ESDS Reference Architecture for the Decadal Survey Era

ESDS Reference Architecture Working Group

January 6, 2012
Version 1.1

Table of Contents

Change History

Version	Date	Section	Summary of Changes
v0.6	*2/2/2011*	*All*	*Initial release to review team. Includes fully annotated outline, Table of RA Views, initial Stakeholder Survey, Mission Use Case Survey, Function diagram, Glossary and Acronyms*
v0.7	*2/18/2011*		*Insert Change History*
			Add Header
			Augment annotations with input from sjk
		5.0	*Separate Archive and Distribute functions. Provide initial draft of section text.*
		5.1.1	*First draft of text and diagram*
		5.1.2	*First draft of text and diagram*
		5.1.3	*First draft of text and diagram*
		5.1.4	*First draft of text and diagram*
		5.1.5	*First draft of text and diagram*
		5.1.6	*First draft of text and diagram*
		5.1.7	*Added subsection, First draft of text and diagram.*
		Acronyms	*Added OGC*
v0.8	*4/11/11*		*Augment annotations with input from ICESAT-2*
		6.x & 7.x	*Switch System and Information View sections*
		7	*Rename to System Service View*
		6	*Initial content, including domain model figure and table of definitions.*
			Augment annotations with input from James Gallagher
			Augment annotations with input from Patrick Denny
		Table 2	*Clarifying descriptions: System Architect, Program Management*
		5	*Wording changes for clarification*
		5.1.3	*Improve wording*
V0.9		*7.x*	*Clean up heading levels*
		7.1	*Initial content, add Table*
		7.2	*Initial content, add Table*
		7.3	*Initial content, add Table*
		Table 1	*Reorder rows, update definition of System Service View*
		6	*Update wording*
		Table 5	*Update definitions for: Collection, Dataset, QualityMetadata, Subscription, WebServiceInterface*
		Table 4	*Update description of Archive Data*
		InformationView	*Add ComputingResource and Statistics*
			Remove Product Value Chain View

V 1.0	*6/30/11*	*1*	*Complete section text*
		2	*Complete introduction text. Complete Table2.Removed Roles and Concerns subsections. Completed Views subsection and table.*
		3	*Completed text.*
		4	*Complete introduction text*
		Figure 6	*Added step for "Develop Algorithm Improvement"*
		10	*Remove "Implementations" section.*
V1.1	*12/31/11*		*Complete Technology View*
			Incorporate Security drivers
			Emphasize opportunities for real-time access and predictions

1 Overview

1.1 Background

Over the last few decades, advances in geophysical measurement technologies have led to an explosion in the volume and scope of Earth Science data sets. Implementation of the instrument systems that acquire these measurements requires commensurate development in data systems to enable the collection, processing, archiving and distribution of the resultant large volumes of data. As more and more of the data systems come on line, their common features have become increasingly apparent. NASA's Earth Science Data System Working Group (ESDSWG) recognizes these trends and has asked its Standards Process Group (SPG) to develop a reference architecture for NASA's Earth Science Data Systems (ESDS's). That reference architecture is captured in this document. The NASA ESDS Reference Architecture describes those common features and patterns, and thus provides guidelines for individuals and teams who plan to construct new NASA ESDS's.

The driver behind this effort is the upcoming set of Earth Science Decadal Survey Missions that NASA plans to implement. The instruments that this effort proposes will generate an unprecedentedly large volume of data that could not be handled a decade ago. The provision of guidelines based on previous experience should enable stakeholders to plan and design a complete and functional system more quickly and thoughtfully, and thus reduce implementation effort and cost. Although the initial intent of the architecture is intended for use in the Decadal Survey Missions, the components of the reference architecture are adequately flexible for use in any data system that handles NASA's Earth Science data, large or small, complex or simple.

1.2 Context and Purpose

This Reference Architecture elucidates a data model, a set of functions and services that are common among data systems that operate on NASA's Earth Science data. The reference architecture thereby provides individuals and teams who wish to construct, or interoperate with, Earth Science data systems with a concrete strategy for the development and implementation of higher-level design. The architecture specifies sets of uncoupled, exchangeable and replaceable services. These uncoupled services provide users of the architecture with adequate generality and flexibility to compose the requisite system that adheres to project requirements. Implementations based of these flexible components ultimately enable system designers to update and modify their systems as requirements and technologies change and evolve.

It should be noted that this document is focused on ESDS's within the purview of NASA. Therefore any reference to Earth Science Data, ESDS's, requirements and other facets of

ESDS's are intended solely to be in the context of NASA's investments and resources in the Earth Observation and Science fields.

1.3 Fundamental Concepts

Contemporary programs and the systems that support them can be very complex. The current state of practice is to represent these systems from multiple perspectives, in views[1]. Each view is intended to be complete on its own, for the context it represents, but orthogonal to (and therefore not in conflict with) any other view. Mappings are often provided to capture the relationships between various facets of different views.

The following table identifies the architecture views that will be used to represent the ESDS Reference Architecture, and what each of those views represents.

Table 1 - ESDS Reference Architecture Views

View	Description
Function View	*What functions happen within the ESDS, their relationships and consumers of those functions*
Information Architecture View	*The structure and content of the information managed within ESDS.*
System Service View	*The abstract services supported by ESDS implementations for both system and user interaction.*
Technology View	*The standards and tools that are commonly used within instances of the reference architecture*

2 Stakeholders

Stakeholders represent research and development efforts associated with Earth Science Data as well as the organizations that sponsor those efforts. As such, each stakeholder has specific interests in particular features of a data system. By leveraging principles, patterns and concepts that typically impact data system development, stakeholders dictate specifications and requirements that determine how data system components are implemented.

This Reference Architecture is based on the interests of a specifically identified set of Stakeholders. These stakeholders typically begin their specification of requirements by considering components commonly found in historical data systems as well as by considering their experience as users of those historical systems. Table 2 lists the stakeholders this Reference Architecture identifies. The table then describes the role of each of these stakeholders and addresses their specific concerns as they relate to the Reference Architecture.

[1] For example, the Open Group has defined TOGAF, an architecture framework for representing architectures. (http://www.togaf.org/)

Table 2 - NASA ESDS Stakeholder Survey

Stakeholder	Description	Concerns
System Architect	The System Architect is responsible for establishing the vision of a system that will meet specific needs (within a program), the approach for realizing that vision and ensuring that it is properly fielded. Normally this role is for systems targeting an operational capability instead of a short-term throw-away effort, for example, a prototype. The System Architect typically will evaluate the target system and put it in the context of the Reference Architecture, first ensuring that the Reference Architecture is pertinent, then understanding how best to leverage it.	• Scope of target system • Relationship to Reference Architecture (How does what I am responsible for building relate to the ESDS Reference Architecture? What can I leverage or reuse?) • Integrating with/extending Reference Architecture (How does what I am responsible for building work with other implementations of the ESDS?) • Information and standard reuse
Data Architect	Data Architects are responsible for information modeling within systems. They are concerned with capturing the information used within a system, from conceptual through physical deployment.	• Data (and metadata) exchange • Information model reuse
Security Architect	The Security Architect is responsible for ensuring that the security of an information system's operation is commensurate with the risk and magnitude of harm that might result from a compromise of that system's confidentiality, integrity, and availability.	• Risk Assessment and Mitigation • Information Access • Data Integrity
Program Management	Programs managers establish and track plans, budgets, processes and governance mechanisms within programs. These responsibilities begin during the Inception phase, where potential programs are conceived and funding decisions are made, through deployment and maintenance.	• Formulation of program vision • Scope of target system and its components • Cost estimation by reusing concepts, services, components and standards
Project Scientist	Individual responsible for defining and delivering the science mission requirements. Particular focus on understanding what resources (data, services, applications) are available to leverage and how to deliver new data to users.	• System architecture concepts • Leveraging existing systems and their components
Policy Researcher	Policy researchers are interested in higher level products and applications that can be used for analysis, decisions and formulation of policy.	• Leveraging existing systems and their components
Research	Individual or small team who use Earth observation	• Leveraging existing

Scientist	resources in the pursuit of their own science goals. While the resulting science may be introduced into the system as shared resources that may not happen, or may happen very sporadically.	systems and their components
NASA Management	Responsible for mission fulfillment, policy adherence and budgetary concerns. These managers ensure that maximum value is being derived from existing investments. They identify high value areas of future investments and validate proposed solutions in the context of existing enterprise resources.	• Formulation of program vision • Enterprise Architecture, opportunities to optimize investments • Cost estimation by reusing concepts, services, components and standards

2.1 Views

The NASA ESDS Reference Architecture is represented from a set of views. These views are defined in Sections 5 through 8. Each view represents a different facet of NASA's ESDS's. This section identifies the architecture views that are most relevant to each stakeholder.

Table 3 - Reference Architecture Stakeholder mapping to views

Stakeholder	Functional	Information	System Service	Technology
System Architect	X	X	X	X
Data Architect		X		X
Security Architect	X	X	X	X
Program Management	X			X
Project Scientist	X		X	X
Policy Researcher		X	X	
Research Scientist	X	X	X	X
NASA Management	X		X	X

3 ESDS Reference Architecture Principles

The following principles provide a basis for decision making of architectural choices for ESDS systems.

3.1 Architecture Drivers

Architectural Drivers for ESDS Reference Architecture come from several sources:

- NASA Earth Science Data System Needs identified in the NRC Decadal Survey
- The Science Data Systems in the Decadal Survey Era Workshop held on June 25-26, 2009
- NASA ESDS Level 0 and 1 Requirements
- NASA Information Security Policy (NPD 2810.1D) and associated NASA Procedural Requirements (NPR 2810.1A)

The NRC Decadal Survey contains several recommendations that directly affect the architectural choices for upcoming ESDS data systems. These recommendations include provisions for:

- new and updated processes,
- flexibility to implement changes in data architectures
- new functional capabilities that handle applications that have not yet been realized, and
- the adoption of new technologies.

In addition, the ESDS Level 0 and 1 Requirements document covers a broad set of system and architecture level requirements. Since the release of these documents, designers and developers have discussed the challenges facing the Decadal Survey Missions with respect to technology, reuse and standards at various venues.

Ongoing documentation and discussion lead to a specific set of architectural drivers. The following bullets summarize those drivers:

- **An Explosion in Data Volume**
  NASA ESDS processing and storage requirements will increase significantly. Handling these data will require more automation, scalable high capacity computational and storage systems as well as advanced data movement techniques.
- **More Complexity**
  Requirements for mission operations, scientific instruments, and data are becoming more complex. To meet these requirements, new data systems will need an improved and well-defined information model for managing, processing, and archiving diverse and distributed products as well as a modern online data dictionary.
- **More Interfaces**
  Each additional mission adds to the diversity of data products and data providers. As data sources and contents diversify, the need for a streamlined and an easy-to-use standard architecture becomes more critical. Disparate systems will need to deliver data in standard data formats. Methods that manage data across platforms as well as assimilation and analysis tools will be necessary to design, generate, validate, and deliver quality data sets.

- **Greater User Expectations**
  Earth Science researchers and data users anticipate seamless and easy access to multiple data streams from the Decadal Survey Missions, as well as from other operating missions and current and heritage data held in EOSDIS. Modern data users expect well-documented data to be readily available via text-based or graphical search systems. Users also expect data delivery in a variety of formats compatible with their data processing systems accompanied by tools that display and analyze discipline-specific data.
- **More Collaboration**
  The current ESDS DAACs operate somewhat autonomously and independently. The DAACs share little of their technical expertise in this heterogeneous environment. ESDS needs to establish operational specifications that enable distributed and shared services across all organizations and missions, and thereby ensure a set of common tools that can plug into all local environments. Thus, common data infrastructure services should be provided where they make sense. Examples of common services that might benefit include ingest, cataloging, discovery, distribution, assimilation and analysis.
- **Dynamic Environment**
  In the dynamic environment of NASA Earth Science endeavors, securing transactions, data, and infrastructure components can be complex. The architecture must provide the security elements needed to support the evolving IT infrastructure, emerging legislative regulations, and ever-increasing threats. An effective security architecture combines policies and leading practices, technology, as well as a sound education and awareness program.

3.2 Architecture Principles

Large-scale software systems typically have explicit principles that guide the evolution of components and the systems for large enterprise teams. The following principles have influenced the development of the ESDS Reference Architecture:

Data Model Driven: Data design and maintenance should be based on a conceptual information model that is implementation independent. The model employs a common vocabulary so data are described consistently throughout the system and the definitions are understandable and available to all users.

Usability: Software must be easy to use. An intuitive user interface for software and associated tools is critical. Data should be accessible in a timely manner for users to perform functions, regardless of where the data are located. Data lineage and provenance must be readily available. Software tools should enable integration of data from multiple sources.

Reliability: Maintain ESDS operations regardless of system interruptions. Nominal operations should recover quickly after hardware failures, natural disasters and detected data corruption.

Modular and Evolutionary Development: To facilitate phased development and deployment, the system should be decomposed into manageable elements and components. Separate, self- contained components, offering access to their capabilities through services, enable the evolution of the system while preserving the other operational parts.

Technology Independence: Software must be independent of specific technology choices and therefore can operate on a variety of technology platforms. Separation of software from underlying technologies enables cost-effective development and maintenance.

Scalability: The system must be scalable to accommodate increased data volume and complexity. Extensible systems enable easy integration of new tools, components and COTS products.

Common Use Software: The development and integration of system-wide applications and APIs is far more effective than duplicate development of common functions.

Interoperability: Software and hardware should conform to appropriate standards that promote interoperability for data, applications, and technology. Use of industry standards promotes consistency and maintainability of services as well as encourages the adoption of open source or vendor solutions.

Security: Information technology security measures are in place. The system is compliant with the NASA Information Security Policy. Security procedures must incorporate ongoing risk analysis and mitigation to balance user expectations against policy constraints.

4 Reference Requirements

While each NASA ESDS has its own unique requirements, the role of a reference architecture is to identify those mechanisms, patterns and approaches that are in common across implementations of the reference architecture. The NASA ESDS reference architecture has identified a set of mission-level use cases that represent the common scope of all NASA ESDS's. Table 4 offers a survey of those mission-level use cases. The requirements represented by these use cases provide a set of measurable constraints on the architecture by which conformance can be determined. Not all ESDS's will be responsible for all the functionality in all of these use cases. However, any system that may want to leverage this reference architecture should have some overlap with at least a subset of these use cases.

The Use Case Survey in Table 4 labels and identifies the six basic use cases, offers a short description of them and identifies the set of stakeholders engaged in the realization of each use case.

Table 4 - ESDS Mission Level Use Case Survey

UC #	Name	Description	Stakeholders
ESDS-1	Receive Sensor Data	Remote Sensing resource sends data to Ground Systems.	System Architect Data Architect Project Scientist Data Provider
ESDS-2	Develop Products	Using algorithms and other services, data is transformed into higher value products. Includes mechanisms to generate real time or near real time data products.	System Architect Data Architect Project Scientist Data Provider Program Management
ESDS-3	Distribute Products	Generated products are made available to end-consumers. This includes the discovery and access of data using tools, as well as the actual delivery of data to endpoints. Includes mechanisms that enable access to real-time, or near-real time data products, if and when appropriate.	System Architect Data Architect Data Provider Program Management
ESDS-4	Develop Predictions	Using data products and models, predictions are generated, saved and made available through the system, much like any other data product.	System Architect Data Architect Research Scientist Policy Researcher
ESDS-5	Manage Remote Sensor/Instrument	The control and command of remote sensing resources.	System Architect Project Scientist Program Management NASA Management
ESDS-6	Data Stewardship	Keep data resources preserved and available for the long term (beyond mission life).	System Architect Data Architect Data Provider NASA Management

5 Functional View

The Functional View is intended to identify those high-level functions that are supported by systems that implement the reference architecture. The functions identified within the Functional View are intended to be comprehensive in that they cover all the functions that a fully capable instance of the reference architecture will offer.

For the ESDS Reference Architecture a set of functions have been identified to cover the lifecycle of earth science information, from acquisition to archival, from algorithms to access, and supporting stakeholder interactions with ESDS's. There are a number of functions that support broader enterprise goals, and are closely related to ESDS goals. In order to focus this reference architecture to those functions that are the primary responsibility of ESDSs, only the functions that are wholly the responsibility of ESDS's are included in the Functional View.

Figure 1 represents the entire suite of functions considered in the development of the ESDS Reference Architecture. Those that are colored blue are important, but not entirely within the responsibilities of an ESDS. Those that are brown are those considered fully ESDS functions and comprise the ESDS Functional View.

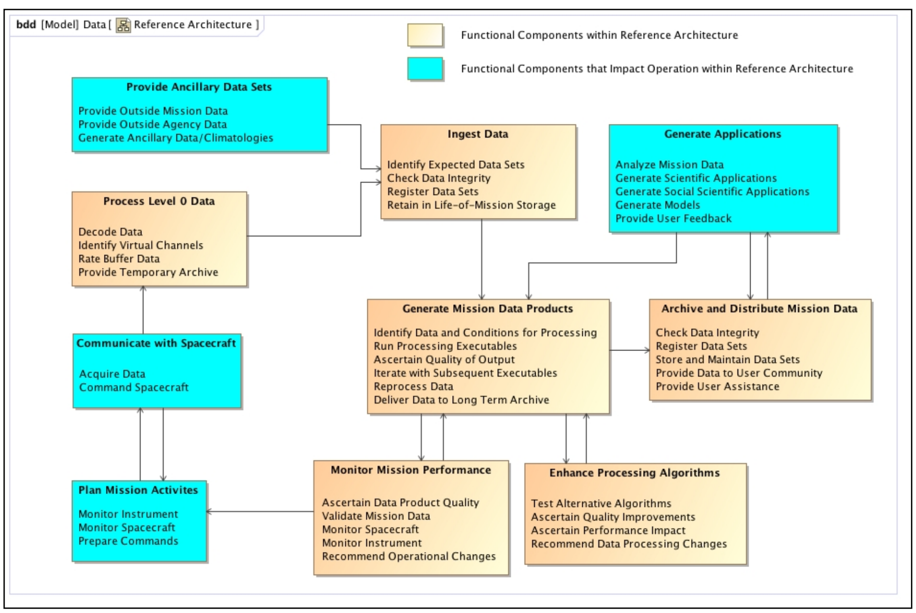

Figure 1 - ESDS Reference Architecture Functions

The following table, Table 5 – ESDS Function Descriptions, provides summary descriptions for the ESDS functions. Each of these functions is discussed in more detail in the remainder of Section 5.

Table 5 – ESDS Function Descriptions

Function	Description
Process Level Zero Data	Decode, separate and rate buffer data. Provide temporary storage
Ingest Data	Identify requisite data, check data integrity, register and store data in Life of Mission Storage
Generate Mission Data Products	Identify available data for processing, run executables, check for successful completion, iterate through processing sequence, upon approval deliver data to long term archive, reprocess data, generate real time or near real time products.
Monitor Mission Performance	Ascertain data quality, validate mission data, monitor spacecraft and instrument performance, recommend operational changes
Enhance Processing Algorithms	Test alternative algorithms, ascertain quality improvements, ascertain performance impact, recommend data processing changes
Archive Data	Check data integrity, register, store and maintain data sets, metadata and documentation.
Distribute Data	Provide data and commensurate assistance to the user community to support access to that data. May include receipt of real time or near real time data products.

5.1 ESDS Functions

The following subsections provide further discussion on each of the six functions that comprise an ESDS. Every instantiation of this reference architecture may not need to provide all of these functions, nor any one function in its entirety.

5.1.1 Process Level Zero Data

Data within satellite-based remote sensing systems begins at the sensors. Sensor data is collected on a satellite, an aircraft or perhaps from an in-situ location. Satellite data is normally the most voluminous. Those data are transmitted to terrestrial systems as raw data in IF signals. Antenna systems and network environments support the movement of that raw data into the realm of Earth Science Data Systems. The initial step of the conversion of these data into useful Earth Science is the function of Processing Level Zero Data.

The remote sensing community has a well-established practice of categorizing satellite data into "levels". These levels correlate to a nominal level identify various stages of processing – from raw sensor data to highly processed and integrated data products. Level 0 data are full-resolution raw instrument/payload data. They are intended for further processing to higher levels, not for direct use.

The following figure depicts the nominal activities that comprise the Process Level Zero function.

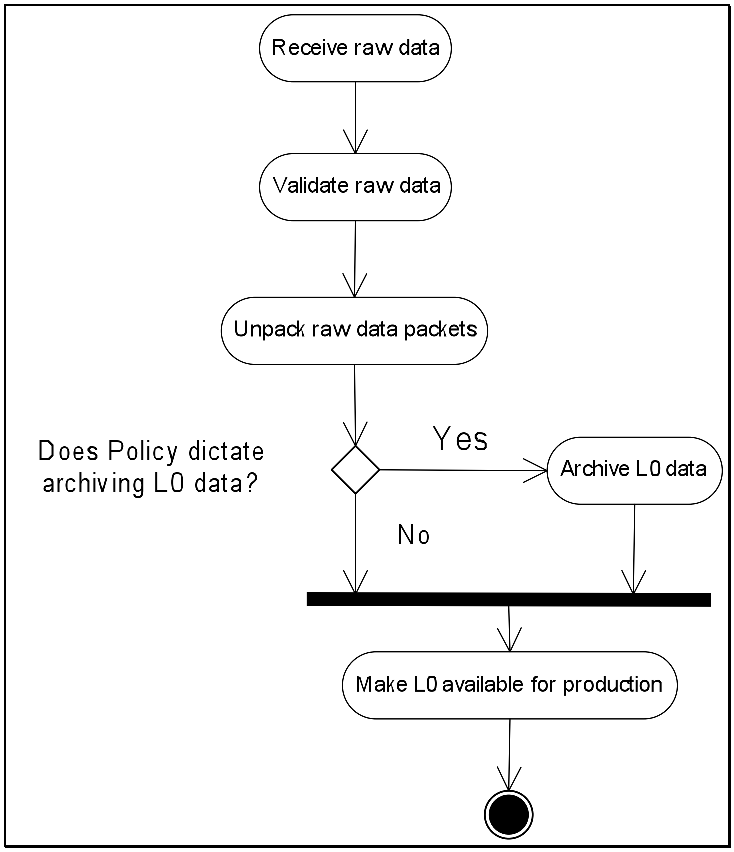

Figure 2 - Process Level 0 Data Activity Diagram

5.1.2 Ingest Data

Systems that process large quantities of data require considerable organization of resources. Without such organization, the system will not be able to identify individual data sets and employ them in appropriate processing steps.

Large numbers of data sets may arrive at the processing system facility. These include data sets associated with the mission instrument, as well as ancillary data sets that are required to generate various science data products.

The system identifies each arriving data set, and determines whether these new data are needed for processing. If the new data are needed, the system checks the integrity to ensure that the contents are usable. If the newly arrived data are indeed required and usable, the system catalogs each data granule with identifying information, and makes the data set available for processing.

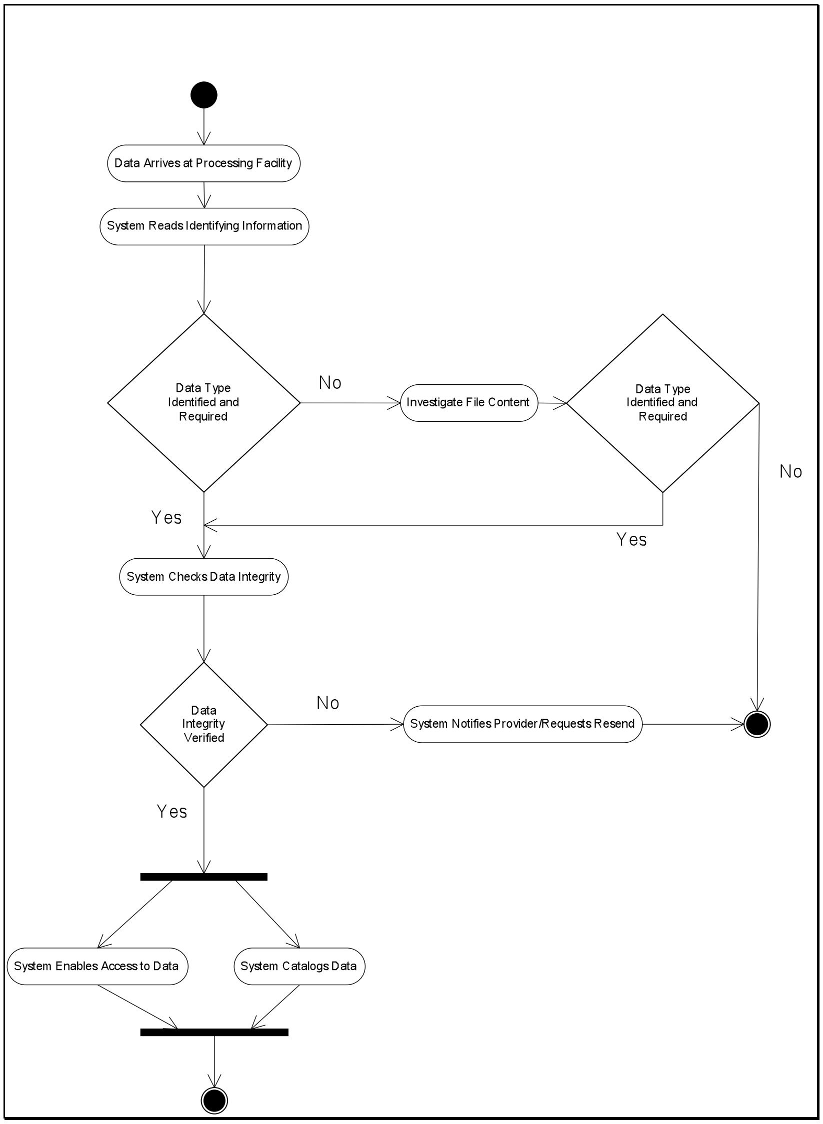

Figure 3 - Ingest Data Function Activity Diagram

5.1.3 Generate Mission Data Products

Systems that generate quantities of data require a means to organize each process. Without such organization, the system cannot ensure that all incoming data are processed through every required step of the processing chain.

The Generate Mission Data Products function describes the instantiation of a single step in a processing chain. In a typical architecture, this step repeats multiple times. Each instance represents one process in a workflow or a pipeline that transits the incoming data into various higher-level data products.

The function checks whether all conditions are in place to run a particular process. In most cases, conditions that constitute readiness are based on the availability of requisite data sets. If detected conditions indicate readiness to run, the system employs a set of rules to assemble all the necessary materials in a configuration that the processing step requires. The system then runs the process.

Upon completion, the system ascertains whether the process ran successfully. Regardless of the level of processing success, the system determines whether the process output displays acceptable quality. The system ultimately records and stores the output. At the same time, the system registers the output in the system catalog with information that denotes processing success and product quality.

This ESDS function provides the capability to generate real time or near real time data products.

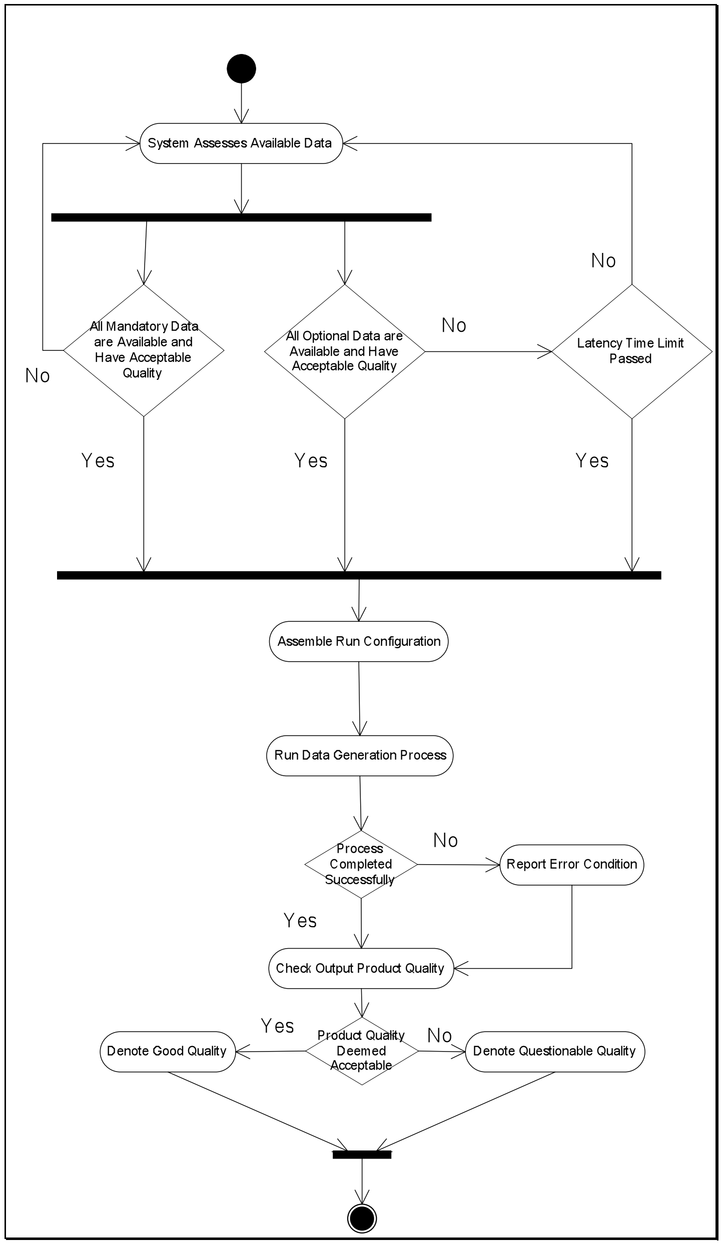

Figure 4 - Develop Mission Data Products Activity Diagram

5.1.4 Monitor Mission Performance

The previous functions allude to problems with data storage and processing that might take place. Those functions do not provide specific procedures that the system might adopt if processing should fail, or if data assessment should indicate unacceptable quality.

The Monitor Mission Performance step provides guidelines to handle these situations. Operational decisions depend on the nature of the detected problem. For instance, anomalies that do not appear to be systematic are most often due to some peculiarity in the input data. When situations are not systematic, and the data associated with those conditions normally require annotation but no additional work. Problems that appear to be systematic require further investigation. The root of systematic problems may rest in ground processing, but sometimes these problems are more closely aligned with the flight or instrument systems. Thus, the larger mission team cooperates to assess the source of the detected problem, in order to ascertain a workable resolution for subsequent data acquisition and processing.

If the quality of already acquired data can be improved using process modification and reprocessing, the team may choose to take those steps. If the problem cannot be resolved for data that were acquired during the anomaly period, the system denotes the problem for the associated data in the catalog. If those data are deemed usable, the system will make them available to the user community with appropriate warning information.

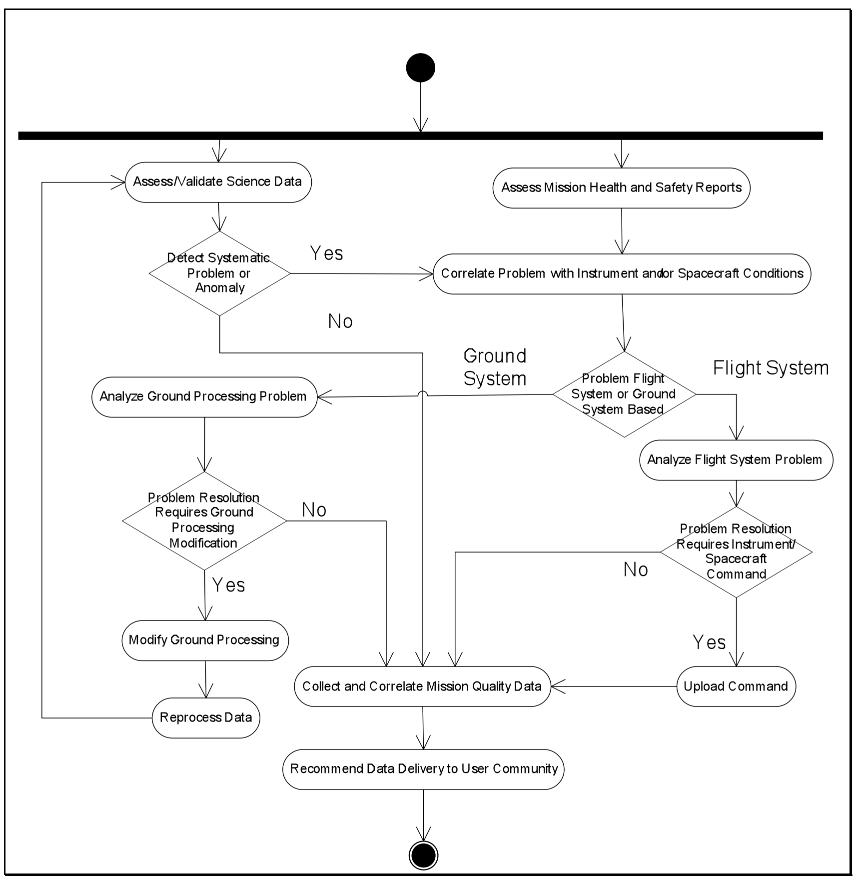

Figure 5 - Monitor Mission Performance Function Activity Diagram

5.1.5 Enhance Processing Algorithms

Processing Algorithms within an ESDS run whenever their required input data is available. Over time opportunities for improving, or enhancing the algorithms become apparent. The recognition of these opportunities and the activities that result in making these enhancements constitute Enhance Processing Algorithms function.

Mission engineers and scientists regularly monitor the processing algorithms that generate products. If members of the mission team identify a flaw in the algorithm or a means to enhance sagging algorithmic performance, the team may choose to investigate algorithmic improvements. While the data system continues to operate, mission team members attempt to design and implement alternative algorithms that can resolve the perceived problem.   If the implemented improvement shows promise, the team will integrate the modified algorithm in a test environment and validate the changes. A validated upgrade must be approved before it becomes operational.

The following figure depicts the nominal activities that comprise the Enhance Processing Algorithm function.

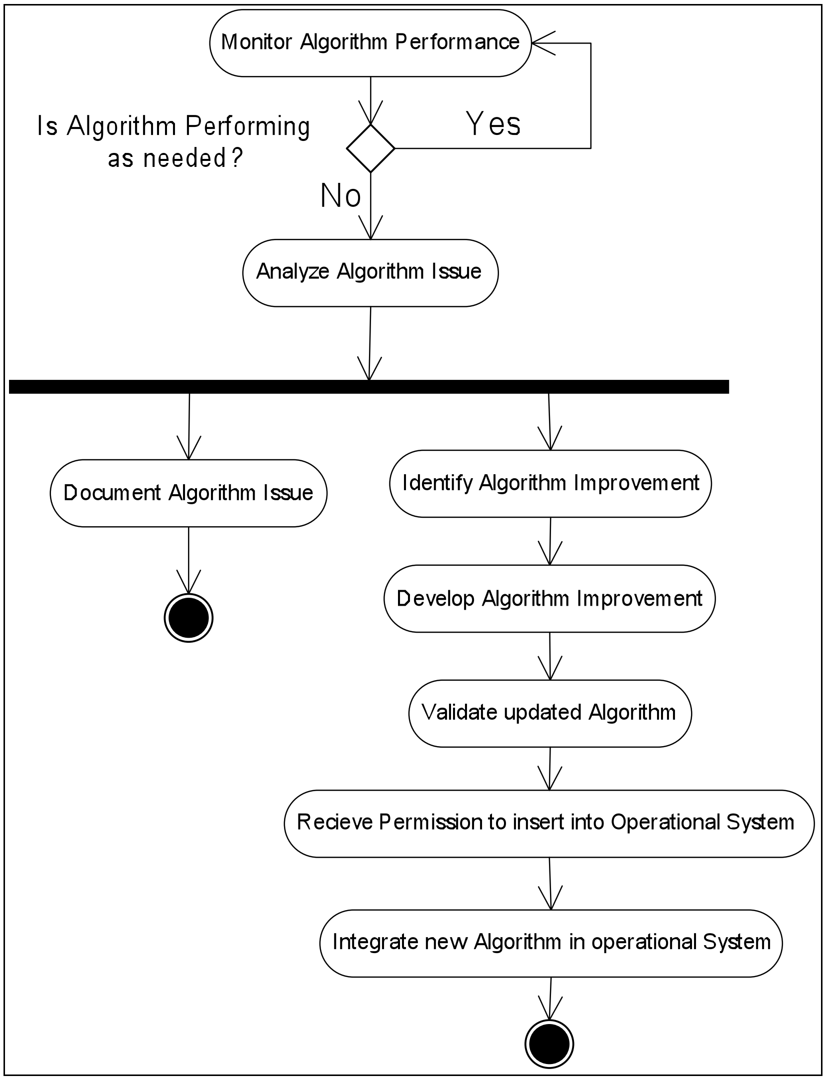

Figure 6 - Enhance Processing Algorithm Activity Diagram

5.1.6 Archive Data

Archive Data is the fundamental process for securing the Earth science data for perpetuity within ESDS, as well as cataloging and making the data available for access and distribution. The archive function also includes examination of the data and associated descriptive information to ensure that quality meets ESDS requirements

The following figure depicts the nominal activities that comprise the Archive Data function.

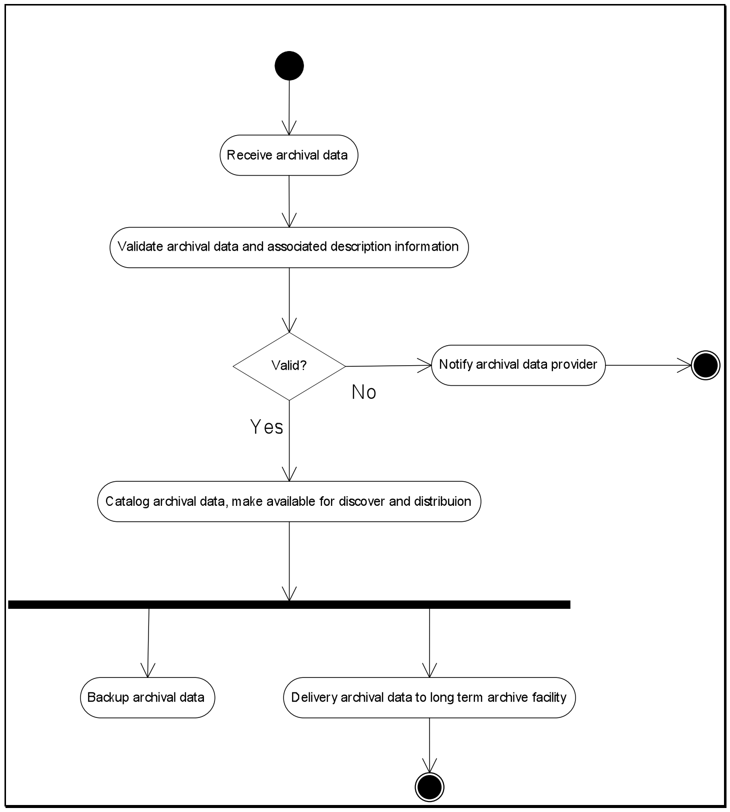

Figure 7 - Archive Data Activities

5.1.7 Distribute Data

The ultimate purpose of the ESDS is the generation of data sets that enhance Earth Science knowledge and applications. Distribute Data is the fundamental process that enables users to identify, locate and receive the data sets they need for their investigation and research. Distribute Data manages when, how and to whom data are to be distributed, as well as provides users with the capability to query the data holdings to find desired data. Access provides the ability to extract specific pieces of information from a data holding, then repackage that information in a different format and make it accessible to a user. The Distribute data function also provides resources to answer user's questions regarding issues of the data access. In some cases, distribution of data may employ broadcasting technology (either terrestrial or satellite-based), to support real-time access.

The following figure depicts the nominal activities that comprise the Distribute Data function.

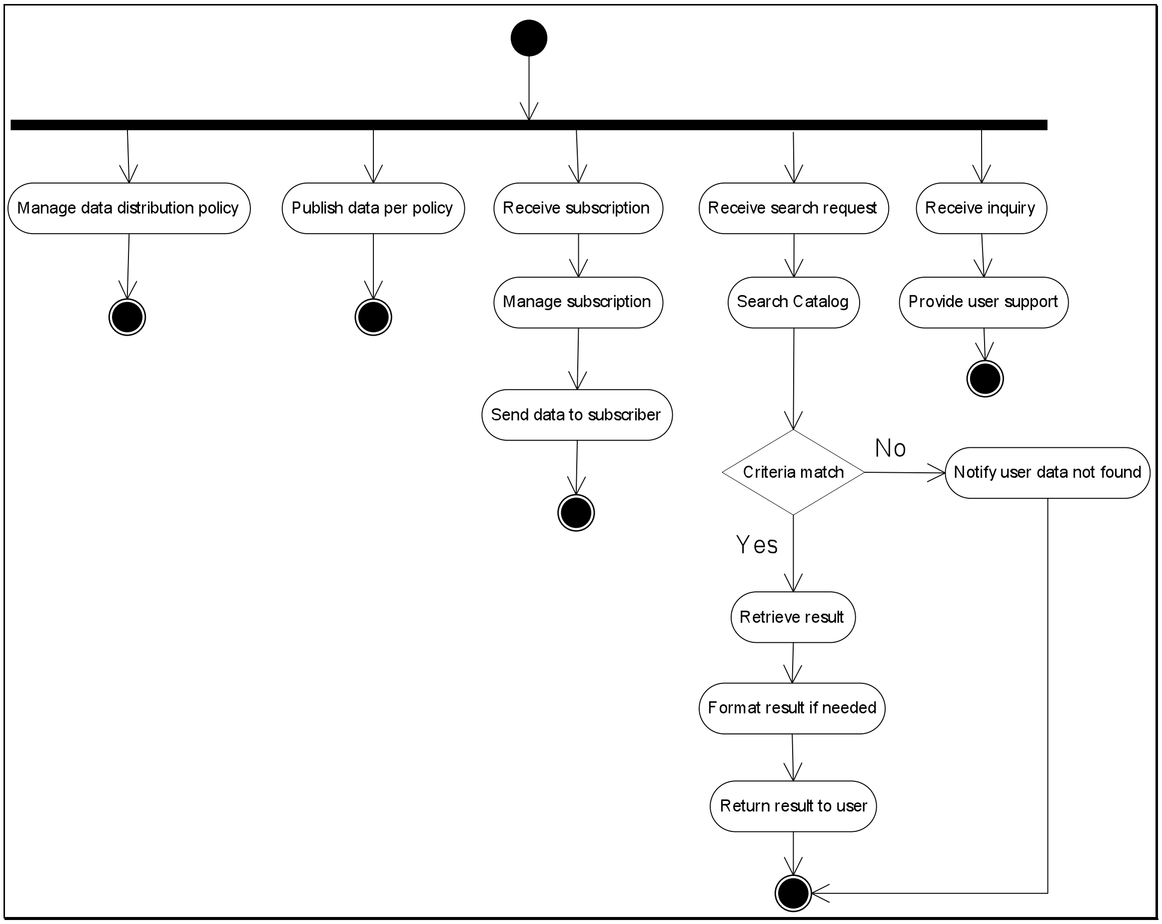

Figure 8 - Distribute Data Function Activity Diagram

6 Information Architecture View

The Information Architecture View is intended to identify the high level information entities that are implemented within, and exchanged between systems that implement the reference architecture. It also defines the relationships between those information entities. Consistent with the goal of providing guidance rather than a prescriptive solution, the entities within the Information Architecture View are intentionally captured at a high level of abstraction. This Information Architecture View will provide a common language between ESDS implementations, thereby maximizing the potential for interoperability and harmonization of those systems and the information they exchange.

The ESDS Reference Architecture Information View is based on a domain model established by the Working Group on Information Systems and Services in the CEOS community. There is significant commonality between this domain model and the legacy experience of NASA's Earth Science information systems. The Information Architecture View has been captured in two forms. Figure 9 - ESDS Information Model depicts the

ESDS Information Model. This shows the entities and their relationships. Relationships captured include composition, aggregation, generalization/specialization and simple linked relationships. Where appropriate, the multiplicity of those relationships is included. Table 6 provides definitions of each entity within the ESDS Information Model.

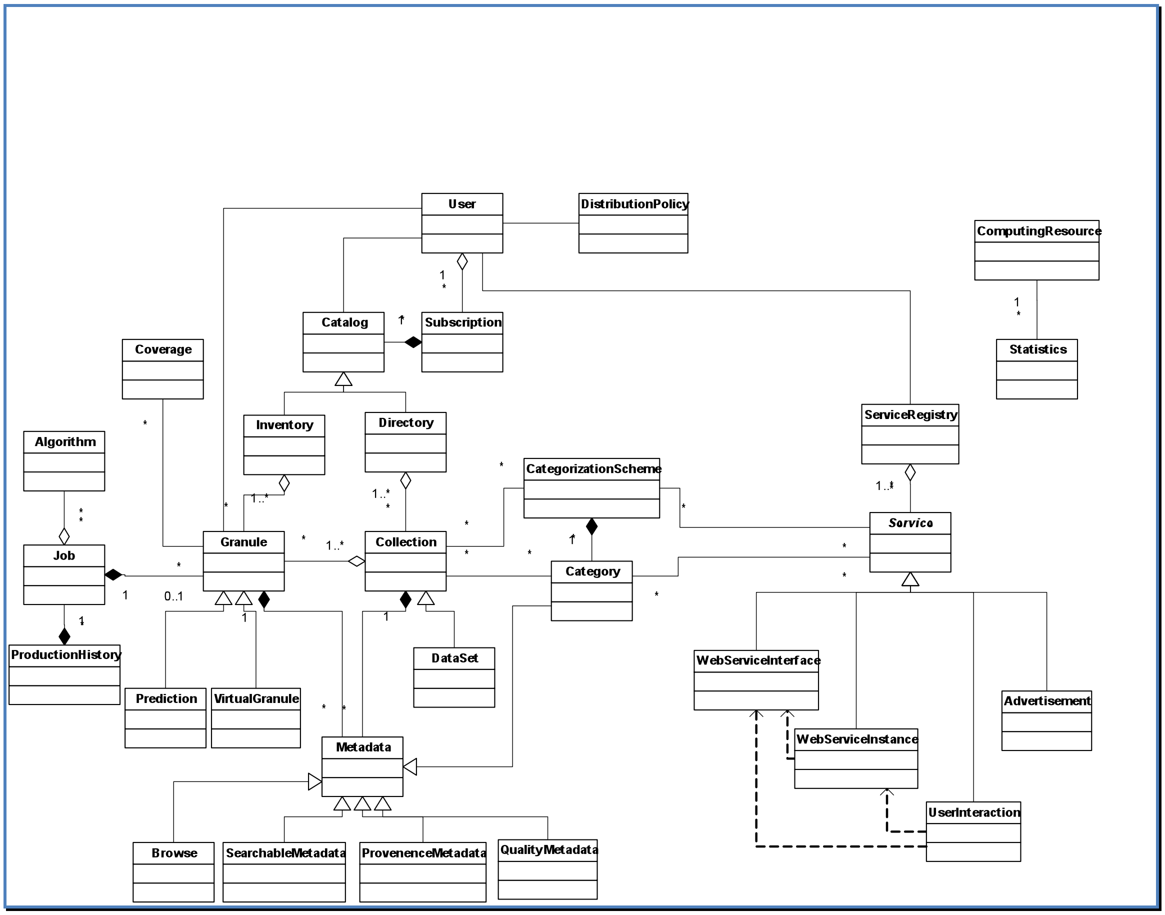

Figure 9 - ESDS Information Model

Table 6 – ESDS Information Model Definitions

Element	Description
Advertisement	A representation of a service that does not directly interface with software systems through interfaces. An example of an Advertisement might represent an organization's expert ability to utilize MODIS Ocean data, or the ability to deliver sensors in the

	Mozambique Channel.
Algorithm	A well-defined procedure that is used to calculate values within a data product.
Browse	A type of metadata that provides a visual representation of a particular data object to facilitate a quick evaluation of usability. There can be many types of browse, and a data object can have multiple browse items.
Catalog	A generalization of containers that represent data resources (datasets or granules). Catalogs provide mechanisms that support user discovery of data resources.
CategorizationScheme	A means of organization that facilitates classification. CategorizationSchemes enable more efficient discovery, and provide insight into understanding the relevance of an item. Two examples of CategorizationSchemes that apply to automobiles might be color and manufacturer. Every instance of an automobile can be classified in both CategorizationSchemes. Note that individual categories in these schemes can change over time.
Category	A particular element of classification within a CategorizationScheme. An example in the AutomobileColor CategorizationScheme might be “Yellow”. An example within the AutmobileManufacture CategorizationScheme might be “Ford”. Items may be classified in multiple Categories within a Categorization Scheme. For example, a two-tone car might be both “Yellow” and “Black”.

Collection	A set of granules that are grouped together due to common attributes. An example of common attributes might be a common research application, a common sensor source, or a common process level.
ComputingResource	Physical computing device where implemented ESDS functionality is deployed. An example of a ComputingResource might be a server or a disk array.
Coverage	A portion of data that contains values representing a set of spatial, temporal or spatiotemporal range.
DataSet	A special type of Collection, where all of the member granules represent the same processing level of the same mission or campaign.
Directory	A searchable container for information about types of data (Collections)
DistributionPolicy	The rules governing the behavior of data access. DistributionPolicy is established by an organization.
Granule	A specific data object. Granules belong to Collections. This is the atomic level of data, as it is represented in a data system or Inventory. Granules may be further de-composed, but those further decompositions are not represented in an Inventory.
Inventory	A searchable container for information about granules.
Job	The directives and parameters that collectively determine a processing stream that, when run successfully, results in data production.
Metadata	Data that describe data. In this information mode, metadata describe Collections and Granules. Metadata is most valuable for

	discovery and understanding of objects.
Prediction	A data granule that provides expected conditions, locations or operations rather than data based on actual observations.
ProductionHistory	A record of a system's operational activities that generated the associated data products.
ProvenenceMetadata	The metadata that describe the full set of input used to generate a specific data granule.
QualityMetadata	The metadata that reflect the viability of a granule for user consumption, application development or subsequent processing.
SearchableMetadata	A subset of metadata that enable discovery of data that users seek for investigation or further processing.
Service	The various mechanisms that consumers employ to access functional elements within a data system. There are four types of Services: WebServiceInterface, WebServiceInstance, UserInteraction and Advertisement.
ServiceRegistry	A searchable container for services.
Statistics	Specific measures generated by ComputingResources which reflect the state and usage of that ComputingResource.
Subscription	A request for an action upon the occurrence of an anticipated future event. For example, a Subscription might send all granules within a predefined collection to a requestor upon publication.
User	An individual (or representation of an individual) that interacts with one or more data systems.

UserInteractionService	UserInteractionServices provide a computer based interface, that enable humans to interact with other services. UserInteractionServices may be dedicated to specific WebServiceInstances, or may enable interaction with any WebServiceInstance that meets a WebServiceInterface. An example of a UserInteractionService is a web site that provides controls for users to manage and interact with an on-line application.
VirtualGranule	VirtualGranules appear to users as though they exist, but do not physically exist as represented. Users can interact with VirtualGranules just as they would with Granules. Systems and services generally create the Granules "on demand", or when requested to by users.
WebServiceInterface	A WebServiceInterface defines the "contract" that a service provider will employ to deliver the functionality of a service. The WebServiceInterface defines the operations, parameters, types, exceptions, as well as some non-functional attributes. An example of a WebServiceInterface is the OGC Web Coverage Service specification (WCS).
WebServiceInstance	A specific location that supports an identified WebServiceInterface. A WebServiceInstance is often the address of a machine that listens, responds to, and/or messages with an IT protocol, in accord with the format defined in a WebServiceInterface. An example of a WebServiceInstance is a

	specific OGC WCS server, which responds to conformant request.

7 System Service View

The System Service View of the ESDS Reference Architecture is intended to identify the key system-level services that are implemented within ESDS's. This representation is based on the Service-Oriented architecture pattern. In this architecture pattern, services are defines as the fundamental points of interface and interaction between components, as well as between components and their users.

Within ESDS's, these are high level services. When implemented each of these services will have a number of operations, and ESDS's may have a number of lower level services. Within the ESDS Reference Architecture, these are identified as the fundamental set of mechanisms required build compliance instances or pieces of ESDS's.

7.1 System Services

This section identifies those system services that are fundamental to NASA's ESDS's. The services are identified and described in Table 7, below.

In the second column of the table, system services are categorized by type. Those types are "User Service" ("U"), or "System Service" ("S"). In this context User Services are those that are offered to the community outside ESDS implementations. These users may be either people or applications. System Services, in this context, are those that occur within the ESDS implementation, and are fundamental to operating those computing systems. Some services are classified as both "user' and "system". In these cases the very same services may be used both internal and external to the ESDS which implements them.

Table 7 - ESDS System Services

System Service	Type	Description
AccessData	U/S	Provides mechanisms for users and applications to retrieve data they identify, moving copies of the data of interest from ESDS resources to their own. These services may provide some data manipulation capabilities, where data is transformed before being delivered to users.

AnalyzeData	U	Provides capabilities for users to study and manipulate data.
AuthorizeUser	U/S	Determines whether or not a user has access to a managed resource, based on system policies.
BroadcastData	U/S	Provides data to users in a real-time manner, via streaming or other technologies.
DiscoverData	U/S	Provides the mechanisms for users and applications to find requisite data resources (both Collections and Granules)
DiscoverService	U/S	Provides the mechanisms for users and applications to find requisite Services.
GenerateReport	S	Provides access to a set of reports (including ad hoc?) that indicate state and trends of the ESDS.
IngestData	S	Provides mechanisms to move data into the ESDS, or move data from one component to another within ESDS instantiations.
InspectData	U	Provides a mechanism to retrieve a data object's metadata, for the purpose of understanding its applicability without actually having to retrieve the entire data object.
InvokeService	U/S	Provides a mechanism to execute a published service on behalf of a user.
Login/Logout	U/S	Initiates or terminates a managed session with an ESDS.
ManageResource	S	Provides mechanisms to control computing resources that support the operation of ESDS components.
DiscoverData	U/S	Provides the mechanisms for users and applications to find requisite data resources (both Collections and Granules)
DiscoverService	U/S	Provides the mechanisms for users and applications to find requisite Services.
GenerateReport	S	Provides access to a set of reports (including ad hoc?) that indicate state and trends of the ESDS.
IngestData	S	Provides mechanisms to move data into the ESDS, or move data from one

		component to another within ESDS instantiations.
InspectData	U	Provides a mechanism to retrieve a data object’s metadata, for the purpose of understanding its applicability without actually having to retrieve the entire data object.
InvokeService	U/S	Provides a mechanism to execute a published service on behalf of a user.
Login/Logout	U/S	Initiates or terminates a managed session with an ESDS.
ManageResource	S	Provides mechanisms to control computing resources that support the operation of ESDS components.
MonitorStatus	S	Provides mechanisms to observe the state of resources within the ESDS or individual components in the ESDS.
ProcessData	S	Provides capabilities to process existing data and generate new data by executing a set of algorithms.
PublishData	S	Makes data resources publicly available.
PublishService	U/S	Makes service resources publicly available.
Subscribe	U	Services that publish events as well as enable users to request notification or a predefined action taken, upon the occurrence of the stipulated event.
TransformData	U/S	Services that operate on data. TransformData services typically convert data from its original state to something more appropriate and/or usable for an end-user’s application. Examples of TransformData services include subsetting, subsampling and reprojection.
VisualizeData	U	Provide mechanisms to display data and interactively manipulate the presentation of that data in various tools.
WorkflowService	U/S	Provides mechanisms to define, compile, configure and execute sequences of services.

7.1.1 Mapping to Reference Requirements

This section offers a mapping of the System Service View to the Use Cases identified in Section 4. In this reference architecture, these Use Cases represent the high level definition of functional requirements of ESDS's. Table 8 demonstrates that each of the system services supports at least one of the ESDS functional requirements.

Table 8 - ESDS Services to Use Cases Mapping

	Receive Sensor Data	Develop Products	Distribute Products	Develop Predictions	Manage Remote Sensor/ Instrument	Data Stewardship
System Service	**ESDS-1**	**ESDS-2**	**ESDS-3**	**ESDS-4**	**ESDS-5**	**ESDS-6**
AccessData		X	X	X		X
AnalyzeData			X	X		
AuthorizeUser	X	X	X	X	X	X
BroadcastData		X	X	X		X
DiscoverData		X	X	X		
DiscoverService		X	X	X		
GenerateReport	X	X	X	X	X	X
IngestData	X	X		X		X
InspectData		X	X	X	X	X
InvokeService	X	X	X	X	X	X
Login/Logout	X	x	X	X	X	X
ManageResource	X	X	X	X	X	X
MonitorStatus	X	X	X	X	X	X
ProcessData		X	X	X		
PublishData	X	X				X
PublishService		X	X	X		X
Subscribe	X	X	X	X		X
TransformData		X	X	X		
VisualizeData			X	X	X	X
WorkflowService	X	X	X	X	X	X

7.1.2 Mapping to Information Architecture

This section offers a mapping of the System Service View to the Information View identified in Section 6. In this mapping, only the key Information Model entities that support the services are identified. Specific instances of these services may support interaction and exchange of other entities. For example, instead of identifying all subtypes, wherever applicable, only generalized base entities are identified.

Table 9 - ESDS Services to Information Model mapping

System Service	Key Information Usage
AccessData	Granule, Collection
AnalyzeData	Granule, Service, Collection
AuthorizeUser	User, Service
BroadcastData	Granule, Collection
DiscoverData	Catalog, Granule, Collection
DiscoverService	ServiceRegistry, Service
GenerateReport	Statistics
IngestData	Catalog, Granule
InspectData	Granule, Collection
InvokeService	Service
Login/Logout	User
ManageResource	ComputingResource
MonitorStatus	ComputingResource
ProcessData	Collection, Granule, Service
PublishData	Catalog, Granule, Collection
PublishService	ServiceRegistry, Service
Subscribe	User, Catalog, ServiceRegistry
TransformData	Granule, service
VisualizeData	Granule, Collection, UserInteractionService
WorkflowService	Service

8 Technology View

The Technology View of the ESDS Reference Architecture provides guidance on the role and selection of technologies and standards. This document does not and should not specify any particular technology or standard. Project-related design and constraints should drive technology decisions. Furthermore, the lifecycle of technologies and standards is not easily congruent with any Reference Architecture.

NASA's ESDS capabilities would not be possible without advances in technology and standards. The pace of emerging technologies and standards will continue to accelerate and provide tremendous value to upcoming NASA ESDS-related programs. Wise decisions with regard to the adoption of appropriate technology are a key challenge for each of these programs. Clearly some programs can tolerate the risk of adopting new technologies at a different level than others.

The NASA Earth Sciences enterprise has a goal to maximize interoperability of the systems and capabilities under its purview. The role of technology and standards are key in achieving this goal, and thus enabling an effective service oriented enterprise. Programs should be highly aware of appropriate technologies and standards from their inception, as they work to participate and contribute to this enterprise goal.

The Earth Science Data Systems Standards Process Group (SPG) focuses on the area of standards for ESDS's. The SPG has a well establish process for identifying and vetting standards. More information on the SPG and the standards the process of evaluating and endorsing standards can be found at http://www.esdswg.org/spg.

The following tables provide lists of various types of technologies and standards that programs need to consider.

Table 10 – Technology Areas for NASA ESDSs

Area	Description
Networking	Protocols that enable connectivity within and among systems
Communication	Approaches and usages of networking resources that enable exchange of information between individuals and systems
Service Description and Invocation	Approaches for describing service capabilities and the interface mechanisms for accessing those capabilities
Computing Language	Source language used to write and compile computer software
Computing Resources	Hardware and operating system foundations for computation, storage and communication
Human Interface	Technology that supports the presentation and human interaction with available data as well as the status of data processing
Collaboration	Technology that supports the sharing of ideas and resources among individuals and communities
Security and User Management	Technology that supports the authentication and authorization of users, as well as the technology that protects computing resources

Table 11 – Standards Areas for NASA ESDSs

Area	Description
Data	Formats and structures of data products
Metadata	Fields that describe data products
Access Services	Services that enable users (software and human) to send data products to designated computing resources.
Discovery Services	Services that enable users to find and understand resources within the ESDS enterprise
Archiving Standards	Approaches that organize and retain data resources for long term storage and accessibility
User Management	Mechanisms that enable tracking and storage of the identity of users as well as enable common exchange of authentication and authorization

9 Glossary

Actor – A UML construct that is used to represent something, or someone, that interacts with a system. Actors are external to a system – they are not part of it. Actors typically stimulate the system to perform some behavior, or are the recipients of a systems work product.

Components – Physical entities (hardware or software) that have one or more specific tasks within a system. Components offer their capabilities, and interact with other components through interfaces (often APIs). Within the ESDS Reference Architecture, components are representations of capabilities that will be realized by actual software or hardware.

Functions – High-level activities that systems support. The components of a system collaborate to provide these high-level functions. Within the ESDS Reference Architecture, functions are those enterprise or program level things that occur within the Earth Science Data System domain. An example of an Earth Science Data System function is Monitor Mission Performance, where many portions of an ESDS will work together to accomplish this function.

Mission Level Use Cases – High-end Use Cases (analogous to Business Use Cases) that describe the responsibilities of projects and the organizations that support them, in the language of its user community and in terms that are independent of technology or solution. Fundamentally, these use cases describe the processes and activities that fulfill mission goals, with mission being defined as an Earth Science Data System and its goals. Mission Level Use Cases are modeled and communicated using techniques from business process engineering. This is in contrast to System Use Cases, which present the capabilities of the system, in terms of interactions between pieces of the system and the stakeholders who use them. Within the ESDS Reference Architecture, Mission Level Use Cases help define the high level requirements that are shared by all instances of the reference architecture. An example of a Mission Use Case within an Earth Science Data System is Develop Products.

Scenarios – Low level interactions between actors and components within a system. Scenarios support the realization of portions of System Use Cases, and are often reusable in many differing use cases. Scenarios represent low level interaction between components of systems, and model the interactions at the message or API level. An example of a Scenario within an Earth Science Data System would be “Login to System”.

Services – Functionality offered by components, provided to other components and actors within a system. Services are accessed via interfaces, which are described through interface definitions. Within the ESDS Reference Architecture, services are standardized in their interfaces, allowing many different component instances to contribute and participate in instantiations of the ESDS Reference Architecture. An example of a service within an ESDS is an OGC Web Mapping Service, which has a well-defined interface and provided by a specific component.

System – An organized composition of Information Technology components (software, hardware, network and data), processes and people that work together to provide capabilities. Users gain access to a system through a well-defined and managed set of interfaces. Within the ESDS Reference Architecture, systems realize some subset of the ESDS Functions, and provide access to that functionality in a manner consistent with the reference architecture.

System Use Cases – Use Cases that describe a system's responsibilities. The responsibilities are captured in the context of how users interact with that system and what those users can expect of the system. System Use Cases describe a series of steps, which are interactions between users and various portions of the system. An example of a System Use Case in an ESDS is, "Discover data granules".

Use Case – A mechanism used to define a set of interactions that lead to a specific goal being realized. Use Cases are captured in the language of the user (called an Actor in UML) and are technology independent. Typically Use Cases are used to capture the functional requirements of a system or organization. There are three levels of Use Cases discussed within the ESDS Reference Architecture: Mission-level Use Cases System Use Cases and Scenarios.

10 Acronyms

API	Application Programming Interface
CEOS	Committee on Earth Observation Systems
COTS	Commercial Off the Shelf
DAAC	Distributed Active Archive Center
EOSDIS	Earth Observation System Distributed Information System
ESDS	Earth Science Data System
L0	Level 0
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NRC	National Research Council
OGC	Open Geospatial Consortium
PB	Petabyte
SPG	Standards Process Group
UML	Unified Modeling Language
WCS	Web Coverage Server