary considerably between Producers. In fact, it is possible that the preservation WBS could vary considerably between different products for a given Producer. Flexibility is afforded to the LP DAAC to create a preservation WBS that fits each use case.

For Producers typified within the Making Earth System Data Records for Use in Research Environments (MEaSUREs) Program, the Inception Phase proceeds in two stages.

Inception Phase

Producer team roles

- Justify need for specific science
- Acquire sponsorship for science research
- Visualize science results
- Define sensors
- Provide essential resource infrastructure
- Acquire low level science data
- Develop science algorithms
- Produce science products (perform the experiment)
- Validate science products
- Provide quick turnaround to restricted user base (latency)
- Provide short-term preservation
- Establish archival responsibilities with ancillary and low level data providers
- Establish relationship with Distributed Active Archive Center science team
- Support migration planning
- Migrate to Active Archive

Data and information products (subsections from National Aeronautics and Space Administration earth science data preservation content specification, Ramapriyan and Moses, 2013)

- Pre-operations calibration (3.1)
- Science data products (3.2)
- Science data documentation (3.3)
- Mission data calibration (3.4)
- Production software (3.5)
- Algorithm inputs (3.6)
- Validation artifacts (3.7)
- Tools (3.8)
- Project documentation (3.9)

*Migration of data and information to **At-Large User Community***

Active Archive Phase

Active Archive roles

- Capture important science
- Establish embedded relationship with Producer and Long-Term Archive
- Provide mid-term preservation
- Provide for adherence user community policy
- Advocate for maximum interoperability
- Provide infrastructure sized for user community needs
- Lead migration planning
- Acquire information and data from producer
- Enhanced user experience
- Provide reasonable turnaround to user base (latency)
- Provide production as needed
- Repeat the experiment by inspection of checklist of needed items
- Anticipate Long-Term Archive phase
- Lead Long-Term Archive migration planning

Data and information products

- Series of algorithm packages for released products (3.3.3)
- Ancillary not archived elsewhere (3.6)
- Raw data not archived elsewhere (3.2.1)
- Most current range of data products (3.3.3)
- "Golden copies" of past data products (3.3.3)
- Preservation of cited data (3.3.1)
- Project documentation (3.9)

*Migration of data and information to **Permanent Archive Services***

Long-Term Archive Phase

Long-Term Archive roles

- Preserve important science
- Provide archive policy
- Support inception phase planning
- Support Distributed Active Archive Center science team for capacity planning
- Provide preservation infrastructure
- Support migration planning
- Acquire essential data and information
- Repeat the experiment by inspection
- Advertise (generalized user discovery)
- Media based data management

Data and information products

- Final version algorithm package (3.3.3)
- Final ancillary data not archived elsewhere (3.6)
- Raw data not archived elsewhere (3.2)
- Final version high-level products (3.2.1)
- Preservation of cited data (3.3.1)
- Select project documentation (3.9)

Figure 1. Three-phased product lifecycle model.

Table 2. Grouping producers by National Aeronautics and Space Administration preservation mandate.

[NASA, National Aeronautics and Space Administration; LP, Land Processes; DAAC, Distributed Active Archive Center; SRTM, Shuttle Radar Topography Mission; ESDS, Earth Science Data System; ESDR, Earth Science Data Record; EOSDIS, Earth Observing System Data and Information System; ASTER L1T, Advanced Spaceborne Thermal Emission and Reflectance Radiometer Level 1 Terrain and Precision Corrected; NAALSED, North American ASTER Land Surface Emissivity Database; MEaSURES, Making Earth System Data Records for Use in Research Environments; VIP, Vegetation Phenology and Vegetation Index Products; WELD, Web-enabled Landsat Data; GFCC, Global Forest Cover Change; GDT, Global Digital Topography; ASTER Japan, Japanese Advanced Spaceborne Thermal Emission and Reflectance Radiometer Science Team; MODIS, Moderate Resolution Imaging Spectroradiometer; SIPS, Science Investigator-led Processing Systems]

Producer group	Producer characteristics	Examples
Community offerings (outside NASA programmatic mandate)	Sometimes considered a data rescue activity for valuable land processes products (LP DAAC mandate for preservation may need to be negotiated). Inception phase most likely has been completed before LP DAAC involvement. LP DAAC may have had few real-time opportunities to influence Producer decisions. Often not supported with the continuous and rigorous infrastructure needed to provide permanent preservation. Typically an informal Data Management Plan. Instrument development artifacts may be part of the information not preserved elsewhere. Raw data likely not preserved elsewhere. Service Level Agreements may not exist with ancillary data providers to assure preservation on other systems.	SRTM.
ESDS program assigned	LP DAAC shares mandate for preservation with producer. Draft Data Management Plan with preservation roles defined was provided with Principal Investigators (PI) project proposal. LP DAAC Data Scientists have enough time to become "embedded" in PI team relative to Data Management Plan and preservation. Instrument development artifacts are preserved elsewhere (sometimes these may not be available anywhere, unless planned for preservation ahead of time.) Raw data are preserved elsewhere. Equivalent of "raw data" may actually be ESDRs that are preserved in other systems. Product formats and contents follow EOSDIS standards. Service level lgreements usually exist with ancillary data providers. Links together multiple ESDRs or utilizes individual ESDRs in a new way to create new parameters or new climate data records.	ASTER L1T, NAALSED, MEaSURES (VIP, WELD, GFCC, GDT), ASTER Japan, MODIS SIPS.

At the first stage, NASA accepts the Principal Investigator's (PI's) proposal[4] to develop a new product or suite of products and NASA assigns the product suite to the LP DAAC. In the second stage, the proposed work begins and the PI becomes the Producer of a new product. Optimally, LP DAAC Data Scientists are informally embedded with the proposal team early in the first stage. Early involvement allows the PI to better benefit from LP DAAC knowledge of the at-large user community, and allows the LP DAAC to support the PI in writing the Data Management Plan generally required after proposal acceptance. By the second (production) stage, embedded LP DAAC Data Scientists are signed contributors to the Final Data Management Plan and lead the Product Assessment process required by NASA to allow new products to migrate to the Active Archive Phase.

In some cases, the product already exists and is assigned to the LP DAAC after production either is well underway or has been completed. This is often true for Community Offerings and legacy products that have been recommended by the science community for active management or long-term preservation. The development of management plans for such products is treated on a case-by-case basis, depending on existing documentation and descriptions of data such as metadata quality and availability, Algorithm Theoretical Basis Documents (ATBDs), algorithm code, and ancillary production data.

[4]NASA may require that any data obtained through an award be deposited in an appropriate public data archive as soon as possible after calibration and reduction. If so, NASA will negotiate with the organization for appropriate transfer of the data and, as necessary, may provide funds to convert the data into an easily used format using standard units. (NASA Research Announcement or Cooperative Agreement Notice Proposers' Guidebook)

Critical Data and Information

Ramapriyan and Moses (2013) provide considerable detail on the types of data and product information that need to be considered for long-term preservation. The content items are divided into eight categories: Preflight/Pre-Operations Calibration, Science Data Products, Science Data Product Documentation, Mission Data Calibration, Science Data Product Software, Science Data Product Algorithm Input, and Science Data Product Validation and Science Data Software Tools. Ramapriyan and Moses (2013) breaks down each of the categories, providing item descriptions as well as rationale relative to why future users of the products will need each item. LP DAAC Data Scientists have synthesized this extensive list into an LP DAAC Collection Inception Checklist (appendix 2). The checklist can be used to both identify needed items and help facilitate application of standards.

Both NASA and the National Oceanic and Atmospheric Administration (NOAA) (1999) acknowledge that it is necessary to plan for the archival of critical documentation to accompany data and algorithms as early as possible in the Inception phase, which is a consistent theme present in the literature addressing preservation: "Data archiving makes the generation of new geophysical results possible; however, preserving complete documentation along with data is of absolute and fundamental importance for proper study. Every example of successful re-analysis of archived data has depended upon quality documentation…[Relative to life-cycle data management] It must be an ongoing process—not stopping when an initial implementation is made" (National Aeronautics and Space Administration and National Oceanic and Atmospheric Administration, 1999, page 2). Both NASA and NOAA (1999) cite several examples of cases where the existence of a data and information product archive has allowed reprocessing to produce new products (such as the ozone hole example) (National Aeronautics and Space Administration and National Oceanic and Atmospheric Administration, 1999, page 12). To the extent practical, LP DAAC Data Scientists will become embedded in Producer teams at Inception to support the Producer in designing data and information products that can successfully move through the lifecycle to meet the preservation requirement.

Applicable Standards

During the proposal stage of the Inception Phase, LP DAAC Data Scientists help the PI to define the standards framework needed to describe proposed data and information products. In the production stage, LP DAAC Data Scientists help Producers "design to" the proposed standards framework in such a way as to maximize interoperability with the LP science community as well as downstream archive tools and services.

One of the critical design elements for production systems is the metadata structure used to describe data and information in catalogues such as data dictionaries and data inventories critical to access and discovery operations. Metadata records include core library catalog elements such as the title, abstract, and publication data; geographic elements such as geographic extent and projection information; and database elements such as attribute label definitions and attribute domain values (Federal Geographic Data Committee, 2013). LP DAAC Data Scientists can make a substantial contribution to the production team by providing an understanding of necessary metadata and how the metadata should be structured to best serve post-production stakeholders.

NASA and USGS rely on the application of data standards endorsed by the FGDC to consistently provide services to the LP science community. Because the FGDC has endorsed ISO 19115-1[5] and associated standards, LP DAAC Data Scientists advocate for ISO metadata using the ISO XML format wherever practicle. One of the benefits of ISO XML is the support available to the science community through a wide range of tools such as ISO metadata editors used for metadata creation and update. However, presidence exsists to grandfather in products defined using the Global Change Master Directory (GCMD) standards which preceded ISO 19115-1.[5] For example, grandfathering includes the use of Object Description Language (ODL) in the Hierarchical Data Format headers for data products reliant on legacy product chains and for products that will be used by legacy production systems.

Working closely with the Producer, LP DAAC Data Scientists focus on the organization of metadata used to describe the data and information products to begin the process to formulate object containers known as Earth Science Data Types (ESDTs). As part of the migration process, ESDTs are compiled into the EOSDIS Core System (ECS) and delivered to the LP DAAC to take maximum advantage of Active Archive services. Ultimately, ESDTs enable standardized interfaces between production systems and the Active Archive. Because of the adherence to standards common to both NASA ESDS and EROS, the interface between the Active and Long-Term Archives becomes cost-effective to implement because of the Producer's use of standards at the beginning of the product lifecycle.

LP DAAC Data Scientists also contribute to the production team by confirming the Producers' expectations as to what standards for product content and format best serve the LP science community. These standards are well established and all stakeholders are recommended to adopt them to take advantage of the many tools and services in existence used to manipulate products in research efforts.

---

[5]The forthcoming, ISO 19115-1, despite its numbering, was developed after ISO 19115-2 and is an updated version of the base standard (ISO 19115). In addition to changes to the structure of some base elements and the content of some domains, ISO 19115-1 expands upon the former standard by providing more fields to describe geospatial data services (ISO 19119), multi-dimensional gridded datasets, modeling results, etc., and enabling entity/attribute descriptions developed using ISO 19110: Feature Catalog to be associated with or integrated into the metadata record. (Federal Geographic Data Committee, 2013)

Planning for Transition from Production Systems to LP DAAC

For LP DAAC work breakdown purposes, the Inception Phase is subdivided into two stages: the Planning Stage and the Production Stage. A Planning Stage WBS development guide, provided in table 3, is used by the DAAC Data Scientist and Producer to lay the initial groundwork during the Planning Stage to ensure critical items on the LP DAAC Collection Inception Checklist are planned for development in the Production Stage because they are essential for downstream support. The Planning Stage is initiated following NASA's assignment of the product suite to the LP DAAC.

A Production Stage WBS development guide, provided in table 4, is used by the DAAC Data Scientist and Producer to liaison to the science community, support production, and plan transition of products to the LP DAAC. The DAAC Data Scientist derives many requirements for migration and preservation activities from workflows to produce artifacts needed by NASA approval authorities. The NASA Data Template is a key work instruction with necessary workflows provided by ESDIS Science Operations Office (SOO) for Product Developers, DAACs and ESDIS-SOO, and Science Community Participants.

Active Archive Phase

This section provides the framework necessary for the LP DAAC Data Scientists to plan Active Archive activities relative to enabling the most cost-effective method to ensure preservation of new Producer data and information products.

Active Archive Characterization

In the Active Archive Phase, the LP DAAC Data Scientists bridge the Inception Phase to the Long-Term Archive Phase. At the initiation of the phase, the LP DAAC receives authorization to migrate approved products and services from the Producer into the Active Archive. Development of the request for authorization materials (NASA Data Template) begins early in the Inception Phase but is completed after production begins to provide products for assessment by the LP DAAC Data Scientist, the LP DAAC User Working Group, and ESDIS.

Once approval to migrate is received, the LP DAAC sets up and operates core system interfaces needed to acquire and ingest data and information products from the Producer as well as to enable tools and services for discovery and access by the at-large LP science community. To enable core system interfaces, the LP DAAC Data Scientist creates new ESDTs for each collection to migrate from the Producer. Often, new ESDTs can be modeled after existing ESDTs (in some cases it may be possible to completely reuse an ESDT already in place).

ESDTs have a development lifecycle of their own. Currently, the LP DAAC Data Scientist "codes" a Draft ESDT in an ASCII descriptor file in ODL format to describe metadata and services. ODL is used to convey relationships between product attributes and their characteristics. The ODL file provides "keyword=value" statements that can be correlated to ISO XML using available translation tools. ESDTs are passed to the ECS maintenance team for review and compilation. The compiled ESDT descriptor files are added to the core systems Dynamic Link Library (DLL), enabling services (such as ingest, search, and archive) for the new product. As part of the process, the LP DAAC Data Scientist works with the Producer to select one of many available "ingest methods" used to pull or push products into the Active Archive. The LP DAAC Data Scientist acquires some representative samples of the new product and examines the implementation in a test mode of the core system. Corrections are made and the process is repeated until all testing is successful.

Once testing is complete, the ESDT is released to the LP DAAC as a Test Executableor in the next formal delivery of ECS. The LP DAAC Data Scientist creates a timeline to enable ingest interfaces and begins the process to migrate data from the production system to the Active Archive. At a mutually agreed time, additional functions are enabled in the core system so the new products are eligible for extended services such as publication in community-wide user client systems and ultimately export to other compatible systems.

Downstream in this phase, the LP DAAC Data Scientist receives authorization to migrate approved products to a EROS project for long-term archive. Migration includes any additional work to seek NSLRSDA acceptance for the new products to be permanently archived at USGS. The actual work to lead LP DAAC operations to migrate the new products to the Long-Term Archive is included in the Long-Term Archive Phase.

Critical Data and Information

Ramapriyan and Moses (2013) provide considerable detail listing the types of data and product information that need to be considered for long-term preservation. Section 3 of Ramapriyan and Moses (2013) breaks down each of the categories providing item descriptions as well as rationale relative to why future users of the products will need each item. LP DAAC Data Scientists have synthesized this extensive list into an LP DAAC Collection Inception Checklist (appendix 2). By mutual agreement, the Producer and the LP DAAC Data Scientist select those items on the checklist that must migrate to the Active Archive for interim preservation.

Applicable Standards

Ramapriyan and Moses (2013) discuss existing archive standards for preservation of information content: "NASA and the international Consultative Committee for Space Data

Table 3. Inception phase work breakdown structure guide—Planning stage.

[WBS, work breakdown structure; LP, Land Processes; DAAC, Distributed Active Archive Center; USGS, U.S. Geological Survey; EROS, Earth Resources and Observation Center]

WBS activity identification	Description of LP DAAC data science activity
Embed in Producer team	Provide data science specialization to Producer team. Attend Producer team conferences and telecommunications. Meet face-to-face with Producer team.
Provide a Data Management Plan	Co-author final Data Management Plan for Producer products providing detail on management of data characteristics, such as: • Naming conventions (collection, granule, information packages). • Formats (tile/granule/document). • Product type/level. • Versioning approach. • Spatial extent, temporal coverage. • Content (scene coverage, bands, resolution). • Compression. • Latency. • Map projections. • Metadata (field names, types, header embedded or standalone). • Browse. • Resource utilization. • Unique services (such as subsetting). • Digital object identifier information. Identify artifacts on Collection Inception Checklist for preservation: • Designate artifacts for Active Archive interim preservation. • Designate artifacts for USGS EROS long-term preservation. Develop preservation policies: • Items archived by other partner agencies. • Version obsolescence rules. • Golden copy sub-setting of each version. • Algorithm history. • Retention of data used in publications. • Differences in depth and breadth between Active Archive and Long-Term Archive of data and information products.

Systems (CCSDS) member space agencies have long recognized the importance of developing information standards for use in long-term preservation of space-related data collections. Volunteers have developed recommendations titled the Reference Model for Open Archival Information System (RMOAIS). Subsequent activities continue to expand through a range of related interests that reach toward more practical guidance for developing agency standards. They include provider-archive interchange recommendations (2004) and packaging of data and metadata (XFDU), to facilitate information transfer and archiving (2008). The most recent update to the OAIS Reference Model is the Recommended Practice "Magenta Book" (2012), and supersedes the Blue Book of 2002. The CCSDS also has developed ISO 16363 that specifies requirements for certification of trustworthy digital repositories, based on the OAIS Reference Model, and ISO 16919 that describes how to audit archives for compliance with the requirements."

Table 4. Inception phase work breakdown structure guide—Production stage.

[WBS, work breakdown structure; LP, Land Processes; DAAC, Distributed Active Archive Center; EOSDIS, Earth Observing System Data and Information System; USGS, U.S. Geological Survey; EROS, Earth Resources and Observation Center; NASA, National Aeronautics and Space Administration; ESDT, Earth Science Data Type; ESDIS, Earth Science Data and Information System; SOO, Science Operations Office]

WBS activity identification	Description of LP DAAC data science activity
Liaison to LP science stakeholders	Serve as advocate for LP science community needs. Attend science team conferences. Routinely liaise with EOSDIS and other sponsors on LP-related science. Obtain DAAC user working group support for new products.
Provide LP DAAC Collection Inception Checklist	Provide checklist for data and information products (appendix 2 of this report) Support tailoring of checklist content as necessary.
Support production	Attend production team conferences and telecommunications. Meet face-to-face with production team. Provide understanding of community standards. Support development of data and information product release plan. Provide understanding of Active Archive and LTA project requirements.
Plan to repeat the experiment	Feasibility study to provide processing capabilities on Active Archive or USGS EROS resources (decide on processing by demonstration or inspection). Capture and package artifacts needed for production processing.
Determine approach to unique tools/services	Identify product unique tools or services. Determine applicable clients. Determine packaging of downloadable tools. Determine migration/support of online tools or services.
Seek authorization to migrate to Active Archive	Receive request for supporting new data products. Receive request for supporting new tools/servic.
Provide NASA Data Template materials	Support development of Data Template for Product Developers. • ESDT title and long names. • Product formats (such as HDF5 or GeoTIFF). • Geophysical parameter specifics. • Descriptions including resolution and temporal/spatial extent. • Algorithm theoretical basis. • Science Need. • Quality and accuracy. • Intended and appropriate use. • Science value endorsements. Provide development of Data Template for LP DAAC and ESDIS SOO. • Heritage (data source, authorization/rationale to support). • Capacity requirements (volume and count metrics). • Active Archive data management principles. • Levels and types of service (user service, ingest method, discovery portals, distribution methods, production support, Intermediate archive approach). • List of product documentation. • Current involvement/responsibility (LP DAAC, Producer). Support development of Data Template for science community. • Science research value relative to climate data record/ earth science data record. • Suggested NASA data management priority, level, and type of service. Forward product and tool assessment materials to ESDIS project for support. Forward product and tool assessment materials to LP DAAC user working group for support. Review/update rejection justifications, support appeal process if necessary.

Beyond generally accepted standards for providing interim preservation, the LP DAAC and Producer agree on specific policies for data management of collections:

- At what point during ingest does a specific version of a product reach significant critical mass and become representative of the Producer collection and allow advertisement of the new version in user client systems and clearinghouses (such as the Earth Observing System Clearing HOuse [ECHO])?
- What Service Level Agreements (SLAs) or Memoranda of Understanding (MOU) exist to cover data and information products that are already permanently archived in other systems?
- How many versions deep shall the archive be for consecutive iterations of the product?
- When can redundant or supplanted versions be purged from the Active Archive?
- What is the policy for retaining a consistent subset of versions (sometimes called a Golden Copy) and associated information (including algorithms) for "version-to-version" change assessment?
- What method is used to preserve products that have been used in scientific publications and ensure complete lineage/provenance tracking?
- When does the Active Archive become the primary source for the products with respect to the LP science community at large?

Transition from Production to LP DAAC

An Active Archive Phase WBS development guide for the DAAC Data Scientist and Producer to execute the migration of products from production systems to the LP DAAC is provided in table 5.

Planning for Transition from LP DAAC to Long-Term Archive

An Active Archive Phase WBS development guide for the DAAC Data Scientist to plan migration of products from the LP DAAC to the Long-Term Archive is provided in table 6. An internal EROS administrative procedure (J.L. Faundeen, unpub. data, 2012) provides the process to obtain approval to migrate an Active Archive collection toEROS sysyems. In a recent change, the final bullet (approval by Programmatic Lead at headquarters) in appendix A of that document is no longer required. At this point in time, Level 1 ASTER and Level 2MODIS products have already been approved for migration but has not commenced. Derivatives of ASTER and MODIS products are not considered approved and will need to follow the procedure for acceptance on a case-by-case basis.

Long-Term Archive Phase

This section provides the framework necessary for the LP DAAC Data Scientists to plan Active Archive activities relative to enabling the most cost-effective method to ensure permanent preservation of new Producer data and information products in the Long-Term Archive.

Long-Term Archive Characterization

All collections offered or maintained at EROS are reviewed through the EROS scientific records appraisal process in order to qualify for EROS LTA archival services. Data and information products must align with the EROS mission to best serve the land remote sensing research community. LP DAAC Data Scientists shepherd new products through the process by offering them to the EROS.

LP DAAC Data Scientists bridge the Active Archive and the EROS LTA Project by executing a plan that begins with a successful appraisal resulting in the EROS Center Director providing the EROS Archivist and LTA Project Manager a decision memo indicating concurrence to assume physical and legal ownership of a NASA product. As part of the plan, the LP DAAC and EROS LTA Project collaborate in the generation of agreement documents such as an MOU defining topics such as the migration timeline, LTA levels of service, modes of access, preservation policy, and sunset policy. The LP DAAC Data Scientist assists the EROS Archivist by providing documentation and materials in accordance with an internal EROS administrative procedure (Faundeen, unpub. data, 2012) to an EROS Appraisal Team, which in turn provides its recommendation to the EROS Center Director. The materials include itemized labor and materials estimates, cost estimates, and formal or informal budget estimates. If the EROS Center Director rejects the plan, the EROS Archivist provides assistance to the LP DAAC in finding a new home for the data, which could include placement in the National Archives and Records Administration.

Upon successful appraisal, the next step is referred to as the Accession process, whereby the collection transitions to the EROS. With regard to LP DAAC accession, the EROS LTA Project will be engaged to determine necessary resources using mutually agreed upon methods. The LP DAAC provides several methods to achieve physical transfer of products to EROS that are compatible with LTA Project techniques. Typically, the metadata for the long-term archive database and the active archive database are nearly identical as well as the data file content and formats. LP DAAC holdings of significant size migrate on a timeline that could range from weeks to many months. LP DAAC Data Scientists monitor and control the migration according to the plan and provide detailed coordination with the LTA Project staff.

Table 5. Active Archive phase work breakdown structure guide—Transition to Land Processes Distributed Active Archive Center.

[WBS, work breakdown structure; LP, Land Processes; DAAC, Distributed Active Archive Center; ESDIS, Earth Science Data and Information System; SOO, Science Operations Office; ICD, Interface Control Documentation; ESDT, Earth Science Data Type; ECHO, EOS Clearing House]

WBS activity identification	Description of LP DAAC data science activity
Obtain authorization to migrate to Active Archive	Assess early versions of products for LP science community need and utility. Determine capacity and throughput requirements. Complete NASA data template, submit to LP DAAC user working group for approval and on to ESDIS SOO where required. Accept authorization to implement new data type.
Plan migration method	Plan the migration timeline. Estimate and budget LP DAAC resource. Provide ICD to Producer. Acquire sample products to drive migration testing. Collaborate with Producer to outline data management policies. Submit migration plan to LP DAAC Operations.
Install new product into LP DAAC core system	Build ESDTs for each new collection. Submit ESDTs to core system maintenance for review and compilation. Enable migration environment in core system test model. Beta test migration according to the ICD. Install new ESDTs in core system operational mode.
Migrate new products	Provide monitor and control services for migration projects to LP DAAC management. Apply mutually agreed policies for data management (such as versioning).
Advertise availability of new products	Apply mutually agreed policies for product release. Release product information to clearinghouse systems (such as ECHO).
Assume primary access and discovery role for new products	Issue notification to stakeholders to release Producer from access and discovery responsibilities.

According to the migration plan, the EROS LTA Project advertises the new products for Access in one or more user client systems as the migration reaches critical mass. During the migration, the LP DAAC continues to provide user access; it is likely that both the LP DAAC and EROS LTA Project will provide user access for a significant period of time as part of the bridging process. At a designated time following the migration, the LP DAAC simultaneously provides notice to the user community that services have transitioned to the EROS LTA Project and discontinues user access services.

The EROS LTA Project is a preservation archive by definition. For a period of time immediately following the migration, both the Active Archive and EROS LTA Project provide dual preservation services for the new products. Eventually, the Active Archive discontinues its preservation role in favor of the EROS LTA Project. Preservation for the new product does not necessarily occur on-site at EROS indefinitely. To achieve permanent preservation status with EROS resources, a new product must be approved to meet criteria supporting the NSLRSDA level. If not so approved, the new product will eventually sunset to the National Archives and Records Administration. An internal EROS administrative procedure (Faundeen, unpub. data, 2012, Appendix B) provides the additional Appraisal criteria used by the EROS Archivist to nominate new products to receive NSLRSDA approval.

Critical Data and Information

Ramapriyan and Moses (2013) provide considerable detail listing the types of data and product information that need to be considered for long-term preservation. LP DAAC Data Scientists have synthesized this extensive list into an LP DAAC Collection Inception Checklist (appendix 2).

The EROS scientific records appraisal process includes steps to prioritize those data and information products that will be permanently archived with EROS resources. Not all data in the Active Archive are of sufficient priority. In coordination with the EROS Archivist, the LP DAAC Data Scientist advocates for the land process science community so that the community is involved in making the decision of prioritizing data collection elements and level of service provided. The

Table 6. Active Archive phase work breakdown structure guide—Approval to transition to the Long-Term Archive.

[WBS, work breakdown structure; LP, Land Processes; DAAC, Distributed Active Archive Center; LTA, Long-Term Archive; USGS, U.S. Geological Survey; EROS, Earth Resources and Observation Center; ICD, Interface Control Documentation]

WBS activity identification	Description of LP DAAC data science activity
Obtain authorization to migrate to LTA	Officially propose a collection for review by USGS EROS Archivist using scientific records appraisal process (select essential data and information products). Provide capacity and throughput requirements. LP DAAC Data Scientist supports and monitors process to migrate to LTA. • Archivist designates a specific USGS EROS project as a candidate for archival responsibility (for LP DAAC this is the LTA project). • Archivist calls for assembly of USGS EROS appraisal assessment team. • Support Archivist in loading key rationalization information into the records appraisal tool • Appraisal Assessment team assesses products for USGS EROS mission alignment and science community utility. • Archivist generates recommendation for EROS senior staff comment. • EROS Director receives comments and accepts, rejects, or modifies the recommendation (rejection results in process to find a new home, which could be the National Archives and Records Administration). • If product is accepted, the USGS EROS Project Manager assumes scope to integrate and accept fiscal responsibilities.
Plan migration	Accept authorization to migrate new data type to LTA project. Collaborate with Archivist to outline data management policies in memorandum of understanding format. Plan migration export/import method. Submit migration plan to LP DAAC operations and LTA project. Plan the migration timeline. Estimate and budget LP DAAC resources. Provide ICD to LTA project. Provide sample products to drive migration testing.

process carefully evaluates the primary aspects of a data collection which are science relevance, spatial coverage, temporal coverage, distribution potential, supporting metadata, risk of loss and availability.

Applicable Standards

USGS is a foundational member of organizations that define standards used throughout the product lifecycle, including the FGDC, ISO, and OGC. In agreement with the ESDIS Project, EROS staff are active in the Earth Science Information Partners (ESIP) Federation's Data Preservation and Stewardship Cluster in proposing and developing emerging standards for Provenance and Context Content. Data provenance, sometimes called "lineage" or "pedigree," refers to the ability to trace and verify the creation of data, how it has been used or moved among different databases, and how it has been altered throughout its lifecycle. Much of the EROS scientific records appraisal process goes to establishing and maintaining data provenance critical to future research. LP DAAC Data Scientists bring awareness of this requirement to Producers at the Inception Phase so that it is accounted for in the Long-Term Archive Phase.

Transition to Long-Term Archive

A WBS for the Long-Term Archive Phase that can serve as a guide for the DAAC Data Scientist to migrate data and information products from the LP DAAC to the Long-Term Archive is provided in table 7. An internal EROS administrative procedure (Faundeen, unpub. data, 2012) provides the process to obtain approval to migrate an Active Archive collection to the EROS LTA Project. The procedure also covers subject matter normally in Operations Agreements (OA) or Interface Control Documentation (ICD) that would be required between external agencies; in this case, the agreement is between several organizations internal to the USGS.

Table 7. Long-Term Archive phase work breakdown structure guide—Transition to the Long-Term Archive.

[WBS, work breakdown structure; LP, Land Processes; DAAC, Distributed Active Archive Center; LTA, Long-Term Archive; EROS, Earth Resources and Observation Center; NSLRSDA, National Satellite Land Remote Sensing Data Archive; USGS, U.S. Geological Survey]

WBS activity identification	Description of LP DAAC data science activity
Enable migration to LTA	Receive approval from center director by way of EROS Archivist. Enable methods to physically transfer products from LP DAAC resources. Confirm methods are in place in LTA to receive products Enable test environments; select representative test cases. Test transfer; benchmark performance and capacity. Test access methods; benchmark user performance; test dissemination.
Execute migration to LTA	Confirm reasonableness of migration timeline. Monitor and control migration.
Advertise availability of new products on LTA resources	Confirm access performance as migration nears completion. Issue notification to stakeholders that LTA provides access. Provide dual access for period beyond end of migration.
Transfer primary access and discovery role for new products to LTA	Issue notification to stakeholders to release LP DAAC from access and discovery responsibilities. Levels of service may differ between the LP DAAC and the LTA because the LTA serves a very wide base of users and products; the notification identifies any major levels of service provided by the LP DAAC that have been discontinued due to transition to the LTA.
Obtain authorization for new products to become part of NSLRSDA	Officially propose a collection for review by the USGS EROS Archivist for NSLRSDA certification. LP DAAC Data Scientist supports development of assessment materials.
Sunset products no longer providing scientific value	LP DAAC Data Scientist supports EROS Archivist to develop materials needed to transfer products to the National Archives and Records Administration (if not NSLSRDA approved).

Selected References

Centre for Atmospheric Science, 1998, Part I—The history behind the ozone hole: accessed March 20, 2014, at *http://www.atm.ch.cam.ac.uk/tour/part1.html*.

Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, and Zolly, L.S., 2013, The United States Geological Survey science data lifecycle model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., *http://dx.doi.org/10.3133/ofr20131265*.

Federal Geographic Data Committee, 2013, FGDC endorsed ISO metadata standards: accessed March 20, 2014, at *http://www.fgdc.gov/metadata/geospatial-metadata-standards*.

Gamble, Lay'wan, 2009, Mission operation procedures for the EMD Project—EOSDIS maintenance and development project: Riverdale, Md., Raytheon Inromation Solutions. (Also available at *http://edhs1.gsfc.nasa.gov/waisdata/rel7/html/611emd220.html*.)

McDonald, Kenneth R., McKinney, Richard A., Smith, Timothy B., and Rank, Robert, 2004, The lifecycle of NASA's earth science enterprise data resources, *in* Conference on Preservation and Validation of Data European Space Research Institute (ESRIN), Frascati, Italy, October 3–8, 2004, Proceedings: Frascasti, Italy, NASA, accessed accessed March 20, 2014, at *http://nix.nasa.gov/search.jsp?R=20040182475&qs=N%3D4294966819%2B4294961986%2B4294908834*.

National Aeronautics and Space Administration, National Oceanic and Atmospheric Administration, 1999, Global change science requirements for long-term archiving, USGCRP, Boulder, Colo., October 28–30, 1998, National Center for Atmospheric Research, 78 p.

Ramapriyan, H.K., and Moses, J.F., 2013, NASA earth science data preservation content specification: Greenbelt, Md., NASA Goddard Space Flight Center. (Also available at *https://earthdata.nasa.gov/about-eosdis/requirements*.)

Tilmes, Curt, and Fleig, Albert J., 2008, Provenance tracking in an earth science data processing system, *in* Freire, Juliana, Koop, David, and Moreau, Luc, eds., Provenance and annotation of data and processes—Lecture notes in computer science: Berlin, Springer-Verlag, v. 5272, p. 221–228 accessed March 20, 2014 *http://dx.doi.org/10.1007/978-3-540-89965-5_23*.

Welch, Charles, 2014, Ozone hole history: accessed March 20, 2014, at *http://www.theozonehole.com/ozoneholehistory.htm*.

Appendix 1. Product Lifecycle Phase Characterization

Each product lifestyle phase deals with phase-characteristics, provided in table 1–1, from a slightly different point of view. The point of view could affect how Land Processes (LP) Distributed Active Archive Center (DAAC) Data Scientists plan activities or create policies that bridge from one lifecycle phase to another depending on the needs of the specific science community being served during the phase. For example, the Inception phase serves a specific science team, the Active Archive phase serves an expanded user base that may not be as familiar with the data and information products as the original science team, and the Long-Term Archive serves users who may come back to the products considerably past the point when they were being used actively in science activities.

Table 1–1. Comparing phase characteristics across the product lifecycle.

[PI, Principal Investigators; RMA, Reliability, Maintainability, Availability; <, less than]

Phase characteristic	Inception	Active Archive	Long-Term Archive
Mission driver	Create science products	Stewardship Expand user base	Permanently preserve data and information.
Opportunity announcement	Identify stakeholder needs	Serve new stakeholders	Stakeholder legacy.
PI proposal	Justify the experiment	Explain the experiment	Preserve the legacy of the experiment.
Sensor	Develop the sensor (if built explicitly for this experiment)	Characterize the sensor domain	Interoperability in the sensor domain.
Calibration	Validate sensor ranges	Characterize accuracy of measurements	Preserve the experiment.
Algorithms	Create the experiment	Repeat the experiment	Inspect the experiment.
Lowest level data	Produce low level data	Capture lowest level data not archived elsewhere with reasonable access latency	Capture lowest level data not archived elsewhere.
Infrastructure	Provide essential science production focused resources	Provide general high RMA resources Provide resources for extended user base	Provide maximum RMA resources. Provide resources governed by long-term user service policy.
Latency requirement	Near real time (<minutes) Active Archive provides services on released products	Quick turn-around (<hours)	Reasonable turn-around (<days).
Development	Provide product algorithms	Capture versions for each major release	Capture final release.
Ancillary inputs	Identify sources	Identify commitment by sources to provide archive and future access Capture inputs from sources not archived elsewhere	Capture inputs from sources not archived elsewhere needed to validate preserved products.
Data management, data backups	Establish naming conventions Establish versioning processes Purge unused data	Provide access to several versions Purge duplicates Provide backup and recovery	Provide access to best version Provide disaster recovery.
Media migration, technology refresh	Media chosen for product development period unlikely to change	It is likely that technology refresh will occur, thus data and information products will be migrated from old media to new media	Technology refresh will occur, thus data and information products will be migrated from old media to new media.

Table 1–1. Comparing phase characteristics across the product lifecycle. —Continued

[PI, Principal Investigators; RMA, Reliability, Maintainability, Availability; <, less than]

Phase characteristic	Inception	Active Archive	Long-Term Archive
Production	Produce products using iterative version cycle	Capture current best products Capture "golden" segments of previous versions	Capture final products. Capture data used in major publications.
Validation	Validate to support publications	Validate to support dataset quality	Document to support data quality.
User base	Restricted to science team during validation	Unrestricted to expanded user base after release	Unrestricted.
Level of service	Focus is on science team needs Some tools may be specifically designed for science team but not applicable to community at large	Focus on science community at large Attempt to preserve science team expectations but limited ability to replicate specialized tools	Level of Service is standardized to range across a large base of data types. Some tools used by the science team may no longer be supported.
Anticipate migration needs	Anticipate Active Archive requirements Anticipate Long-Term Archive requirements	Provide Active Archive requirements Anticipate Long-Term Archive requirements	Provide Long-Term Archive requirements.
Migration planning	Support migration planning to Active Archive	Lead migration planning Support migration planning to Long-Term Archive Monitor and control migration timelines	Support migration planning from Active Archive.
Publications	Provide full range of documents for goals, justifications, publications, and validation as science evolves	Capture document sets related to series of migrated data versions Provide new documentation to meet wider user community expectations	Capture document sets related to final versions migrated.
Data provenance	Gather sources Document workflows	Document provenance in metadata	Ensure provenance is well documented.

Appendix 2. LP DAAC Collection Inception Checklist

Table 2–1 provides a mapping of the LP DAAC Collection Inception Checklist to the most applicable NASA Earth Science Data Preservation Content Specification subsection (Ramapriyan and Moses, 2013).

Table 2–1. Mapping the inception checklist items to preservation specifications.

[LP, Land Processes; DAAC, Distributed Active Archive Center; NASA, National Aeronautics and Space Administration; ESD, Earth Science Data; PCS, Preservation Content Specification]

LP DAAC collection inception checklist	NASA ESD PCS
1.0 Program solicitation	
1.1 Program solicitation description	3.3.1
1.1.1 Definition of program intent, authority, eligibility, opportunity, and scope	3.3.1
1.2 Program solicitation Web site	3.3.1
1.2.1 Official program solicitation Web site	3.3.1
2.0 Project solicitation proposal	
2.1 Project abstract	3.3.1
2.1.1 Overview of project proposal	3.3.1
2.2 Project team	3.3.1
2.2.1 Project team members including name, affiliation, role, and contact information	3.3.1
3.0 Project understandings	
3.1 Proposal requirements	3.3.2
3.1.1 Project requirements as detailed in the program solicitation	3.3.2
3.2 Proposal agreements	3.3.1
3.2.1 Project proposal agreements as related to requirements outlined in the program solicitation	3.3.2
3.3 Proposal modifications	3.3.1
3.3.1 Ongoing notification(s) of any project modifications to the original project proposal agreements	3.3.1, 3.3.3
3.4 ESDIS data template	3.3.1
3.4.1 Completion of data providers sections, which includes science justification, data quality summary and verification(s), intended/appropriate product use, and science value	3.3.1, 3.3.2
3.5 Algorithm theoretical basis document	3.3.4, 3.5
3.5.1 Description of data product creation process	3.3.4
3.6 Citations	3.3.6
3.6.1 Data producer citation	3.3.1, 3.3.6
3.6.2 Non-data producer related citations such as product approach, methods, and development	3.3.1, 3.3.6
3.7 Policies	1.1, 1.2
3.7.1 Description of data access constraints, use constraints, or fees	1.1, 1.2
3.8 Communication	3.3.1, 3.9
3.8.1 Data provider official Web site	3.3.1
3.8.2 Data provider public FTP location of pre-LP DAAC collections available to the public	1.1, 3.3.1, 3.9
3.9 Media	3.3.1, 3.3.6
3.9.1 Materials used for community communication such as journal articles, presentations, posters, conference papers, or workshop packages	3.3.1, 3.3.6

Table 2–1. Mapping the inception checklist items to preservation specifications.—Continued

[LP, Land Processes; DAAC, Distributed Active Archive Center; NASA, National Aeronautics and Space Administration; ESD, Earth Science Data; PCS, Preservation Content Specification]

LP DAAC collection inception checklist	NASA ESD PCS
4.0 Data product materials	
4.1 Complete information for all data suites per project agreements	3.2.2
4.1.1 Type/format	3.2.2
4.1.2 Name convention	3.2.2
4.1.3 Version description	3.2.2, 3.3.3, 3.3.4
4.1.4 Processing location	3.2.2, 3.3.1
4.1.5 Platform(s)	3.2.2, 3.4.2
4.1.6 Instrument(s)	3.1.1, 3.2.2
4.1.7 Band description(s) including name, data type, numerical range, void value, and unit of measurement	3.1.1, 3.2.2
4.1.8 Temporal coverage	3.2.2, 3.5
4.1.9 Geographic extent	3.2.2, 3.5
4.1.10 Scene coverage	3.2.2
4.1.11 Image dimensions including rows and columns	3.2.2
4.1.12 Total number tiles/granules	3.2.2
4.1.13 Data file statistics including minimum, maximum, and average file size	3.2.2, 3.5
4.1.14 Title volume	3.2.2
4.1.15 Compression type	3.2.2
4.1.16 File format(s)	3.2.2, 3.3.2
4.1.17 Map projections and datum	3.2.2, 3.3.4
4.1.18 Quick start	3.3.1
4.1.19 User guide	3.3.1
4.1.20 Product keywords	3.3.1
4.2 Digital object identifier information	3.3.1
5.0 Product data	
5.1 Complete product versioned data delivery	3.3.3, 3.3.5
5.1.1 Data files	3.2.1, 3.3.2
5.1.2 Ancillary data/files	3.3.4, 3.6
5.1.3 Masks	3.2.1, 3.3.4, 3.6
5.1.4 Headers	3.2.2, 3.3.2
5.2 Algorithm source code	3.3.4, 3.8

Publishing support provided by:
Rolla Publishing Service Center

For more information concerning this publication, contact:
U.S. Geological Survey Earth Resources Observation
and Science (EROS) Center
47914 252nd Street
Sioux Falls, South Dakota 57198
(605) 594-6151

Or visit the EROS Center Web site at:
http://eros.usgs.gov/

ISSN 2331-1258 (online)
http://dx.doi.org/10.3133/ofr20141139