U.S. Geological Survey

Earth Resources Observation and Science (EROS) Center

Architecture Study Team (EAST) Final Report

August 11, 2015

EROS Architecture Study Team (EAST)

Concurrence and Approval

Principally Authored By:

Doug Daniels	Frank Lindsay
EAST Member	EAST Member
The Aerospace Corporation	NASA GSFC

Approved for Submittal to the Director, USGS EROS, by:

Jim Nelson	Ken Klinner
EAST Lead	EAST Co-Lead
Engineering & Development Manager	CITT Manager
USGS EROS	USGS EROS

Table of Contents

Table of Figures

Table of Tables

1. Executive Summary

The Director of the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Dr. Frank P. Kelly, established an EROS Architecture Study Team (EAST) in October 2014 to execute a 9-month study and assessment of the vision for and roadmap to an EROS system and infrastructure architecture that is best positioned to meet the Center's upcoming strategic challenges. The resulting study provided a high-level concept and roadmap for a systems architecture, required infrastructure, and business processes required to meet these strategic objectives and ensure the Center's future information technology (IT) systems operate in an effective manner as possible.

The challenge to the EAST was multifaceted. First and foremost, the challenge was to provide a high-level concept for the systems architecture, infrastructure, and processes required to meet EROS strategic objectives. A key future objective of EROS is to enable Land Change Monitoring, Assessment, and Projection (LCMAP), a capability to provide data that are "analysis-ready" to users and to provide a continuous monitoring capability. Further, the challenges for the EAST included the definition of high-level concepts, considerations, assumptions, risks and benefits, and alternatives for the future EROS architecture and infrastructure. The study objectives included the consideration of new technologies and cost efficient approaches, as well as potential international and private sector partnerships. Finally, the EAST needed to consider refined or enhanced capabilities requested by stakeholders and multiple internal and external user communities as part of its overall assessment.

The architecture study approach was divided into three distinct phases. The first phase encompassed definition of the problem and challenge, as well as characterization of the user communities and use cases. The second phase included identification and definition of the As-Is architecture along with alternative architecture concepts. This phase concluded with selection of a single target architecture concept. Finally, the third phase constituted generation of a roadmap to achieve the future architecture vision. During this study, a parallel and independent comparative cost assessment also was completed.

Initial analysis by the study team readily verified that without a centralized architectural establishment, the architectural direction is determined by individual projects, branches, or funding sources and the architecture is allowed to evolve in decentralized fashion. Furthermore, the information collected from the use-case survey identified a number of overarching issues that highlight limitations of the current architecture framework. These limitations helped form the basis for potential architecture drivers for the EROS near- and long-term architecture evolution. Most users, both internal and external, identified the substantial increase in data volumes as a strain on data delivery mechanisms, storage capacity, and processing capabilities. Acknowledgement that independent, project-specific storage and processing capabilities impede the efficiency of science efforts was a common theme among respondents. Limitations on throughput capabilities, process automation, virtual machine environments, and commercial licensing were all cited as limitations that inhibit effectiveness. Finally, the lack of direct access to analysis-ready data (ARD) was identified as a substantial limiting factor in the timeliness and spatial scope to which science research and operations are able to function.

Using the As-Is architecture assessment as a baseline, the EAST identified three additional candidate architectures:

1) **As-Is architecture:** represents the "do nothing" decision to continue with the current highly decentralized and independent architecture evolution;

2) **Projectized Matrix:** emphasizes minor changes from the As-Is architecture through limited centralized governance or system of systems guidance offered to projects;
3) **Enterprise:** stipulates an effective centralized governance and system of systems approach be established and used to manage and oversee the Center's architecture evolution; and
4) **Cloud-Centric:** provisions almost all Center compute and storage resource capabilities off-site to private or public cloud providers.

At the conclusion of Phase 2, the study team recommended pursuit of the Enterprise Architecture alternative for Phase 3 assessment and roadmap development. This architecture option ranked the highest among the measures of success, was determined to be the lowest risk alternative, and best met the EAST challenge statement.

During Phase 3, the team developed an incremental roadmap to transition from the As-Is state to an enterprise architecture, which will require many activities to be run in parallel within each area of the architecture. Implementation planners will need to carefully consider the complementary nature of the approach along with dependencies to ensure all needed pieces are evolving at the correct pace for enterprise business, technology, applications, and information lifecycles to be used in an effective manner. An implementation plan outlining the approach, roles and responsibilities, and methods for achieving the enterprise architecture will need to be generated as a follow-on activity to further develop the details of the roadmap.

The study team concluded its activities by making three primary observations concerning the architecture study activities:

1. **EROS is currently effective as a collection of semi-independent projects, but inefficient at a center level.** In many cases projects are efficient within their project boundaries and in some cases effectively share experiences and capabilities with other projects. However, EROS could be more efficient from an overall resource sharing, common services, and business objectives perspective.

2. **The EROS Enterprise Architecture approach could provide many benefits.** A Center-level system of systems view of EROS architecture enables effective strategic planning and a centralized security responsibility and capability improves the overall security posture. Additionally, virtualization and other enterprise services help accommodate various types of projects, from small science projects to large projects like Long-Term Archive (LTA) and Landsat, to new project endeavors like LCMAP.

3. **Transitioning from As-Is to target architecture should be an evolution.** Business model, information model, and infrastructure changes are needed early on to accommodate eventual transition to an enterprise architecture. The business model must be agile, efficient, and cost effective for projects and the Center to realize cost benefits. Finally, regular updates to the architecture roadmap and implementation plan will be important to a successful transition.

The following key recommendations are offered for implementation planning and execution:

1) Develop an enterprise architecture implementation plan based on the provided roadmap;
2) Fully develop and implement a Center-wide agile and nimble business model;
3) Transition to the target architecture in an evolutionary versus revolutionary way; and
4) Build on interagency partnerships established during EAST implementation.

2. Introduction

The following document describes the activities, processes and outcomes of the Earth Resources Observation and Science (EROS) Architecture Study Team (EAST). This introduction provides the context and necessary discussion materials that underpin the study and the recommendations offered. The discussion herein is intended to link the mission of EROS, the goals for the architecture study, and the subsequent sections that characterize the facets of the EAST's approach. For the purposes of this document, use of the term "current" refers to the year 2015.

2.1. Document Organization

The EAST Final Report is divided into 11 sections. The first three sections are introductory and approach in nature and are intended to provide insight and context for the reader to discern the intent of and underlying organizational process for the EROS architecture study and resulting study artifacts. Section 4 addresses the science overview and science drivers including the identification of land science user communities and characteristics, along with likely requirements for data. Section 5 summarizes needed study inputs to ensure the architecture study activity is fully informed. Section 6 describes the architecture viewpoints and standardized approach to architecture definition and standards the EAST applied. Section 7 describes and highlights the existing architecture as the basis for examining future technology pathways.

Section 8 provides several pertinent models and concepts that aid in the definition of future architecture alternatives. Additionally, section 9 describes the process and result of selecting a single concept for further study. Section 9 addresses the 2021 vision for the selected architecture, whereas section 10 includes a roadmap. Section 11 contains the study team's final observations and recommendations for follow-on study and implementation activities. Finally, appendix 1 summarizes an independent cost study, and appendix 2 captures the list of EAST deliverables.

2.2. Reference Materials

The following documents and reports provide the technical and organizational context for the EROS Architecture Study.

1) Federal Enterprise Architecture (FEA), Office of Management and Budget (OMB), (http://www.whitehouse.gov/omb/egov/a-1-fea.html)
2) IEEE Std 1362-1998 "IEEE Guide for Information Technology—System Definition—Concept of Operations (ConOps) Document", (http://ieeexplore.ieee.org/servlet/opac?punumber=6166)
3) ISO/IEC/IEEE 29148 "Systems and software engineering. Life cycle processes. Requirements engineering", (http://ieeexplore.ieee.org/servlet/opac?punumber=6146377)
4) ANSI/AIAA G-043A-2012 "Guide to the Preparation of Operational Concept Documents", (https://www.aiaa.org/StandardsDetail.aspx?id=12878)
5) Information Technology Infrastructure Library (ITIL), (https://www.axelos.com/best-practice-solutions/itil)
6) The Open Group Architecture Framework (TOGAF) Version 8.1.1, (https://www.opengroup.org/togaf/index811.htm)

7) Control Objectives for Information Related Technology (COBIT) – IT Governance Framework, (http://www.isaca.org/COBIT/pages/default.aspx)
8) NPR 7123.1B, "NASA Systems Engineering Processes and Requirements", (http://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7123&s=1B)
9) Burkett, V.R., Kirtland, D.A., Taylor, I.L., Belnap, Jayne, Cronin, T.M., Dettinger, M.D., Frazier, E.L., Haines, J.W., Loveland, T.R., Milly, P.C.D., O'Malley, Robin, Thompson, R.S., Maule, A.G., McMahon, Gerard, and Striegl, R.G., 2013, U.S. Geological Survey climate and land use change science strategy—A framework for understanding and responding to global change: U.S. Geological Survey Circular 1383–A, p. 43. , (http://pubs.er.usgs.gov/publication/cir1383A)
10) NOAA Big Data Request for Information (RFI): (https://www.fbo.gov/index?id=d0a8d9a279c69aeac9108507ff32f22a)
11) NOAA Big Data Cooperative Research and Development Agreement (CRADA) (https://data-alliance.noaa.gov/)

2.3. EAST Charter

The following subsections represent material provided to or discussed with the EAST by the study sponsor, Dr. Frank Kelly, and the EAST steering committee.

2.3.1. Background

The EROS Center (hereinafter the Center) has for more than four decades fostered an Information Technology (IT) environment that is heavily dependent on individual project business cases, objectives, and requirements. With the exception of the Center's IT security posture, governance of EROS IT systems is largely decentralized and has been so since the mid-1990s. As such, the technology, applications, and data management environments within and across projects have evolved in an undetermined manner. In other words, EROS maintains a number of different IT processes and approaches for accomplishing its work and understanding the evolution of these systems from a Center perspective has proven problematic.

Additionally, new development projects such as Landsat 9 and the Land Change Monitoring, Assessment, and Projection (LCMAP) initiative will establish the new cornerstone of the Center's strategic vision. To meet these project challenges while continuing to be successful across the Center's broad spectrum of science activities, an evaluation and assessment of the Center's future IT systems is paramount, especially as the Center evolves from being not only a good steward of land change data, but evolves toward becoming a champion for land change information.

Therefore, the Director of the EROS Center established the EAST to execute a 9-month study and assessment of the vision for and the road map to an EROS system and infrastructure architecture that is best positioned to meet the Center's strategic challenges. The study provides a high-level concept and roadmap for a systems architecture, required infrastructure, and business processes required to meet these strategic objectives and ensure the Center's future IT systems operate in as an efficient manner as possible.

2.3.2. EROS Mission

The EROS overall mission is multifaceted, supporting broad and diverse U.S.-based land science communities and seeking to lead in the understanding of how changes in land use, land cover, and condition affect people and nature. Stated simply, the mission of EROS is to contribute to the understanding of a changing Earth by providing services that monitor relevant land change information and knowledge, assess the trends and consequences of land change, and provide pertinent additional services and support on the use and understanding of land change monitoring products and information. To support these goals, EROS seeks to position itself as the world's primary source of remotely sensed land images of the Earth, being an authoritative provider of land change science information and knowledge.

2.3.3. Study Purpose and Objectives

Based on these overall goals for EROS, the EAST charter stated: "…define and assess candidate architectures that support current needs and allow for the expansion of the EROS mission to include providing land change data, information, and knowledge products, along with a path for evolution from current capabilities." To achieve this overall goal, several simultaneous tasks were undertaken to inform key aspects of the architecture study and subsequent architecture design decisions.

Some tasks focused on the existing infrastructure and organizations to enhance and optimize the EROS As-Is architecture and to identify and streamline opportunities for shared services across project activities. A second set of challenges focused on the need to evolve and support future science activities at EROS. These future-leaning activities included the ability to prepare for next-generation land imaging missions, addressing the capabilities for ready access to EROS data holdings and computing capacity to generate information on land changes as they are detected (that is, LCMAP), and facilitating the evolution of systems and data analytics services needed to enable scientific findings derived from data and modeling.

2.3.4. Sponsors, Steering Committee, and Team Members

The team members who performed this study represented the pertinent USGS, EROS, and partnering U.S. stakeholder agencies. The study sponsor, Dr. Frank Kelly, established a tripartite committee structure to facilitate the communication and guidance across the groups and provide a means for coupling the needs of the study to the responsibilities and expertise of study members. Tables 2-1, 2-2, and 2-3 provide a personnel listing for each of these three groups.

Table 2-1. Stakeholder members for the EAST.

Stakeholder member	Affiliation	Role
Frank Kelly	USGS	Sponsor, EROS Director
Tim Newman	USGS	Land Remote Sensing Program Coordinator
John Hahn	USGS	EROS Deputy Director
Tom Loveland	USGS	EROS Chief Scientist
Dave Jarrett/Steve Neeck	NASA	NASA Earth Science Division Program Management

The key study stakeholders included in Table 2-1 above consist of EROS and USGS senior management and representatives from the National Aeronautics and Space Administration (NASA) Earth Science

Division (ESD). It is from this committee that the study goals, structure, and timeline were produced and the charge to the supporting committees was drawn. Moreover, this group is the recipient of this final report. The steering committee (table 2-2) included the Center's Branch Chiefs and other senior leads. This committee provided vital guidance and feedback to the study team on its activities between formal checkpoint reviews and solicited information from their organizations pertinent to the overall EAST needs.

Table 2-2. Steering committee members for the EAST.

Steering committee member	Affiliation	Role
Tom Kalvelage	USGS	Steering Committee lead, Coordination and Requirements Office Chief
Jenn Lacey	USGS	Observing Systems Branch Chief
Doug Binnie	USGS	Data Services Branch Chief
Dave Hair	USGS	Science Applications Branch Chief
Kim Allington	USGS	Administrative Systems Branch Chief
Steve Covington	Aerospace	Land Remote Sensing Program Representative

The architecture study team shown in table 2-3 consisted of the key technical leads from EROS, supporting experts from Aerospace Corporation, representatives from four separate NASA centers, and National Oceanic and Atmospheric Administration (NOAA) NOAA National Environmental Satellite, Data, and Information Service (NESDIS). This team was co-led by the Engineering and Development Manager, Jim Nelson, and the EROS Center IT Team (CITT) Manager, Ken Klinner. The study team members carried out the tasks and activities called for in the EAST Charter. This group also was responsible for producing all the needed materials and analyses necessary to arrive at findings to help USGS decision makers prepare for the evolution of EROS science data system infrastructure and management.

Table 2-3. Study Team members for the EAST.

EAST Team Member	Affiliation	Role
Jim Nelson	USGS	Study Lead
Ken Klinner	USGS	Study Co-Lead, EROS IT infrastructure
Doug Daniels	Aerospace	Systems Engineering
Mike Budde	USGS	User Needs, EROS Science and Applications
Chris Rusanowski/Chris Torbert	USGS	Data access, archive, and distribution
Chris Engebretson	USGS	Science data processing
John Moses/Frank Lindsay	NASA Goddard Space Flight Center	Science Data Processing, Earth Observing System Data and Information System
Del Jenstrom/Jeff Masek	NASA Goddard Space Flight Center	NASA Sustainable Land Imaging
Dave Alfano/Petr Votava	NASA Ames Research Center	NASA Earth Exchange, NASA Advanced Supercomputer

Rich Doyle/Dan Crichton	NASA Jet Propulsion Laboratory	Big Data, Distributed Data Architectures
Michelle Detomasso/James Holton	NOAA	Science data processing and archive
Tom Sohre	USGS	Management, business models
Mary Covert	Aerospace	Comparative cost analyst

In addition to the members of the EAST, other key contributors included Randy Sunne and Dan Akkerman, both of whom provided enterprise systems engineering support from the EROS Technical Support Services Contractor (TSSC), Stinger Ghaffarian Technologies (SGT).

2.3.5. Measures of Success

The EAST measures of success are criteria by which architecture alternatives were measured, and later ranked and scored using a decision matrix. The Charter for the EAST defined the first four measures of success. The fifth measure, Security, was added by the steering committee during the study:

- **Effectiveness:** The recommended architecture should be capable of sufficient performance in all areas to meet EROS and stakeholder strategic objectives.

- **Flexibility:** The recommended architecture should be scalable, to meet current and future requirements; flexible, to meet a broad variety and scale of EROS requirements; and agile, to be able to provide solutions across EROS with minimum tailoring and re-architecture.

- **Sustainability:** The recommended architecture should provide the solution for the long haul without extraordinary infusions of funds, in a cost-efficient manner as technology, policies, and vendors change.

- **Reliability:** The recommended architecture should be robust, not susceptible to single-point failures, and enable EROS to effectively manage risk.

- **Security:** The recommended architecture should limit potential vulnerabilities and allow EROS to effectively manage necessary changes within a security posture.

These measures were used throughout the study and are referred to frequently in the remainder of this report.

3. Study Methodology

The following sections describe the overall challenge, scope, and methodology used by the study team to address the challenge and objectives put forth by the sponsor and the steering committee.

3.1. Challenge Statement

As discussed in the introduction, the challenge to the EAST was multi-faceted. At a high level, the challenge statement developed by the study team and accepted by the sponsor and steering committee was to define and assess candidate architectures that support current needs and allow for the expansion of the EROS mission to include providing land change data, information, and knowledge products, along with a path for evolution from current capabilities. The challenge was further refined to include the following:

- Enhance and optimize the EROS As-Is architecture;
- Identify and streamline opportunities for shared services across project activities;
- Prepare for next generation land imaging and like missions;
- Address capability for ready access to EROS data holdings and computing capacity to generate information on land changes as they are detected (that is, LCMAP);
- Address evolution of systems and data analytics services needed to enable science from data and modeling; and
- Explore applicability of external public and private partnerships.

In considering these aspects, the EAST also considered, as part of its overall assessment, refined or enhanced capabilities requested by stakeholders and multiple user communities.

3.2. Study Scope

A crucial element in the development and implementation of the EROS Architecture Study was to clearly identify the scope of the study, the various system elements and their relationship to the overall study goals, and acknowledgement of the likely areas where trade decisions will be made. The somewhat short duration (9 months) of the study precluded the addition of several related and pertinent topics. Nonetheless, the scope for the EROS Architecture Study included all the substantial dimensions so that informed and vetted recommendations could be offered for evolving the EROS data system architecture.

Items considered within the study scope were the established central components for the study including the extensive systems and services associated with science data EROS networks, science project systems and services, and emergency operations. The current As-Is system architectural elements served as an essential basis for the analysis. The team performed an analysis of current and future land data holdings, data products (image data processing), product information (information assurance), and knowledge storage and distribution characteristics. Also included in the study scope were end user requirements, user discovery, access, visualization of holdings and partnerships such as the Land Processes Distributed Active Archive Center (LP DAAC).

The following topics were outside the study's scope: EROS Center policies, EROS finance and administration, physical security, end user desktop services (that is, the help desk), EROS communications and public outreach activities, science actions (for example., methodologies, algorithms, and so on), in-situ and field work, and Landsat flight operations and ground stations (including other antennas).

3.3. Study Methodology and Approach

The overall methodology used by the EAST to address the challenge and scope of the study is shown in figure 3–1. Each of the activities in the figure is color coded by phase, as defined in figure 3–1, and ultimately supported the definition of target architecture concepts along with a roadmap to the final architecture representation.

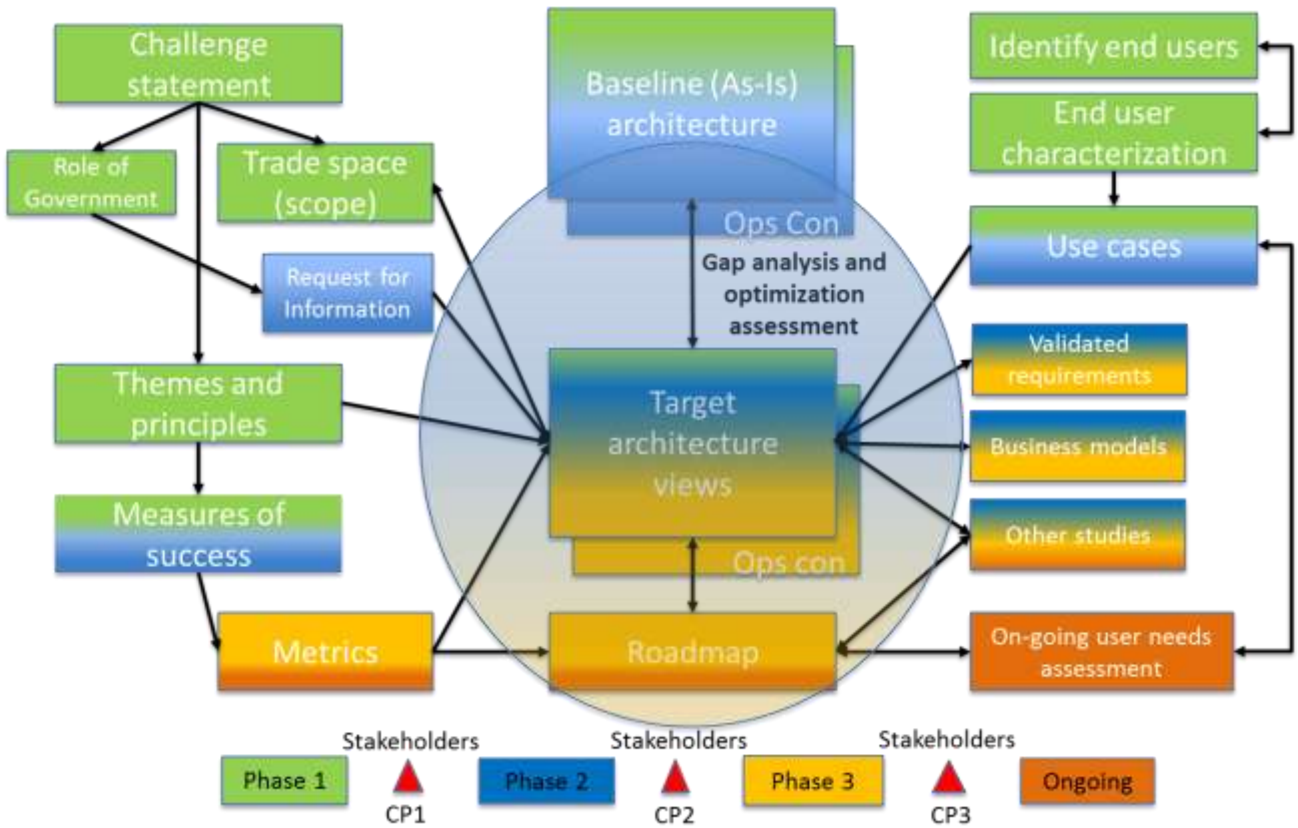


Figure 3–1. Architecture Development Process Methodology.

The architecture study approach was divided into three distinct phases, each completed with a checkpoint review held with the sponsor, steering committee, and stakeholders (fig. 3–2).

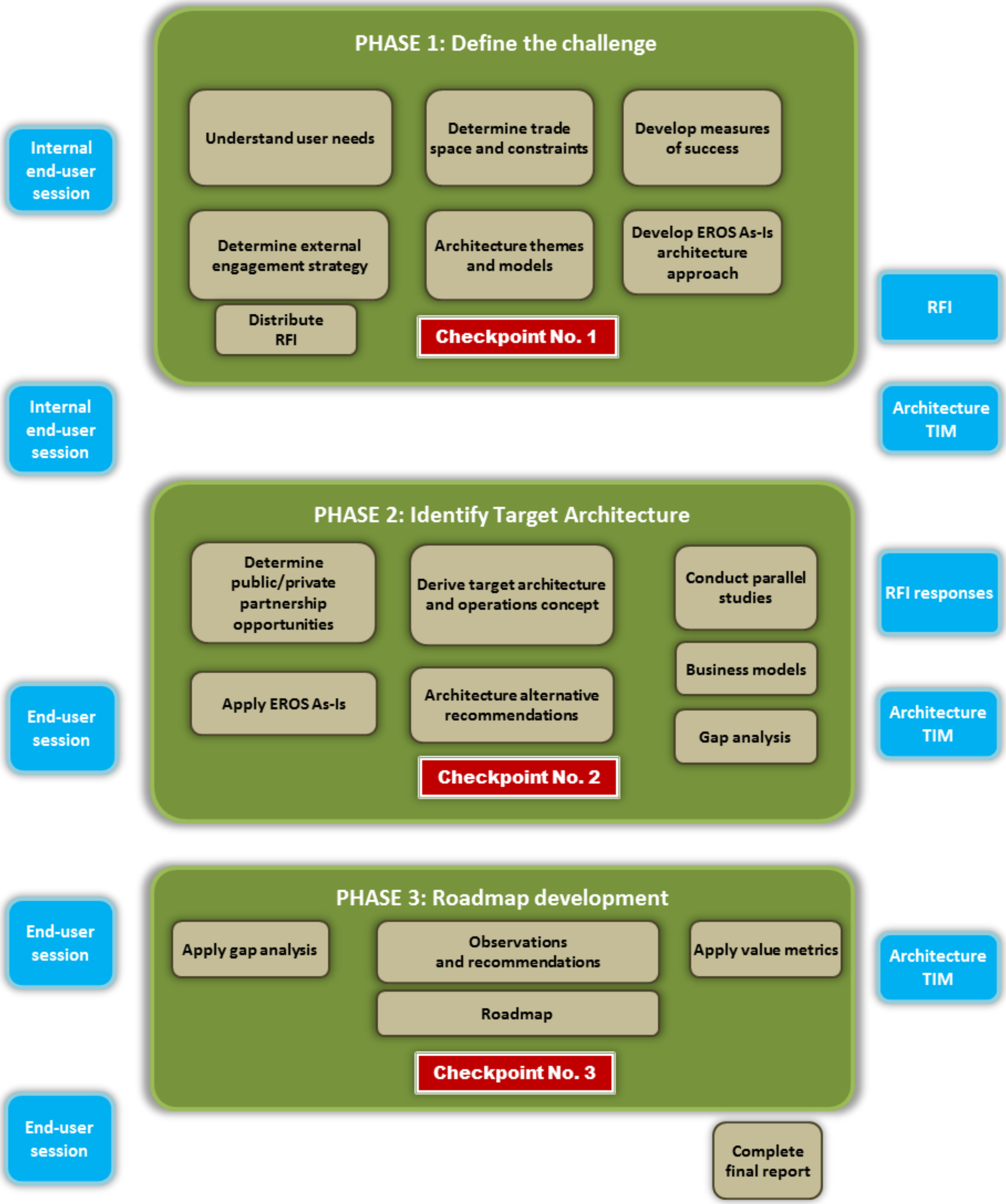


Figure 3–2. Phased EAST Study Approach.

Phase 1 spanned October 2014 through January 2015. The focus of Phase 1 was geared entirely toward clearly defining the problem (challenge). During this phase, the team set out to define needed inputs, develop future needs, define the scope and lines of business, and establish an approach to identifying the As-Is architecture baseline. In support of these phased objectives, the team pursued multiple efforts in parallel including the following:

- Phase 1 started with the first of multiple end user sessions established to determine user needs including user community definition and characterizations, identification of representative use cases, and completion of use case solicitations;
- Determination of trade space (scope) and any constraints;
- Determination of external and internal to EROS engagement strategies;
- Development and refinement of success measures and associated metrics;
- Definition of architecture themes and framework; and
- Development of EROS As-Is architecture identification approach.

During Phase 1, the team also generated and distributed a Request for Information (RFI) to enable Phase 2 assessments of potential public and private partnerships.

Phase 1, completed with Checkpoint no. 1, was held on January 30, 2015. The team also held a technical interchange meeting with the broader EROS stakeholder community in February to convey the team's Phase 2 study objectives and to encourage broad Center participation.

Phase 2 spanned February 2015 through April 2015. The focus of Phase 2 was to identify target architecture concepts along with initial business and governance model constructs. As such, this phase comprised the bulk of the EAST efforts during the duration of this study.

In support of these objectives, the team pursued multiple efforts including the following:

- Finalized EAST architecture framework process and terminology.
- Refined and finalized user characterization and applied use cases.
    - Completed use case evaluation and gap analysis.
    - Surveyed additional projects/applications that addressed gaps in community assessments.
- Finalized EAST measures of success and formulated associated metrics.
- Completed discussions and gathering of partner experiences (for example, lessons learned).
- Gathered information on state of technology.
    - Assessed RFI respondent recommendations for potential private and public partnerships.
    - Assessed technologies for potential architecture application.
- Finalized EROS As-Is architecture applications view and work flow (for example, operations concept).
    - Formulated As-Is architecture observations and assessed against EAST measures of success.
- Iterated and revised target systems architecture scope and views.
- Responded to EAST challenge statement by identifying three additional architecture alternatives.
    - Conducted assessment of architecture alternatives against EAST measures of success and recommended EROS target architecture.
- Developed initial business model and governance constructs.
    - Initiated parallel study for comparative cost analysis in support of understanding historical As-Is costs along with potential public and private partnership costs.

Phase 2 completed with checkpoint no. 2, which was held on May 8, 2015. Before the checkpoint completion, the team also held technical interchange meetings with the broader EROS stakeholder community in February and April to convey the team's Phase 2 study observations, and to encourage broader community feedback on the noted architecture alternatives.

Phase 3 spanned May 2015 through mid-July 2015. The focus of Phase 3 included the following:

- Developed and revised the business model
- Established the framework for the information model and plan forward
- Revised and enhanced the technology architecture and transition states
- Completed the architecture roadmap to 2021
- Completed the parallel cost comparison assessment study.

Phase 3 completed with checkpoint no. 3, which was held on July 10, 2015. Following this checkpoint, all planned architecture study activities were completed.

3.4. Study Timeline

The timeline allotted for this study was, in part, driven by the need for timely analysis and recommendations of the current and possible future state of the EROS science data architecture. Following the approval of the Study Charter, the multiphased EAST study began in early October 2014 with a targeted completion of July 15, 2015, and final report within a month of the study completion. Study activities and interactions were considerably later than this period as illustrated in figure 3–3.

The activities and meetings shown in figure 3–3 reflect the parallel approach the team took in gaining access to information and scheduling numerous face-to-face meetings and user sessions on a regular basis. The study team met in person eight times attempting to provide at least one site visit to the participating NASA centers and EROS. These meetings were significant in assessing where the team was in the process of the architecture evaluation and allowed for course corrections on topics as needed. In addition, the meetings provided a means for greater interaction and involvement of those not participating directly on the EAST. Outcomes from these interactions are addressed more fully in the section 5 Study Input Summary.

Other activities included a host of interactions with key players in the development of the evolving data system, including three science and systems user engagement sessions where the needs and requirements of current and future users were addressed. This included one dialogue session during the Landsat Science Team meeting at NASA's Goddard Space Flight Center (GSFC). The study team also met for three Technical Interchange Meetings (TIMs) that generally coincided with the checkpoint reviews. These meetings allowed the team in delve deeply into the technical issues to gain consensus and insight on later recommendations. An industry RFI also was released early during the study so that the team could capture insights and perspectives from the private and commercial IT business community.

The final task for the study team was to provide high-level system architecture, infrastructure, and process recommendations to the EROS Director by late July 2015, along with a roadmap to achieve the vision and implementation for future architecture.

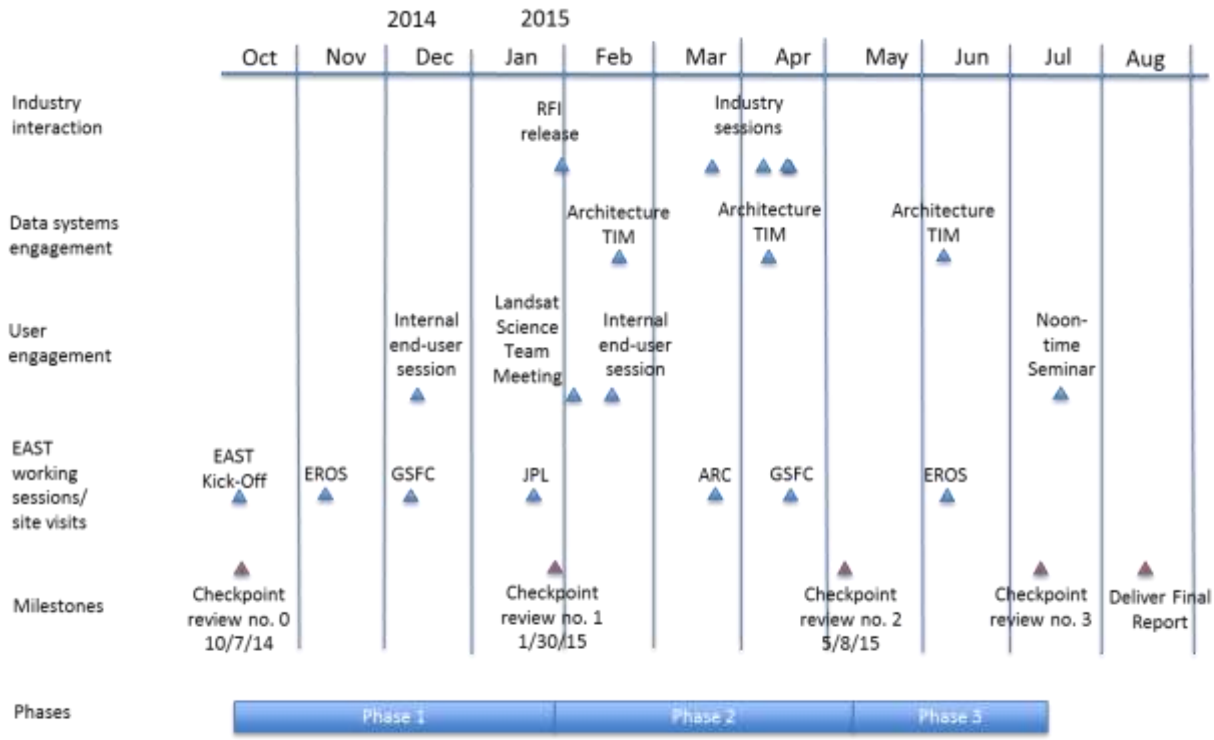


Figure 3–3. EAST timeline.

4. Science Overview and Approach

This science overview represents a description of the methodology used to characterize EROS end user communities and apply information gathered from those users as architecture drivers for the EAST.

Typically, remote-sensing data users have been classified by their affiliation (academia, private industry, federal/state/local government, and so on) or by their primary data use (land use/land cover, agriculture, climate change, fire science, hazards, and so on) and rarely organized by aspects of their data use. The EAST strategy for assessing user needs and requirements considered the typical categories of user classification, but also paid special attention to current and future uses in terms of data volume, types of science data or information products used, and access/distribution requirements.

There are clear distinctions between those users who operate in a bulk data use environment and those who acquire small to moderate amounts of data for specific one-time research applications. Furthermore, there are users who use Earth observation (EO) data in operational modes for real-time or consistent processing through time. Because each of these user groups can span a variety of user affiliations or applications, it was important, while assessing user requirements, to account for different modes of data/product delivery and services.

4.1. End User Characterizations

The approach used in developing our end user characterization included extensive engagement of internal and external science activities. An initial assessment of user groups, based on the criteria identified above, was developed and vetted by a group of EROS scientists. Those members of the EROS in-house science community chosen to review the initial user groups represented a broad spectrum of research and applications projects at the Center, but also had an array of external partner relationships to build on. In addition to the internal science support, the team also received substantial assistance from Dr. Jeffrey Masek, Chief of the Biospheric Sciences Laboratory at the NASA GSFC. The combined efforts of this experienced science community produced a draft list of the data user communities, attributes of their data use, and specific projects or applications that were representative of given user groups. A user survey was developed and distributed to the representative use cases asking for information about data use; current data flows, and suggested workflow improvements. The EAST achieved an approximate 75-percent return on the survey, which provided valuable insight into current use case limitations.

As part of the first technical interchange meeting, held in February 2015, the draft list of user communities, attributes, and use cases was evaluated. Whereas there were no clear objections to what had been assembled, broad representation by the science community was limited. Following checkpoint no. 1, and after consultation with EROS's Chief Scientist, a second effort was made to specifically target the science community and involved participation from approximately 20 members of the EROS Science Research and Applications Branch. The result of the additional session was concurrence that use cases sufficiently represented the end user communities and that the communities themselves are representative of the broad spectrum of EROS archive data users. Also, during Phase 2, an additional use case pertaining to the U.S. Forest Service (USFS) Landscape Change Monitoring System (LCMS) was identified. Warren Cohen and Sean Healey of the USFS submitted a user survey on behalf of the project.

Table 4–1 depicts the EROS user communities identified and organized into seven categories based on data use, access, and derived products unique to each group.

Table 4–1. User Community Characterization.

Data user community	Attributes
1) Large volume science users	High volume, bulk data user Large data storage requirements Broad geographic scope and product range
2) Operational users	High temporal frequency requirements Routine access to data/products Consistently processed data streams
3) Near real-time applications	Rapid access to data is essential Relatively small volumes of data Targeted geographic areas
4) Focus studies	Local to regional investigations Highly diverse product suites desired High in numbers, low data volume
5)Technique developers	Heavily academic in nature Large group, small data volumes Hand off to operations
6) Data providers, commercial enterprise	High data volume, bulk data user Broad geographic scope and product range Small GIS-services companies, agribusiness

7) Derived product users, formal educators, communicators, general public	Includes some GIS analysts Less data and more products/information Little or no remote sensing/image processing Small volumes Probably best served with seamless JPGs

4.2. Applied Use Case Summary

Table 4–2 represents the use cases identified by the science community as representative use cases. Those cases denoted by an asterisk (*) represent the use cases collected and assessed by the EAST. The source of each use case is noted in parentheses after the use case name in the table below.

Table 4–2. Use Case Summary.

User community	Use case
1) Large volume science users	Global forest gains/losses (Hansen/Loveland) NEX Web-Enabled Landsat Data (Votava)*
2) Operational users	USDA National Ag Statistics Service (Mueller)* USDA Foreign Ag. Service (Reynolds)* USGS FEWS NET – ETa modeling (Senay)*
3) Near real-time applications	Monitoring Trends in Burn Severity (Howard)* Emergency Response/Int'l. Charter (Jones)*
4) Focus studies	National Shrub and Grass Fuel Mapping (JV)* Ecology/Vegetation (French)*
5)Technique developers	Landsat albedo algorithm (Schaff)* Landsat ET & STAR-FM algorithm (Gao)*
6) Data providers, commercial enterprise	CEOS-GFOI (Fosnight)* Google/Amazon/ESRI
7) Derived product users, formal educators, communicators, general public	Landsat Look / Data Democracy Initiative* IGETT Program (Allen)

The information collected from the use case survey identified a number of overarching issues that highlight limitations of the current architecture framework. These limitations help form the basis for potential architecture drivers for the EROS architecture evolution. Most users identified the substantial increase in data volumes because of additional sensors with larger file sizes as a strain on data delivery mechanisms, storage capacity, and processing capabilities. Acknowledgement that decentralized storage and processing capabilities inhibit the efficiency of science efforts was a common theme among respondents. Limitations on throughput capabilities, process automation, virtual machine environments, and commercial licensing were all cited as issues that deter effectiveness. The lack of access to analysis-ready data was identified as a substantial limiting factor in the timeliness and spatial scope to which science research and operations are able to function. Projects commonly need similar data and processing resources, but operate in stove-piped environments that impede resource sharing.

A number of architecture requirements were derived based on the limitations identified above. Initiatives such as the USGS Land Change Monitoring, Assessment, and Projection (LCMAP) and the U.S. Forest Service (USFS) Landscape Change Monitoring System (LCMS) will drive the need for production of and access to analysis-ready data. Whereas these efforts are primarily Landsat based, there is clearly a need for incorporating new and existing mission data including, but not limited to, Sentinel–2, Moderate Resolution Imaging Spectroradiometer (MODIS), and Visible Infrared Imaging Radiometer Suite (VIIRS) data. The ever increasing data volumes of new sensors and the opportunities of evolving change detection and assessment methodologies will require the provision of centralized storage capacity and science computing capabilities to efficiently use Center resources. Because not all science projects function in an operational mode, there will also be a need for surge or burst capabilities related to data storage and processing. More efficient use of virtual machine processing along with virtual server and desktop processing is also desired. In addition, there is a need to establish a consistent and predictable processing environment, which includes controlled evolution of applied technology and practices. The architecture solutions recommended by the EAST should enable science projects to focus more on the work they do and less on how to obtain their needed resources.

It would be naïve to expect that the science community will readily embrace all of the recommendations put forth; thus resisters could create potential hurdles in the adoption of a modified architecture. The EAST acknowledges the culture that exists within the Center, whereby individual projects have been "going it alone", and in many cases have developed substantial capabilities that suit their individual needs. The principal goal of the EAST must be to embrace those project-level functions that are working well and not disrupt those proficiencies, while at the same time providing incentives for science-process improvement. Willingness of the science community to actively support and participate in proposed changes is certainly attainable if benefits of doing so are clearly communicated and the procedure is not perceived as being imposed in a top-down manner.

The EAST attempted to gather a broad representative set of science requirements that could be addressed through the implementation of a new Center architecture. There will be an on-going need to engage the internal and external user communities to assure that needs are being met and to incorporate new or unidentified requirements. Along with regular direct engagement, one strategy for accomplishing this goal beyond the EAST would be to leverage the efforts of the USGS Land Remote Sensing Program's Requirements, Capabilities and Analysis for Earth Observation (RCA-EO) group. Through sustained evaluation of user needs, EROS can ensure that the data and information products it provides are being used to their full potential.

5. Study Input Summary

The EAST made use of multiple avenues and activities to capture needed input for the study, including a series of Technical Interchange Meetings (TIMs), discussions with partner members on their applicable experiences, investigations of potential public and private partnerships, and an RFI designed to explore additional technologies and approaches. The following sections address the key input information captured from each of these activities.

5.1. Internal EROS Technical Interchange Meetings

The EAST held two internal EROS TIMs in order to properly engage the broader EROS staff in the work of the EAST. Each TIM had a set of specific goals to be achieved during the discussion.

5.1.1. February Technical Interchange Meeting

The EAST facilitated a TIM with the broader EROS community in February 2015, with a primary intention to focus on the As-Is architecture environment and performance. The goals of the TIM were threefold:

1) Re-engage members of the science community at EROS to assess representative use cases, determine if any gaps exist, and discuss future needs.
    - As a result, the USFS LCMS was identified as an additional needed use case.
    - The user categories and characterizations were confirmed as accurate as the previously developed use cases.
2) Engage EROS As-Is systems curators to explore current and future architecture activities and assess potential CITT architecture and service interfaces with the EAST.
    - CITT "as a service" activities were defined and discussed. Strategic initiatives for CITT spanning current year through fiscal year (FY) 2020 also were discussed and incorporated into the EAST study for planning purposes.
    - Other key project architecture activities were reviewed and discussed, including LCMAP.
3) Engage EROS systems engineers to derive data and operations concept flows along with decomposition of the high-level applications architecture.
    - Key applications pertaining to the EROS mission were identified and folded into the team's applications architecture view.

Perhaps most importantly, this first TIM enabled the broader EROS internal community an opportunity to engage the EAST to better understand the team's objectives and directly support the team's efforts.

5.1.2. April Technical Interchange Meeting

The EAST facilitated an additional TIM with the broader EROS community in April 2015, just before checkpoint no. 2. This technical interchange proved crucial to supporting the conclusion of the EAST's Phase 2 objectives. Specifically, the results of this TIM contributed directly to the definition and preliminary assessments of the target architecture alternatives and potential business models. The primary emphasis of the TIM was to discuss attributes of the alternative architectures. The goals of the TIM were threefold:

1) Discuss LCMAP and the Analysis-Ready Data (ARD) definition including how it pertains to EAST.
2) Review and discuss the "To-be" architecture scope elements based on the architecture framework view.
3) Discuss and iterate "To-be" architecture alternatives including the following:
    - Specific architecture observations,
    - Opportunities for resource sharing, and
    - Governance needs for architecture alternatives.

Of primary interest, TIM participants noted multiple deficiencies because of the lack of centralized and standardized governance related to technology and resource deployments at EROS. In other words, there was a strong bias for moving in the direction of managing EROS resources as an enterprise (that is, managing combinations of hardware, software, and infrastructure as a whole system).

5.2. Key Experiences and Inputs from EAST Partner Organizations

As a part of the EAST phased activities, a series of site visits to participating NASA centers were successfully completed. These meetings allowed the members of the EAST to have real-time discussions with NASA subject matter experts in a range of pertinent information system and technology topics. The NASA-affiliated EAST members identified analogous information systems to be explored during the visit and helped to drive questions about technical trades and lessons learned from these NASA Centers (that is, GSFC, JPL, and ARC). The central messages and experiences from the Centers are captured in the following sections. In addition, interaction with NOAA NESDIS staff added to the collection of partner experiences based on a NOAA architecture study and implementation exercise.

5.2.1. Goddard Space Flight Center (GSFC)

NASA's multimission science data Earth Observing System Data and Information System (EOSDIS) are managed by the Earth Science Data and Information System (ESDIS) Project based out of GSFC but involving teams across several NASA centers. This complex and diverse system of systems has supported many Earth-observing satellite missions beginning in the late 1990s and has undergone multiple periods of change since its inception driven by new technologies, science requirements, and the sheer growth of data volume and types hosted. EOSDIS generates Level 1–4 science data products for Earth Observing System (EOS) missions; archives and distributes data products from EOS and other satellite missions, as well as from aircraft and field measurement campaigns. The architecture team's visit to GSFC was intended to gain understanding of NASA's lessons learned from the planning, development, and evolution of EOSDIS.

An early EOSDIS insight offered to EAST was to manage data system elements at the interfaces. Many diverse stakeholders of a system-of-systems that needed to be loosely coupled became easily managed by controlling interfaces and allowing discipline-specific leadership to develop system architectures tailored to be efficient for their application. The ESDIS Project manages EOSDIS interfaces between system components through the use of interface documentation. These documents include management-level documents such as Working Agreements and Inter-Project Agreements and technical documents such as Requirements Documents and Interface Control Documents.

A significant system design decision made in 2007 for the advancement of EOSDIS to better support a diverse and growing science community was to begin the total migration of the EOSDIS holdings from the tape archives to spinning (hard) disk storage. The higher reliability and lower cost of hard disks made it possible to provide a substantially higher level of data availability through EOSDIS taking advantage of a large number of emerging Web-based services and tools. Another improvement was to rely on commodity hardware rather than specialty systems that required much more expensive, difficult, and cumbersome system-upgrade pathways. Commodity hardware, along with community-vetted standards and protocols, has made EOSDIS a more nimble system, while lowering costs to provide data and services.

Lastly, GSFC membership offered the ESDIS experiences with working in a system of systems, distributed data, and information system architecture. The EOSDIS science operations are executed within a distributed system of many interconnected nodes (that is, Science Investigator-led Processing Systems, or SIPS, and distributed, discipline-specific, Earth science Distributed Active Archive Centers, or DAACs) with specific responsibilities for production, archiving, and distribution of Earth science data products. The DAACs serve a large and diverse user community (as indicated by EOSDIS performance metrics) by providing capabilities to search and access science data products and specialized services. This multifaceted architecture offers challenges but overall offers more advantages in maintaining a system being designed for future requirements and technologies, which underscores the team's observation that no data system is ever fully complete and all architectures should be approached as works in progress.

5.2.2. Jet Propulsion Laboratory (JPL)

From the Jet Propulsion Laboratory (JPL) visit, several topical discussions helped advance concepts and considerations for the EAST study.

Laying out a comprehensive architecture (that is, business, data lifecycle, information models, and software components) is important for developing an organization's data strategy. As data and computing challenges increase, having an architectural strategy upfront to ensure that data, systems, and services can scale and integrate, is essential to an organization. A key aspect of defining that architecture is laying out the methodology used to capture the architecture. This includes defining the principles, stakeholders, and architecture models. In addition, putting the architecture in the context of a larger implementation effort is important to ensure that the architecture can serve as a blueprint for the implementation and not an end in itself. Many JPL systems across space science, earth science, and biomedicine have adopted this approach for ensuring that systems can evolve to support the changing data-intensive needs.

The full data lifecycle perspective is a central consideration in characterizing and resolving data architecture challenges. Typically, 'Big Data' challenges are focused on the archive—the end game—when questions turn to extracting understanding from data. It is important to consider the extended pipeline that reaches back to the point of data collection. As one example, modern sensors and instruments are capable of generating high-volume data streams that can overwhelm capacities to store and move data. In these 'data triage' situations, a user may need to make informed, perhaps irrevocable choices about which data to keep. This imperative may imply moving computational and analytic capability up the data pipeline, perhaps all the way to the point of origin. Note that the burst capacity of cloud services is not relevant in these situations, as cloud services typically are applied only at the archive.

Data stewardship is an important responsibility. Capabilities in support of proper stewardship—such as curation, accessibility, and integrity—have matured favorably at national data centers. Considering favorable maturation, the true end game of the data lifecycle—data understanding—must increasingly support facile and flexible application of data analytics. More and more, multiple and distributed datasets are relevant to addressing scientific, policy, and other questions. As an example to highlight the need to integrate data analytic services, hydrology investigations draw on radar, Global Positioning System (GPS), and in situ well sensors (and more) to address questions in water management regarding activities in deep aquifers. This is the emerging concept of Analytics Centers that builds on the success of Data Centers.

A thorough and thoughtful suite of success metrics is essential to measure whether objectives are being met—both to evaluate data architecture design and to assess effective data system operations. JPL performed an exercise to develop a number of metrics with associated measurements, allocating them to the complementary purviews of data providers, data system, data users, and overall data architecture. Several of these success metrics are included with the EAST study, as well as a set of recommendations.

Undoubtedly, the size, heterogeneity, and distributed nature of datasets will continue to burgeon. This naturally leads to the identification of scalability as perhaps the core architectural objective. Technologies will continue to evolve, and for scalability to be enabled and achieved, it is important to architect data-system solutions with an eye toward the inevitable obsolescence of any given technological embodiment. This, in turn, highlights the importance of the information model, wherein data types, usage patterns, and other defining characteristics of data can be captured and also evolved in a given domain. Coupled with model-based engineering techniques, a data architecture, and particular system solution can be derived from the coherent and stable domain-information model. This information model-driven approach enables the continued use of emerging technologies without having to tear down and rebuild a data system in disruptive cycles—avoiding the implied inefficiencies and costs.

Other considerations collected from the JPL site visit included the following:

- Identifying an architecture methodology to capture principles and different views, and enroll stakeholders;
- Obtaining perspective on the importance of the full data lifecycle;
- Architecting for scalability, extensibility and different architectural topologies;
- Considering a system of systems approach;
- Demonstration of these data capabilities for NASA (Planetary–PDS) and Biomedicine (NIH–EDRN);
- Developing a path forward for integrating data analytics services;
- Open source approaches such as Apache OODT for science data management;
- A suite of success metrics for data architecture design evaluation and operations assessment;
- The importance of an information model-driven architecture;
- Separation/decoupling of different aspects of the systems to enable evolution (for example, adoption of different cloud strategies, support for scalability, and so on); and
- Moving from an architecture to an implementation.

JPL has been working on data architecture challenges for NASA internally (such as, Planetary Data System, various Earth Science activities, Lunar Mapping and Modeling Portal, and so on) and externally (National Institutes of Health [NIH] Early Detection Research Network) for 10 years and more. As an institution, JPL has achieved some progress and success and was pleased to collaborate with the EAST, to both contribute and learn. Many of the discussions have validated the importance of a disciplined approach to architecture definition, along with development of an implementation strategy.

5.2.3. Ames Research Center (ARC)

From the NASA Ames Research Center (ARC) visit, several topical discussions helped advance concepts and considerations for the EAST study, including the following:

- Consider consumers such as Landsat, MODIS, Sentinel–2 and producers such as Web Enabled Landsat Data (WELD) whereby interfaces and network capacity support automation of large-scale data acquisition and distribution. Additionally, consistent product collections aid substantially in the ability to accomplish science.
- Consider data computing capacity in the cloud beyond the high-performance computing as available through NASA or USGS. Strive to support thousands of small users doing data analysis or a few very large ones building global products; the needs and the flexibility of the

system are different for different users. Finally, broaden community engagement in the cloud and accelerate large-scale science using these commercial partnerships.

- Whereas engaging commercial partners can enhance on-demand compute and storage capacity, do not get locked into proprietary application programming interfaces (APIs). EROS must be able to easily move among the systems—this is a fast-changing industry with, at times, a very unpredictable future and business model. Therefore consider proprietary and cost issues carefully during risk planning and mitigation.

While at Ames, the team also spent a good deal of time exploring commercial partnership opportunities based on the Ames model. Additionally, the team discussed and considered application of the Ames Near-Earth Exchange (NEX), NASA's shared high performance computing environment, as part of a potential EROS hybrid architecture.

5.2.4. NOAA's National Environmental Satellite, Data, and Information Service

Several topical discussions with NOAA's NESDIS helped advance concepts and considerations for the EAST study, including the following:

- Ground Enterprise Architecture System (GEARS)
    - One integrated, cross-program, cross-NESDIS team developing a ground enterprise architecture for GOES, POES, NPP, JPSS satellite systems.
    - Transition from stand-alone ground systems with limited interoperability and lack of enterprise approach to future capability development to an enterprise approach with flexible, agile architecture and ops concepts that integrate infrastructure with common services and business processes. This approach aims to improve resource commonality and technology approach for more efficient use.
- Big Data RFI
    - Purpose was to investigate commercial opportunities to help make NOAA's data available in a rapid, scalable manner to the public.
    - An outcome of the RFI was the generation of the Big Data (Cooperative Research and Development Agreement) CRADA, in which multiple commercial providers worked in concert with NOAA to provide access to NOAA's vast environmental data archive. EROS may also benefit by taking advantage of a CRADA to provide broader access to its archives as well.
    - The EAST based its RFI largely on the NOAA framework and experiences.
- Continuation of Comprehensive Large Array-data Stewardship System (CLASS)
    - Potential for hosting CLASS capability at USGS EROS.

During this study period, NOAA was pursuing a very similar investigation track for their NESDIS systems and Big Data architecture. The Big Data RFI served as a good role model for the EAST RFI and NOAA's resulting CRADA proved to be a very interesting topic for the EAST to discuss and assess. Tracking NOAA's progress and experiences throughout their CRADA implementation will benefit EROS IT systems evolution and potentially serve as an example for commercial partnership.

5.3. Potential Public and Private Partnerships

To successfully meet the study challenge statement, the EAST opted to explore potential private and public partnerships. The team established a set of learning objectives and pursued a RFI to engage all types of organizations, including industry, universities, nonprofit organizations, Federal centers, Federally Funded Research and Development Centers (FFRDCs), and international organizations. Learning objectives included the following:

- Potential for public and private partnerships;
- Capabilities pertaining to high throughput and performance computing, storage, data analytics, and information visualization;
- Innovations, products, and opportunities for data and information systems;
- Types of data and information architecture system concepts;
- Limitations pertaining to data transfer, computing, storage, hosting, and so on;
- Provenance methods; and
- Role of government and industry regarding generation of derived information.

5.3.1. EAST Request for Information (RFI)

Working with the team's NOAA representative, the team leveraged NOAA's Big Data RFI and formulated an RFI specific to the EAST's learning objectives. The premise of the RFI was to inform the EAST on the current status of industry sources, technical capacity, operational capability and business practices for potential augmentation or extension to the Center's data and information system architectures and services. As such, the team explored each RFI response for opportunities pertaining to the following:

- Improving access to land imaging data, products, and information;
- Improving land imaging data, product, and information visualization;
- Adding value to land based products and services;
- Enabling surge capacity for high throughput computing and storage; and
- Brokering land based data and services to new user communities.

The EAST received 15 different responses and held one-on-one sessions with many respondents during the EAST working sessions at NASA ARC, USGS EROS, and NASA GSFC during late March and early April.

5.3.2. RFI Findings and Observations

Through analysis and discussions with the RFI responders, the EAST captured a large breadth of capabilities and approaches available for consideration and implementation. Cited respondent capabilities align to the following general categories. Note: none of the following generalized information is proprietary or otherwise sensitive:

- Commercial partnerships for wholesale, scalable cloud and hybrid cloud frameworks, along with technologies for capacity compute and storage, analytic applications, and visualization capabilities.
    - Agile (on demand), resource pooling, elastic and scalable services.
    - Government resources hosted by cloud provider(s).
- Partnerships for provisioned performance compute, storage, and data warehouse capabilities.
- Engineered architecture technology solutions spanning Big Data through analytics.
    - Turnkey system solutions as a service including standard (prefabricated) analytics, visualization, and incorporation of open source applications.
    - Pay-as-you-go use or lease arrangements available (equipment at government site).

- Highly distributed networks for data storage and distribution nodes near user communities to limit latency of high volume data downloads.
- Support services to broker cloud provider "as a service" capabilities.

5.3.3. NOAA Big Data Cooperative Research and Development Agreement (CRADA)

NOAA announced a Big Data CRADA with commercial cloud providers in April 2015. The timing of this announcement coincided with on-going EAST assessments of potential public and private partnerships for cloud-like technology and services. NOAA's CRADA consists of a 3-year partnership with Amazon Web Services, Microsoft Azure, IBM, Google, and the Open Cloud Consortium.

The goal of the CRADA is to promote and facilitate extraction of NOAA datasets to the public cloud to facilitate active, timely data access along with access to co-located cloud-scale computing capabilities. The CRADA provision ensures free and open access to NOAA datasets. Additionally, it allows partners the opportunity to monetize use of the data along with derived products and information. In other words, although not proven, economic opportunities (services) are anticipated in relation to user community value-added processing demands.

Under the terms of this agreement, NOAA implements all of the processing and science computing required to generate its standard datasets. NOAA will also retain a copy of this data and is responsible for all governance and data provenance. Cloud providers then work with NOAA, using its capabilities to extract the data and place a copy within the public cloud. Users may then access this data using the CRADA cloud providers and apply additional value-added processing (analytics) in the cloud environment. This process is monetized by having users pay for access and computing services, including within NOAA projects.

Although this cooperative agreement is innovative, it also presents some risk. Specifically, NOAA and its partners are assuming some cost risk as the economic viability and potential opportunities of the CRADA provisions are largely unknown.

5.4. Input and Experiences Summary

Through the team's interactions with the broader EROS internal community, routine discussions and visits across the team's external partner facilities, and the request for information, the EAST discovered a number of important, applicable concepts including the following:

- There is broad support within the EROS community for enhanced attention to needed architecture evolution, specifically with respect to technology and business models that better enable EROS Center-wide objectives.
- In a similar fashion, the lack of centralized and standardized governance related to technology and resource deployment inhibits effectiveness and flexibility to achieve evolving Center-wide objectives.
- Our partner team members also are grappling with similar challenges and have begun to develop system architecture approaches (specific to technology and business model implementation) to address aspects of these challenges.
- There is a large breadth and depth of potential partner and commercial capabilities readily available for consideration and implementation when strategically applied to EROS architecture alternatives. These capabilities span agile (on demand), resource pooling,

elastic and scalable services, hosting, turnkey solutions, pay-as-you-go models, and highly distributed networks.

- Finally, unique opportunities and solutions are evolving between government and commercial cloud providers that may result in the potential proliferation of EROS Center-wide objectives while becoming economically viable for commercial industry.

6. Architecture Framework

The following section describes the standards approach adopted by the study team to determine and apply architecture views and frameworks to meet the study objectives.

6.1. Framework Overview

Understanding the architectural views and framework of the system is essential to ensuring that an architecture is properly communicated. To support that communication, EAST defined four views around the business model, data and information architecture, application architecture, and the technology architecture elements (fig. 6–1). The comprehensive nature of defining these views of the system helps to establish an overall enterprise view examining the data lifecycle needs, the hardware and software technology needs, the information architecture needs, and the overall business and operations needs for EROS.

Many software architecture frameworks and standards (such as IEEE 1471, TOGAF, DoDAF, and Zachman) identify the need to express views and viewpoints that can be used to articulate an architecture from a specific stakeholder perspective. Building on the view identified and the diversity of EROS, the study team believes that identification of the stakeholders, views and viewpoints are important for communicating architectures using the decomposition that has been developed (fig. 6–1).

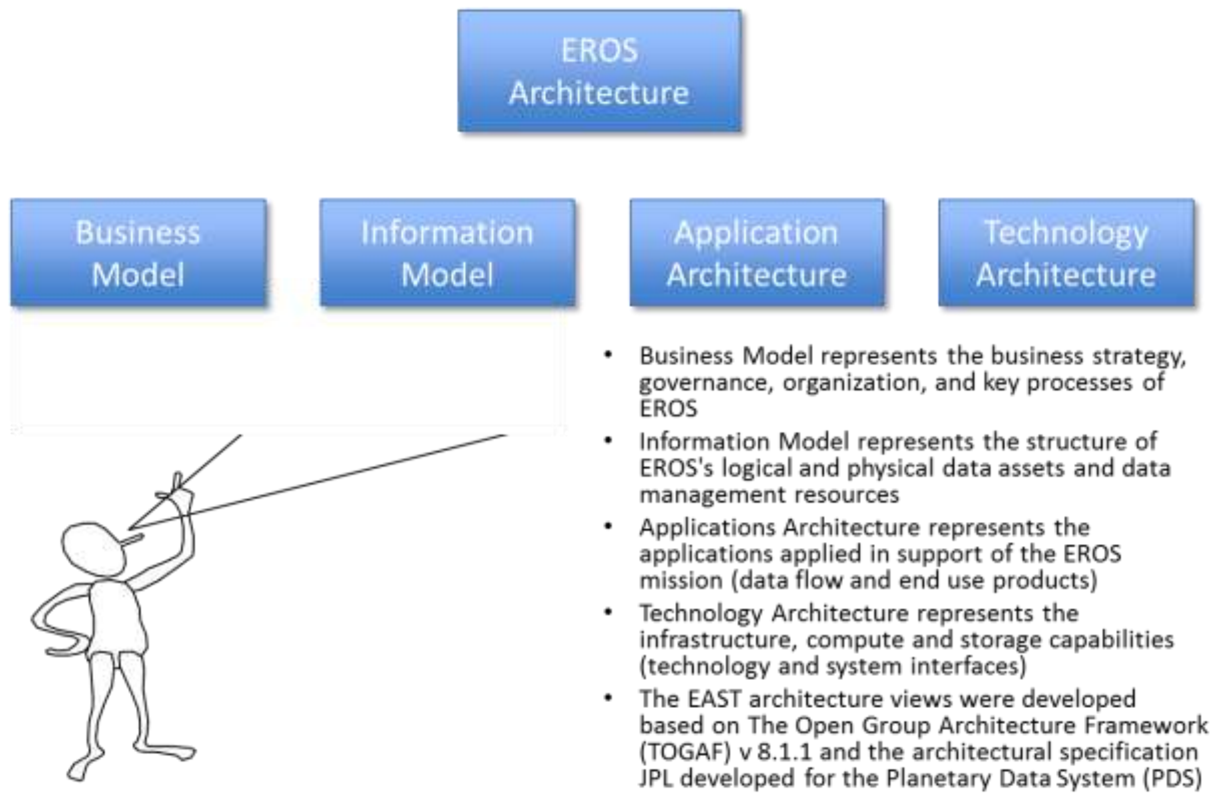


Figure 6–1. EAST architecture views.

To provide a succinct representation of the views for the EROS System Architecture, the planned views were derived from a combination and customization of aspects of the Zachman Framework, TOGAF Framework v. 8.1.1, and the implemented architecture framework of the Planetary Data System developed by JPL. The EAST Architecture Framework diagram is shown in figure 6-2.

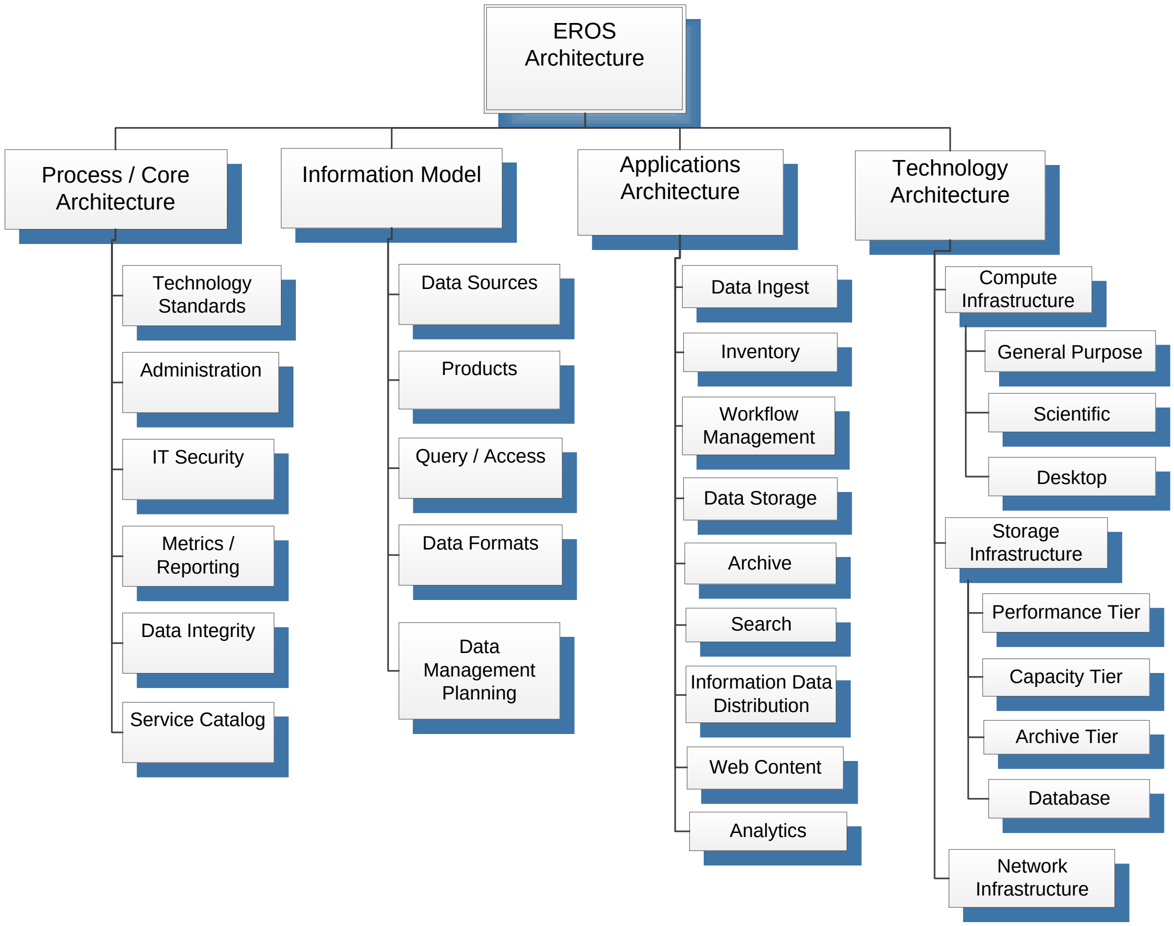


Figure 6–2. EAST Architecture Framework.

Although TOGAF provides great detail regarding how TOGAF maps to the Zachman Framework, the EAST took a much more simplified view. The following sections further define each of the architecture views, specifically identifying the business model, information model, and application and technology viewpoints needed to drive an EAST architectural strategy.

6.2. Data and Information Lifecycle

From the architecture aspect, it is a core perspective to understand the data and information lifecycle. This is an important perspective for developing a comprehensive architecture, and ultimately an operational system, for EROS that can meet the customer performance requirements. The data and information lifecycle perspective also aides in understanding associated software, system, and service needs to support the entire lifecycle of the data. This encompassing approach must include developing capabilities from the EROS data provider all the way to analysis and extracted understanding. This perspective requires architectural considerations for determining and integrating methodologies and infrastructure for capturing and analyzing data across the full lifecycle. To frame the end-to-end concept, the architecture (fig. 7–1) provides guidance for EROS architecture governance, software and system development, and operational concepts that will be required to support effectiveness and scalability necessary to make EROS a national resource.

EROS is designed to be a steward of valuable geoscience research data and information. These data are gathered using many different methods, from lab research to remote Earth observation systems. The lifecycles for these data, however, are the same regardless of their method of discovery.

To properly design a system with this purpose in mind, it is imperative to fully understand the lifecycle of the data from collection to use by the general public. From this, one can properly support the required performance of the system from discovery to scalability and distribution as well as ensure data provenance and integrity for researcher use to support the business needs of the enterprise.

This practice ensures that the data is reliably managed, is supportive of discovery, and is fully used. It is important to point out that data across this entire lifecycle view should not be considered data at rest, but rather data that are discoverable, accessible, and useable to update plans and inform other decisions, and enable science. A well-architected data system from data generation through data analysis and visualization must be in place to support all of these objectives.

Essential areas of the lifecycle (with considerations at each stage) include the following:

- **Data Generation:** the function of acquiring data from an instrument, scientific experimentation, sample gathering, or other source. For the purposes of EAST, it is assumed that this function occurs outside the scope of this architecture.
- **Data Curation and Preparation:** the function of preparing data for use within EROS. This includes definition and annotation of metadata, linking to other data and systems, and ensuring that data is formatted and compliant to EROS standards. Consideration should be made to ensure data has appropriate structure, provenance, and integrity information.
- **Data Transport:** the function of moving data or metadata, or both, from a data provider to EROS. Given the variety of data, it is anticipated that many different technology solutions may be applied for transferring data into EROS.
- **Data Ingest:** the function that provides the services and functions to accept data including metadata, observational data, documents and other data from data providers, and prepares the data for storage and management within EROS. Data Ingest functions include the following:
    - Receiving or retrieving the data or metadata or both from providers,
    - Executing quality assurance on the submitted data, and
    - Harvesting metadata to enable cataloging and linking of data.
- **Data Management:** this function encompasses many different functions including storage management, metadata management, preservation management, data provenance, and management of data citations.
- **Data Search, Discovery, Access, and Distribution:** this function is a major part of constructing the data analytics capabilities for EROS. These functions ensure that data can be located, accessed and distributed to users and other functions within EROS (including data analytics algorithms).
- **Data Analytics:** this function enables the analysis of massive, distributed, heterogeneous data using scalable computing infrastructure and includes the following:
    - On-demand processing of data,
    - User-driven workflows,
    - Integration with scalable-computing infrastructures (high performance computing [HPC], cloud), and

    - Application of computational methods (data fusion, machine learning, data mining, statistical analysis) to support user-driven analysis.
- **Visualization:** this function integrates with data analytics to support rendering and communication of complex data into visual formats to support data understanding and discovery.

6.3. Business Model

The EROS business model recommended by the EAST provides many functions that form the overall operation of the EROS architecture including system and service management. Business functions include governance and establishing, managing, and maintaining the following:

- Architecture artifacts (such as conceptual design, reference architecture, roadmap),
- Standards and policies for data (such as data and metadata standards),
- Standards and policies for system (such as system component interface protocols, data transfer protocol),
- Best practices and policies for technology development and adaption (such as cloud and HPC),
- Best practices and policies for day-to-day operation (such as IT security, system monitoring),
- Guidelines and charters for governance (such as leadership council, various EROS committees),
- Guidelines and processes for resource planning (such as funding and system resource allocation),
- Support for continuity of operations and other risk factors, and
- Performance metrics.

These services and functions are important aspects of operations, evolution, and sustainability of the long-term EROS architecture implementation. Decision making, execution, and communication of these aspects affect all stakeholders in ensuring that EROS remains a long-term productive environment for geosciences.

6.4. Information (Data) Model

The information model is essential to describing the data and information of the overall system, its structure, and organization. A well-defined information model is instrumental in helping to scale to support management, discovery, and analysis of the data.

An information model is a representation of concepts and the relations, constraints, rules, and operations to specify data semantics for a chosen domain of discourse, see the following:

- Entities to be processed,
- Entities that provide context,
- Relations between entities that provide meaning, and
- Definition of key information objects (to be defined).

The information model provides a sharable, stable, and organized structure of information requirements or knowledge for the domain context. Information models play an essential role in not only the data architecture but generally in driving the overall definition of information systems, particularly for information-intensive projects. Given its role in driving systems, it is important that the information model remain independent of its implementation.

In addition, well-defined information models can be used to describe relations of information across multiple projects and domains. At an enterprise level, the information model can provide a framework for inter-relating multiple models. From an EAST perspective, having a multilevel strategy for managing and inter-relating information models will drive forward an information architecture strategy for accepting and integrating data for EROS for many years to come. The information model includes the following:

- Models for the data, formats, and organization,
- Data dictionaries describing core metadata elements,
- Formats of the data,
- Models for querying the data (such as configuring search engines), and
- Models for analytics.

What is important is the relation between defining core aspects of the system, independent of the software, and using that to drive the software and data configuration.

6.5. Application and Technology Architectures

The Application and Technology Architectures form the basis for the hardware and software architecture. The application and technology architecture elements include scalable IT infrastructure services (computing, networking, storage) and the applications that enable processing, management, discovery, and analysis of the data.

7. EROS As-Is Architecture

The As-Is Architecture provided an avenue for analyzing current work flows or use cases which became a solid starting point for the set of observations and recommendations from the EAST. The following section describes the process by which the EROS As-Is architecture was defined and assessed.

7.1. Governance and Management Culture

Although EROS does not maintain a Center-wide architectural picture as of the writing of this report, which is necessary for forming relevant Center-wide study results, an architectural picture of EROS was deemed vital to the architecture study.

Initial discovery verified that without a centralized architectural establishment, the architectural direction is determined by individual projects, branches, or funding sources. So, the first step in creating a Center-wide picture of EROS architecture focused on identifying the owners of architectural pieces. An initial 18 possible sources of architecture information were identified. Some of these were grouped together and external collaborators were eliminated to narrow this down to 11 sources.

A survey was created and sent to each of the 11 possible sources of architectural information. An agreement also was created to allow Technical Support Services Contract (TSSC) support in collating the information on the surveys and subsequent information gathering. This allowed for a shorter turnaround time given the tight schedule required to meet EAST objectives.

Using a basic classification of the architecture based on early target architectural models, a straightforward concept of how to illustrate the architecture and model data flow was created. This was enhanced with functional information to create an As-Is architectural diagram to represent a Center-wide view. After reviewing with EROS staff at the first EAST Technical Interchange Meeting in February, and comparing to the use cases, a high-level diagram was finalized representing the EROS As-Is architecture. This high-level architecture diagram (fig. 7–1) constitutes a deliverable from the EAST to the Sponsor and was considered by many within EROS to be the first successful attempt at deriving this information.

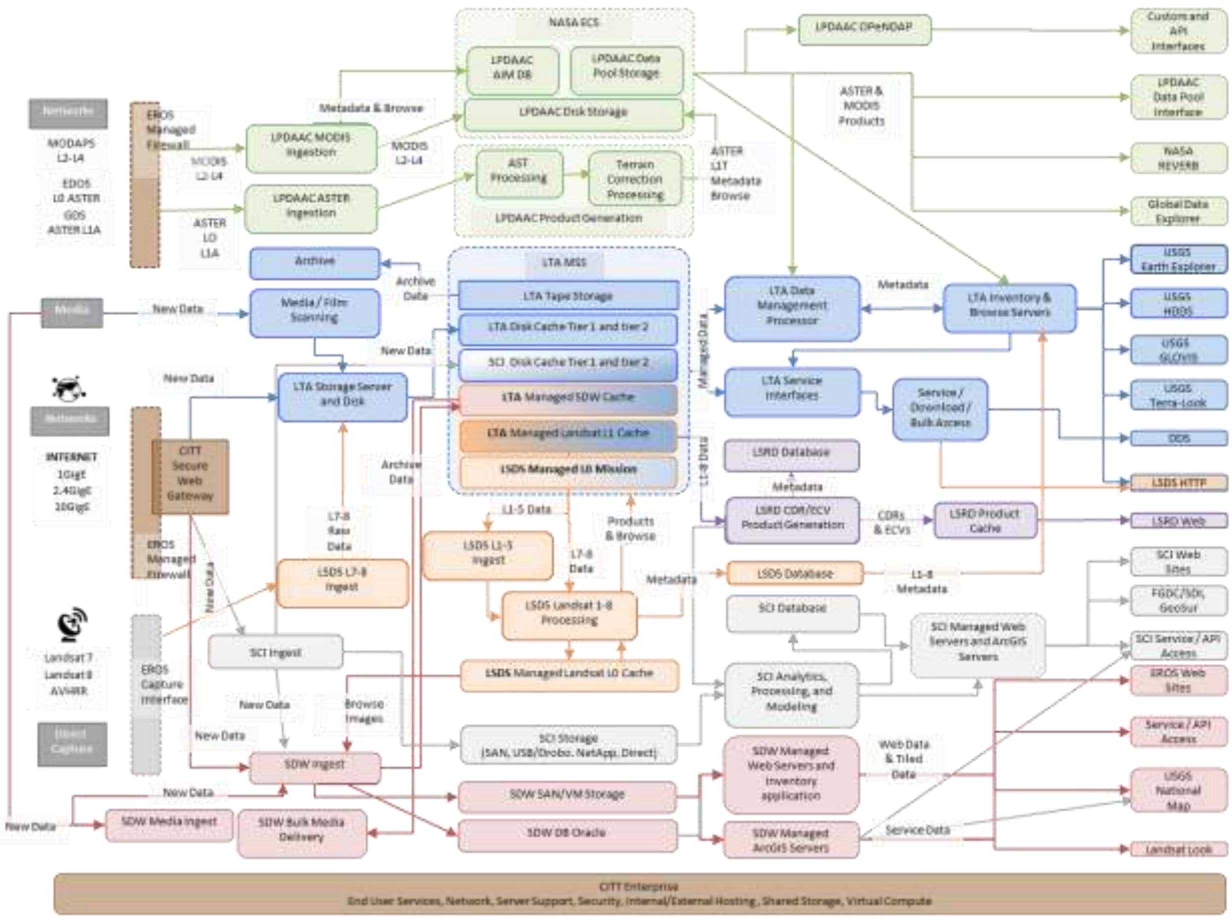


Figure 7–1. EROS High-Level As-Is Enterprise Architecture.

7.2. Curators

After collating the information from surveys and any existing architecture documentation gathered, the number of responsible parties for architectural pieces was reduced to six. The six sources are a mix of branches and projects, so the term curator is used to refer to them, although project is commonly used because most curators are projects. An architecture curator is the group that is responsible for the ownership, management, operation, maintenance, and agreements for a given set of architecture components. In the As-Is diagram, each curator is assigned a different color.

The six curators identified are the Center Information Technology Team (CITT– Brown), Land Processes Distributed Active Archive Center (LP DAAC – Green), Land Satellite Data Systems (LSDS – Orange), Long Term Archive (LTA – Blue), Science (Mostly Research Branch and Applications Branch - RB&AB – Gray), and Spatial Data Warehouse (SDW – Red). Additionally, the LSDS Science Research and Development

(LSRD – Purple) is called out even though it is part of the LSDS systems. This was called out separately to help demonstrate an LCMAP-like data flow.

7.3. Use Case Data Flow Examination

Whereas LCMAP does not yet exist as a project with infrastructure in the As-Is architecture, by examining how current science projects make use of the infrastructure, the team approximated LCMAP potential usage. Figure 7–2 shows a science project (LCMAP) requesting data through an LSRD interface, which, in turn, requests data from the archive, which could request data from LSDS. The science project then needs to transfer data to its resources and process it before it can distribute the final products through Science, SDW, or LTA resources (or multiples thereof).

This shows that the current architecture works, but highlights a few consequences of this curator-driven architecture. There are a number of points where the boundaries between curators are crossed, which means agreements must be made between curators. The project that is attempting to create/setup a processing chain must discover and pursue each agreement separately. The project must also ensure that interoperability remains in place and that any necessary funds exchange occurs. Because the data must cross several curator boundaries, it is often duplicated (at least temporarily), bringing multiple computing resources into use, resulting in a situation that can escalate quickly to a point that is not sustainable or affects other systems. There are few checks and balances, and the data flow is unnecessarily complex.

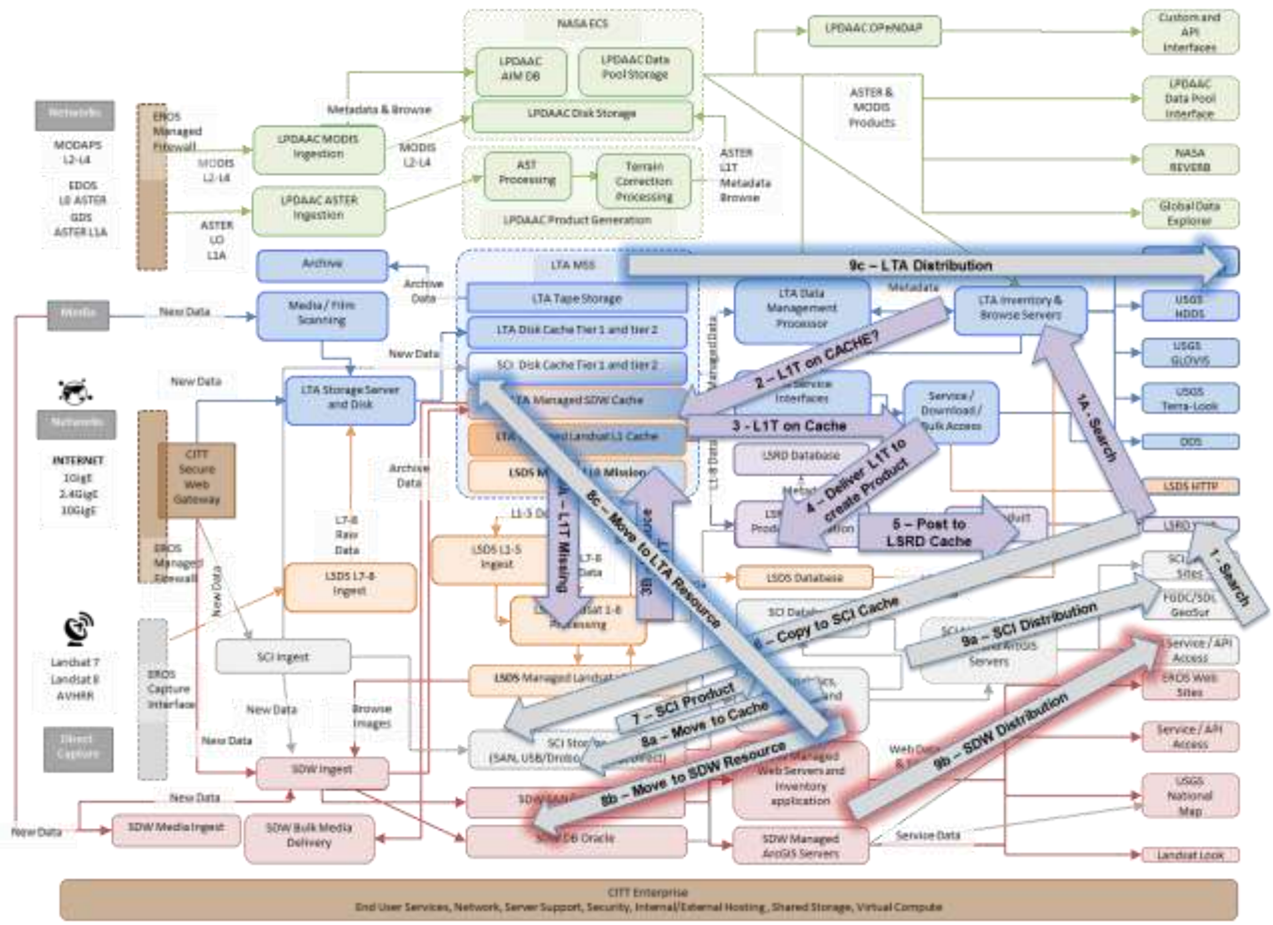


Figure 7–2. Notional LCMAP Data Flow Within As-Is Architecture.

With no over-arching management of this arrangement and agreements not standardized across the Center, simply keeping this architecture running may be unstable. Without service catalogs, or predefined listings, new projects need to either be familiar with the resources available or spend considerable time researching to determine what is available before pursuing individual agreements. Costs are usually shared through agreements with larger projects often shouldering most of the costs (occasionally absorbing costs to support smaller projects).

7.4. Perspectives on the As-Is Architecture

Figure 7–1 represents a limited subset of the architecture to make it reasonable for viewing and analysis. The foremost intention is to focus almost entirely on architecture pieces housed in the computer rooms with only limited desktop/user systems. The underlying facilities are also not represented in this diagram. Of the four major subdivisions of the target architecture framework, this diagram displays technology, with some business model and applications architecture. The information model is not covered within the As-Is architecture as recommendations for this will flow from the Analysis-Ready Data Definition, or other project-led efforts.

Whereas not running a supercomputer or massive dynamic processing resources, EROS does host substantial computing resources that span all the identified curators. One of the underlying components involved with moving from discrete computing systems to a software-defined enterprise is that the new computing architecture requires virtualization. Virtualization is already present in the Center, and is more standardized than the storage; but it is managed in clusters with some shared components. A move to enterprise architecture means some changes would need to take place concerning how the virtualization is managed. Migration from the discrete computing systems presently within curator domains will also involve changing the EROS enterprise business model to allow sharing of the computing resources, especially any excess capacity available between larger processing jobs.

Current EROS computing technology consists of many different underlying pieces, each representing a set of discrete computing resources. In general, the technology within each curator domain tends to be compatible and follows general standardization. Across the Center, however, the technology varies considerably (except for virtualization, which uses VMWare), resulting in more places where changes to accommodate enterprise architecture management might have to occur. In the case of computing capacity, there is little sharing of resources, similar to what was observed in storage technology.

A constraint on the computing technology is that access to data is essential. This means that data transfer by the network and storage of that data may be tied to the computing technology. This particularly comes into play in data analysis, transformation, and any high performance computing (HPC) or clustered computing environments. Currently, there is very limited HPC technology in the EROS architecture.

7.5. As-Is Architecture Observations

The target architecture framework uses four categories of architecture. Although aspects of the Information Model and Application Architecture are visible in the As-Is diagram (fig. 7–1), it primarily focuses on the Technology Architecture category. The Technology category has three subcomponents: Compute, Storage, and Network. Network infrastructure at EROS is mostly managed by CITT as an enterprise resource, so it will not be discussed further in the As-Is section of this document. Networks will still be covered in the recommendations in section 8.5 to show how this integrates into the target architecture, but the architectural pieces will rely on the plans CITT is already developing as of the writing of this report.

The storage in use at EROS includes a wide variety of technology. For the purposes of this study, the following types of storage were considered: Archive, Mass Storage System (MSS), Storage Area Network (SAN), Network Attached Storage (NAS), External Drives, and System Drives. There are some pieces (most notably the Archive and MSS) that, similar to the network, are already managed in a mostly enterprise fashion.

Some of the other types of storage will be more difficult to turn into enterprise resources. The mix of technology and requirements mean that some components will not simply merge into other systems. Change will be required, or multiple systems must be maintained. Because, most project-controlled storage has been obtained to meet specific project requirements, a new requirement, such as a need for extremely high performance processing storage, does not have sufficient resources that can meet it within the building. New resources would have to be obtained rather than merging existing components. Any recommendations considered would still have to support the project-specific requirements that individual components are currently satisfying with a wide variety of technology.

The underlying message is that, from a technology viewpoint, this architecture can satisfy even complex requirements for analysis such as those present in LCMAP. It is quite evident, from looking at a complex data flow, that there are many opportunities for increased efficiencies.

The project- or curator-focused requirements and architecture means directly assigned or assumed requirements are usually well satisfied. Center-level, science-based requirements without good representation are met haphazardly, usually only when a project arranges for the implementation.

Varied architecture implementations lead to more difficulties securing Center resources, or standardizing implementations across the Center. It can even mean reduced opportunities for cost savings by pooling resources for acquisition or usage.

Changes to underlying resources may have ripple effects across many projects that may impact Center activities with little notice, different effects per project, and are often open to interpretation. This means there is a fragility or instability present within the complex data flows that cross many curators.

8. Target Architecture Alternatives and Selection

Using the As-Is architecture assessment as a baseline, the EAST identified three additional candidate architectures that respond to the EAST challenge statement. This section describes the additional architecture alternatives studied and assessed by the EAST, along with the integrated success measures and business objectives used in selecting a single architecture recommendation for EROS.

8.1. Application Architecture View

To develop the target architecture concepts, the EAST team first developed a high-level view of the EROS application architecture. This application view models the flow of data from data sources into the EROS architecture and out to applications and users.

The application architecture view establishes discrete, high-level capabilities for data ingest, processing, query and distribution. It also establishes tiers of storage for various levels of data/processing with a strong underlying inventory and metadata model. It also calls for future capabilities in the form of advanced analytics, primarily in support of the LCMAP concept.

This architecture view was the basis for the development of the specific target architecture concepts described in figure 8–1. The application view is an abstraction of the major functions of EROS that will need to be completed regardless of the target architecture chosen, so the basic application architecture is the same across the various alternatives.

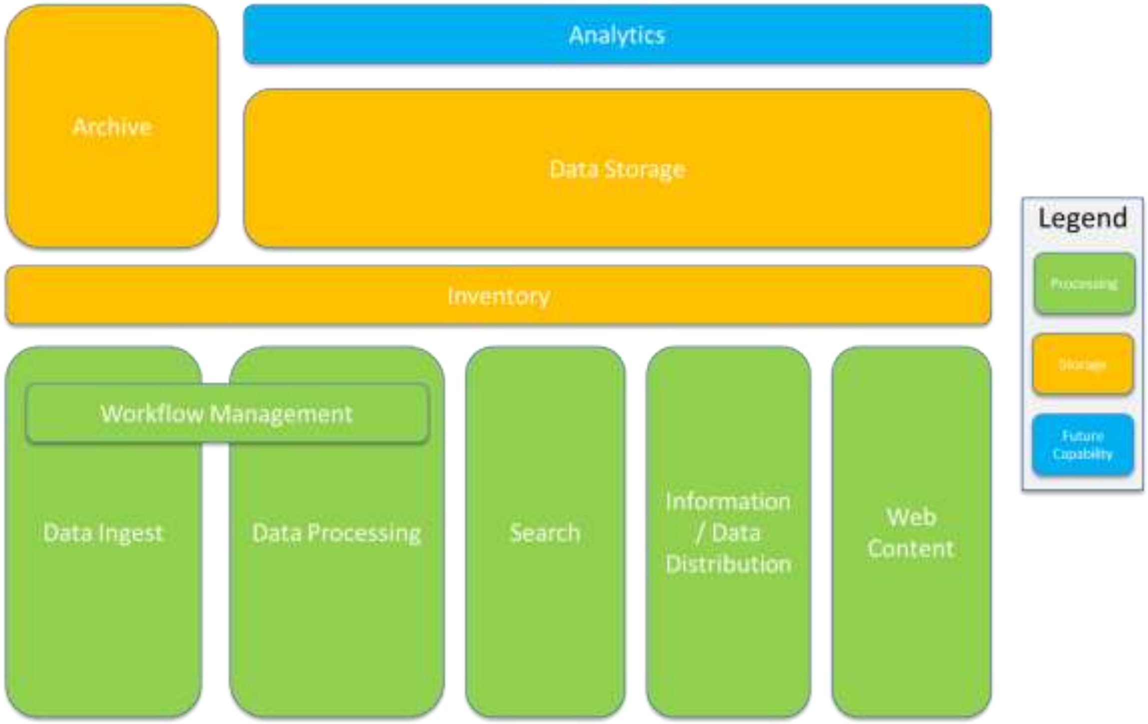


Figure 8–1. EROS Application Architecture.

8.2. Response to EAST Challenge

The study team relied heavily on the following EAST challenge statement:

> *"to define and assess candidate architectures that support current needs and allow for the expansion of the EROS mission to include providing land change data, information, and knowledge products, along with a path for evolution from current capabilities."*

to determine the set of architecture alternatives to consider. Specifically, the EAST focused on candidate architectures that do the following:

- Enhance and optimize the EROS As-Is architecture,
- Identify and streamline opportunities for shared services across project activities,
- Prepare for next generation land imaging and like missions (for example, Landsat 9 and Sentinel–2),
- Address capabilities for ready access to EROS data holdings and computing capacity to generate information on land changes as they are detected (such as science data systems like LCMAP), and
- Address evolution of systems and data analytics services needed to enable science from data.

Additionally, the team applied industry trends discovered through the RFI process, leveraged experiences and lessons learned from EAST partner organizations, and applied knowledge gained from use cases to derive viable alternatives that support the EAST challenge statement. Finally, the team executed a preliminary assessment of the team's success measures: effectiveness, flexibility, sustainability, reliability, and security.

8.3. Architecture Alternatives

The four resulting architecture alternatives feature the following characteristics:

1) **As-Is architecture**: This approach represents the "do nothing" decision and continuing with the current highly decentralized and independent architecture evolution. In this case, there is no centralized governance or system-of-systems approach. There is no strategic, centralized architecture planning for technology, infrastructure, or software planning. Finally, while sharing resources between projects is an option, individual projects decide what resources to share or not share.
2) **Projectized Matrix:** This approach emphasizes minor changes from the As-Is approach through limited centralized governance or system of systems guidance offered to projects. In this case the architecture and its evolution continue to be decentralized (independent). However, greater management emphasis is placed on continued and enhanced sharing of project resources.
3) **Enterprise**: This approach stipulates an effective centralized governance and system-of-systems approach be established and used to manage and oversee the Center's architecture evolution as a whole. Additionally, implementation of "as a service" approaches for infrastructure, software, and platform resource sharing is pursued. This approach also allows for managed and strategic planning of Center architecture evolution.
4) **Cloud-Centric**: This approach provisions almost all Center computing and storage resource capabilities off-site to private or public cloud providers. Specifically, this approach proposes a substantial reduction in Center science computing, storage, and web-enabled capabilities and instead emphasizes a reliance on cloud providers. Center capability is primarily limited to mandated data capture, archive, and specialized processing capabilities. In this model, the Center's architecture posture and configuration is almost solely based on the evolution of vendor services.

The EAST executed an assessment of each architecture alternative against the five success measures provided by the steering committee (effectiveness, flexibility, reliability, security, and sustainability). A detailed explanation of the alternatives and a summary of the related assessments are provided in sections 8.3.1-8.3.4 below.

8.3.1. As-Is Architecture

For a detailed description and general observations of the As-Is architecture, see section 7. The assessment of the As-Is architecture against the EAST measures of success is provided in figure 8–2.

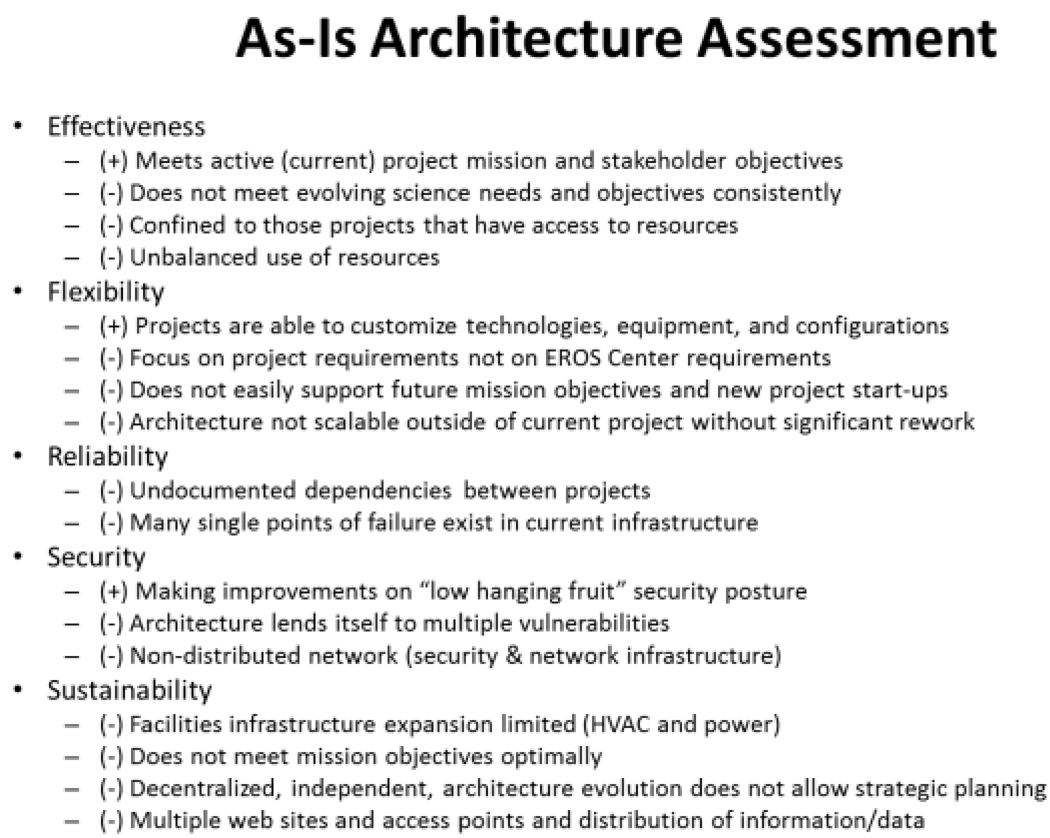


Figure 8–2. As-Is Architecture Assessment Summary.

8.3.2. Projectized Matrix Architecture

The Projectized Matrix architecture is essentially an extension of the As-Is architecture with a more robust interproject communication and limited governance which is intended to encourage sharing of resources and infrastructure between projects. In this model (fig. 8–3), projects retain responsibility for selecting, implementing, deploying, and maintaining their own resources. Projects can leverage their own internal expertise and provide their infrastructure as services to other projects as needed.

The Projectized Matrix architecture does not establish a Center-wide enterprise capability, whereas it does allow for a modest step toward broader strategic architecture control; projects can realize increased resource efficiency by working with each other without requiring complete "enterprising" of computing and storage.

Projects control when and where resources are available, and although the Projectized Matrix architecture encourages collaboration between small projects for storage and compute capacity, ongoing evolution is dominated by large project requirements and agendas. Smaller projects may find it difficult to collaborate with larger projects that have little incentive to support smaller activities, and new projects have to either start from scratch or pick a project to "hang off of."

Because the evolution of the overall technical environment continues to be driven by individual projects, this alternative results in an unpredictable lifecycle environment where resource availability is dependent on individual project lifecycle needs. There is no system-of-systems; the architecture continues to evolve in a decentralized, independent fashion for the long term. Consequently, there is no mechanism to predict architecture evolution or strategically plan for architecture improvement over the long term. Adaptation of applied technology is ad hoc; socializing which technologies work well and which do not work well is dependent on sharing across independent projects.

Finally, projects come and go resulting in the possibility of noncoincident or temporary sharing. For example, a small project could be dependent on a larger project for storage and computing resources; if that larger project concludes and these resources are removed, then the small project is put in the position of either having to procure its own resources or find another project to provide resources.

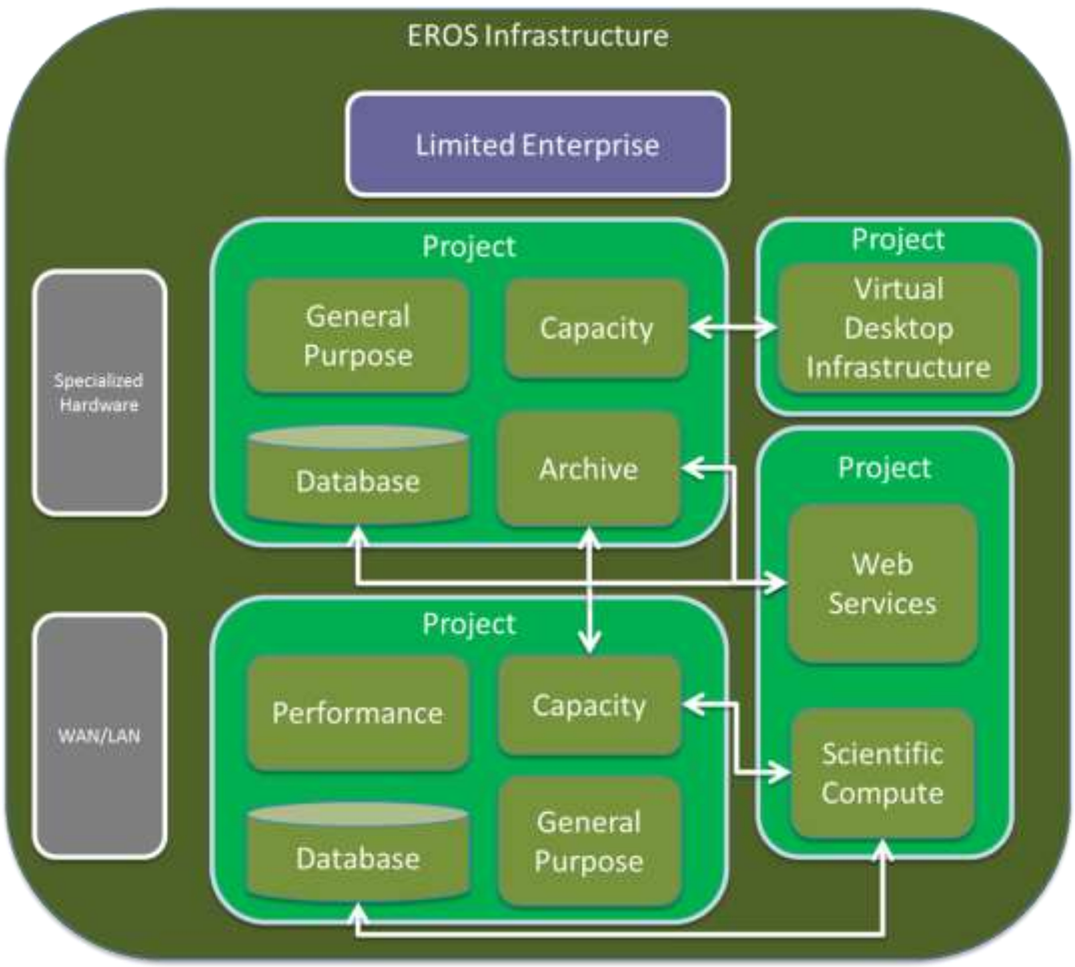


Figure 8–3. Projectized Matrix Technology View.

The following represents the study team's assessment of the Projectized Matrix against the measures of success:

- Effectiveness

  The Projectized Matrix architecture meets active (current) project mission and stakeholder objectives, but it does not meet evolving science needs and objectives consistently. The benefits of this architecture are confined to those projects that have access to resources, but sharing is encouraged. There is a more balanced use of resources than in the As-Is architecture (better understanding of unbalanced resources).

- Flexibility

  Projects are able to customize technologies, equipment, and configurations and focus on project requirements and on Center requirements. The Projectized Matrix architecture does not, however, easily support future mission objectives and new project start-ups. Furthermore, the architecture is not scalable outside of the existing Center scope without substantial rework.

- Reliability

  Although the Projectized Matrix alternative offers an opportunity to better understand the underlying reliability issues with the As-Is architecture by documenting the dependencies between projects, many single points of failure exist in the current infrastructure and this architecture does not change that.

- Security

  The Projectized Matrix architecture allows for improvements on “low hanging fruit” security issues, and the sharing of resources between projects allows for potential efficiency gains in security procedures. However, the continued ad-hoc nature of this architecture lends itself to multiple vulnerabilities, and the use of a nondistributed network introduces certain security and network infrastructure issues.

- Sustainability

  The Projectized Matrix architecture does not meet Center mission objectives optimally. Although the limited governance of this model does allow for some improved strategic planning, the architecture remains decentralized and independent and will evolve as such. Furthermore, this architecture does not attempt to address current or upcoming issues related to EROS facilities infrastructure (such as, heating, ventilating, and air conditioning [HVAC] and power).

8.3.3. Enterprise Architecture

The Enterprise architecture replaces project-managed silos with centrally managed enterprise services and enables managed, strategic planning of architecture implementation. This includes a living service catalog within a managed, controlled enterprise for needed capability using service-level agreements (SLAs.)

In this architecture (fig. 8–4), the enterprise is responsible for balancing resources across projects. Rather than procuring, installing, and managing their own equipment, projects would instead pay for this infrastructure as a service and have it provisioned to them. Implemented properly, this architecture provides the flexibility to expand into trusted partner or commercial cloud providers or both for additional computing and storage resources as needed.

Although this architecture represents a fairly dramatic change from the current Center environment, projects would retain authority over all of their requirements and implementation approaches. In this architecture, projects may have more than one role—as project business and service providers. Specialized hardware (outside of the centrally managed enterprise resources) is allowed as needed in support of project-specific requirements.

The Enterprise architecture results in a consistent and predictable environment with controlled evolution of applied technology and practices. It enables more flexible resource deployment that enables projects (in particular small projects) to focus more on the work that they need to do and less on how to get it done. It enables more effective cross-project use of shared resources, and projects with short-term resource needs (for example, limited studies or small development efforts) can quickly stand up resources without having to go through a lengthy procurement process for equipment that they may only need for a short period of time. In this alternative, it may potentially be more difficult to guarantee a specific project's level of service (compute capacity) than when a project procures its own equipment, but the overall flexibility and efficiency of resource deployment in this model is far preferable to the situation where large amounts of project-specific equipment sits idle most of the day.

Because the Enterprise architecture is a departure from the existing processes at EROS, it will require a new Center business model (that is, including cost models, business practices, and a governance model). This business model may be complex, given the need to distribute costs for varying technologies and service levels.

This model also reduces single points of failure within the EROS architecture, but could create a broader single point of failure given the move to more centralized computing and storage infrastructure. This can be mitigated through proper design and diversity of failover mechanisms and through the implementation of a mesh network that is less susceptible to the loss of a single network path.

Finally, while seamless expansion into the cloud is perceived by some individuals as the "silver bullet" for an enterprise architecture, this approach requires careful consideration because it can be expensive even in more limited capacities, and variability in how this capacity ebbs and flows could present financial (business model) challenges.

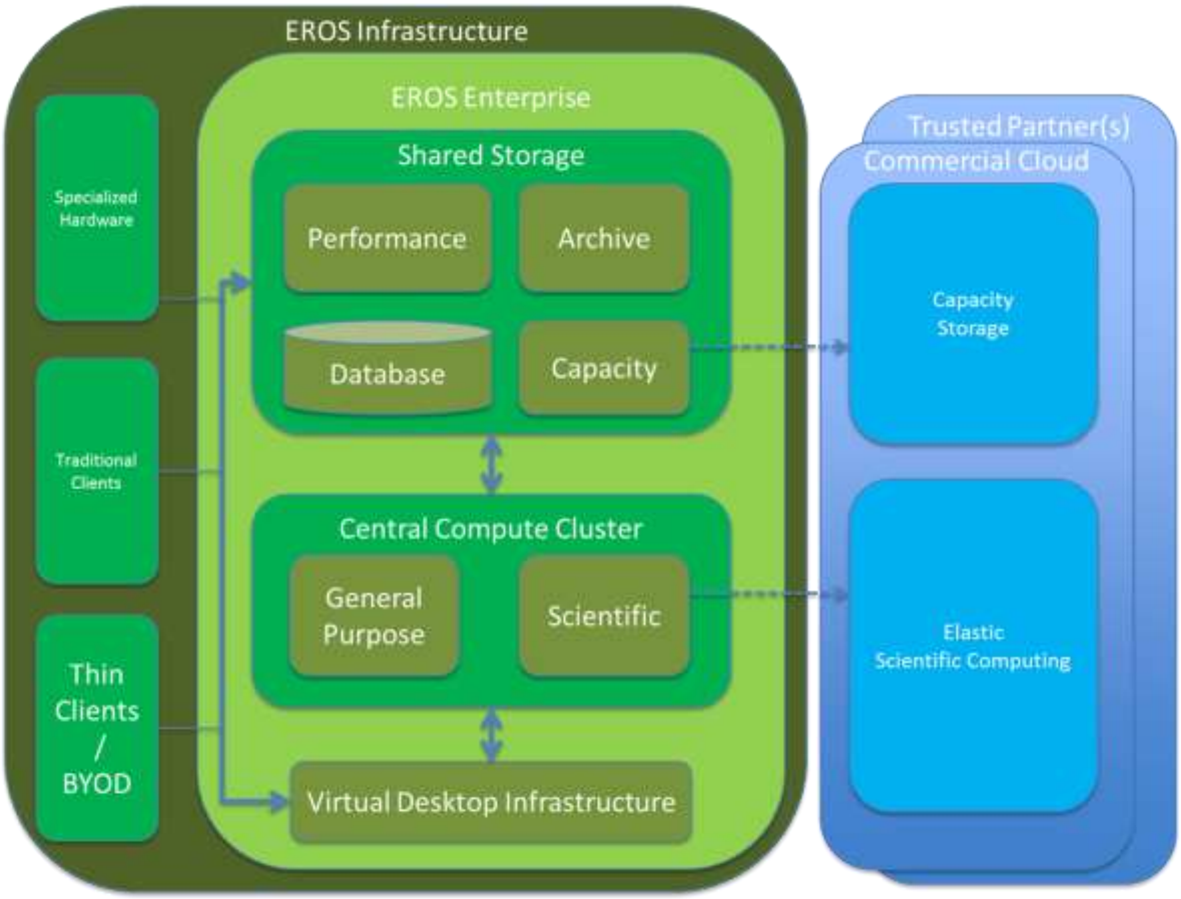


Figure 8–4. Enterprise Architecture Technology View.

The following represents the study team's assessment of the Enterprise architecture against the measures of success:

- Effectiveness

  The Enterprise architecture meets active Center mission and stakeholder objectives and supports evolving science needs and objectives consistently. It ensures that all projects have access to shared resources, and that there is a more balanced use of resources. The Enterprise architecture requires additional overhead for coordination and planning, but this is more than offset by the effectiveness of the architecture.

- Flexibility

  The Enterprise architecture supports future mission objectives and new project start-ups through "as a service" resources. It is scalable outside of EROS's current projects without substantial rework, and it should make it easier to acquire needed resources in a timely fashion. The focus on shared resources and a common service catalog, however, may make it more difficult for projects to customize technologies, equipment, and configurations.

- Reliability

The Enterprise architecture will facilitate the adoption of standard hardware and software configurations to implement the shared computing and storage resources within EROS. Standard configurations such as these (and their associated documentation) should increase reliability and more readily facilitate automated failover. The enterprise architecture also reduces single points of failure that are present in the As-Is infrastructure.

However, the reliance on centralized computing and storage resources also increases the potential for outages to impact multiple projects. This risk can be mitigated with proper infrastructure design (for example, redundancy, automated failover, and so on).

- Security

  As with the reliability observations, reliance on standard configurations makes the Enterprise architecture less susceptible to security vulnerabilities. Additionally, centralized administration will streamline security decision implementations, making things like critical-priority security patches easier to deploy in the required timeframe.

- Sustainability

  The Enterprise architecture meets mission objectives optimally by providing for a managed, strategic planning of EROS architecture evolution. It provides consistency for Web sites and access points (hosting), along with distribution of information/data. Finally, it allows for better use of limited facilities infrastructure (HVAC and power).

8.3.4. Cloud-Centric Architecture

The Cloud-Centric architecture leverages cloud providers for most Center operational mission activities. It reduces the amount of physical infrastructure present at EROS down to only that which is required to ingest and archive the core required datasets. The physical archive remains at EROS, but all higher-level processing and data distribution are handled externally through one or more trusted partners or commercial cloud providers.

The Cloud-Centric architecture (fig. 8–5) enables managed, strategic planning of architecture evolution in the cloud. Rather than procuring, installing, and managing their own equipment, projects would instead pay for capabilities as a service and have them provisioned by way of the cloud. The enterprise would be responsible for facilitating interfaces to trusted partners or commercial cloud providers or both.

As with the previous architectures, projects would retain authority for all requirements and implementation approaches, and specialized hardware would be allowed as needed in support of ingest and archive activities.

Although the Cloud-Centric architecture would provide great flexibility to deploy additional computing/storage resources, EROS would become almost totally reliant on a cloud vendor(s) to implement basic Center functions, and cloud costs would drive architecture solutions. This alternative involves wholesale change to the architecture and business model (that is, revolutionary not evolutionary). The Platform as a Service (PaaS) and Software as a Service (SaaS) capabilities from public cloud vendors provide ample architectural flexibility, if EROS systems are modified to take advantage of it. Because cloud computing is not part of the As-Is EROS infrastructure, there remains much uncertainty, especially considering the constant change in today's (2015) cloud-service offerings, and the fact that Center systems would require substantial architectural changes to move to this environment. Furthermore, these capabilities would be limited to the services that the vendor is willing and able to provide.

Because public cloud resources typically are deployed in a highly available manner in multiple geographic locations, system reliability can be far greater than what can be attained locally. Furthermore, Web sites and other public user content are very easy to migrate to a public cloud environment; however, the current storage capabilities of the public cloud providers do not appear to be conducive to the analysis-ready data (ARD) concept as it is currently defined.

The pay-as-you-go capability provided by the commercial cloud works well for short term or variable use, but not for steady processing/storage scenarios. Moving large, persistent pieces of the EROS architecture completely into the public cloud would likely come at a considerable cost. The potential exists for agreements or arrangements that reduce or shift these costs, but with no real guarantee that they would exist in the long term.

That said, the Cloud-Centric architecture reduces local infrastructure costs because of the bulk of Center capabilities running externally, and a corresponding reduction in required network bandwidth can be realized if product distribution is done from cloud storage.

Finally, although many of the benefits of cloud computing can be applied to any architecture alternative, it is not necessary to move everything to the cloud to take advantage of it.

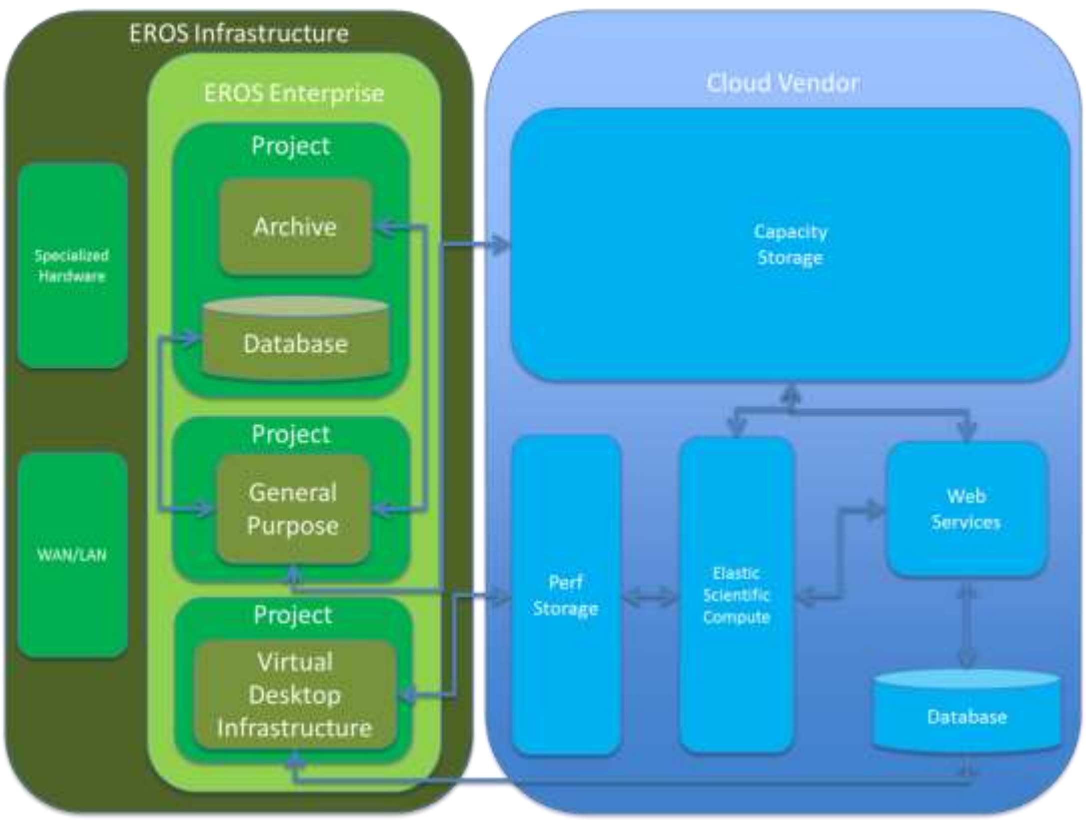


Figure 8–5. Cloud-Centric Technology Architecture View.

The following notes represent the study team's assessment of the cloud-centric architecture against the measures of success:

- Effectiveness

  As with the Enterprise architecture, the EAST believes that the Cloud-Centric architecture consistently meets active Center mission and stakeholder objectives and supports evolving science needs and objectives. It ensures that all projects have access to the same resources. However, this effectiveness comes at a cost; there is additional overhead for coordination and planning, cloud costs will quickly surpass local infrastructure costs for sustained projects, capabilities are limited to what cloud providers offer, and re-architecture is required to use cloud resources effectively.

- Flexibility

  The Cloud-Centric architecture supports future mission objectives and new project start-ups through "as a service" resources. It is scalable outside of current EROS scope without substantial rework, and makes it easier for projects to acquire needed resources in a timely fashion. This alternative provides projects with full access to everything cloud providers offer (assuming that these services are economically feasible), although it is more difficult for projects to customize technologies, equipment, and configurations if these things are not offered by the cloud provider.

- Reliability

  Given the broad use of virtualization, geographic redundancy, and automated failover, the published availability of cloud resources is typically far higher than what is achievable at EROS. Furthermore, the use of cloud resources will reduce the number of single points of failure in the As-Is Center infrastructure.

- Security

  Because the responsibility for security shifts substantially toward cloud vendors in this alternative, the Cloud-Centric architecture is less susceptible to security vulnerabilities, and those that arise will impact EROS resources less than in the other architecture alternatives. Additionally, centralized administration will streamline security decision implementations, making things like critical-priority security patches easier to deploy in the required timeframe.

- Sustainability

  Moving to a cloud-centric architecture would substantially reduce the amount of infrastructure that would have be sustained at EROS, which in turn would lead to much better use of limited facilities infrastructure (for example, HVAC and power). Using cloud services would provide increased consistency for Web sites and access points (hosting), along with distribution of information/data.

  It is currently unknown, however, whether a Cloud-Centric architecture can meet EROS mission objectives optimally. What is known is that the initial transition and any other transitions between cloud providers will likely be cost prohibitive, and that long-term cloud costs are variable and hard to predict.

8.4. Applied Success Measures and Architecture Recommendation

The following subsection describes the process used to select a single architecture recommendation, along with the risk assessments and results.

8.4.1. Assessment Methodology

As described in section 8.3, the architecture study team captured and documented observations and assessments of each architecture alternative, along with input from subject matter experts. The measures of success were applied to help identify differentiators for each architecture option. At this point, the team applied the measures of success, compared and ranked architecture options and evaluated each option based on alternate weightings of the success measures resulting in a comparative ranking of all options.

The resultant ranking was risk-informed by considering factors that were outside of the key measures of success, with particular emphasis on whether an alternative is acceptable to the EROS culture. Throughout the ranking process, the team referenced the challenge statement to inform the evaluation.

8.4.2. Integrated Metrics and Business Objectives

Following the definitions of architecture concepts, the study team integrated the given success measures along with the business objectives of capability, risk, and cost (fig. 8–6). This integration served as the basis for comparative architecture rankings and architecture selection from the four concepts considered.

	Capability		Risk		Cost
	Effectiveness	Flexibility	Reliability	Security	Sustainability
Data Flow	MOP	MOE	MOP	MOE	MOP/E
Data Users	MOE	MOE	MOP	MOE	MOE
System Architecture	MOP	MOE	MOP	MOE	MOP/E

Measures of Performance (MOP) = Quantitative
Measures of Effectiveness (MOE) = Qualitative

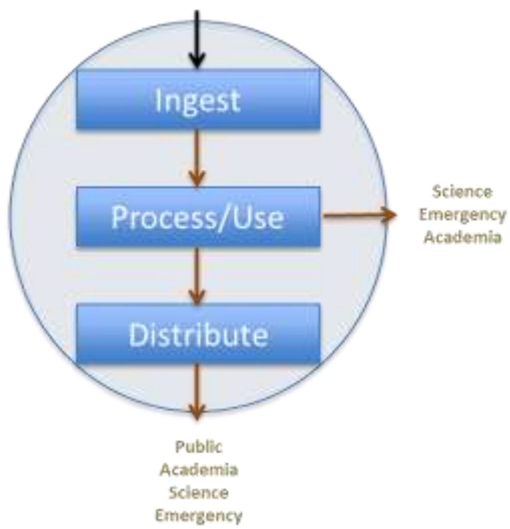


Figure 8–6. Integrated Metrics and Business Objectives.

8.4.3. Decision Matrix Scoring

Table 8–1 represents the evaluation results for the decision matrix scoring.

Table 8–1. Decision Matrix.

	Capability		*Risk*		*Cost*		***Total***
	Effectiveness	**Flexibility**	**Reliability**	**Security**	**Sustainability**	**Transition**	
Weight							
As-Is	1	1	1	1.5	1	4	9.5
Projectized	2	2	2	1.5	2	3	12.5
Enterprise	4	3.5	3	3.5	4	2	20
Cloud Centric	3	3.5	4	3.5	3	1	18

In this way, the study team effectively married the business objectives to success measures plus consideration of transition complexity (that is, acceptability in the EROS culture).

From this point the team weighted each option by numerically varying the emphasis between capability, risk, and cost. This resulted in six different combinations to better view the resulting comparative ranking. Figure 8–7 represents this view along with the “no weight” results.

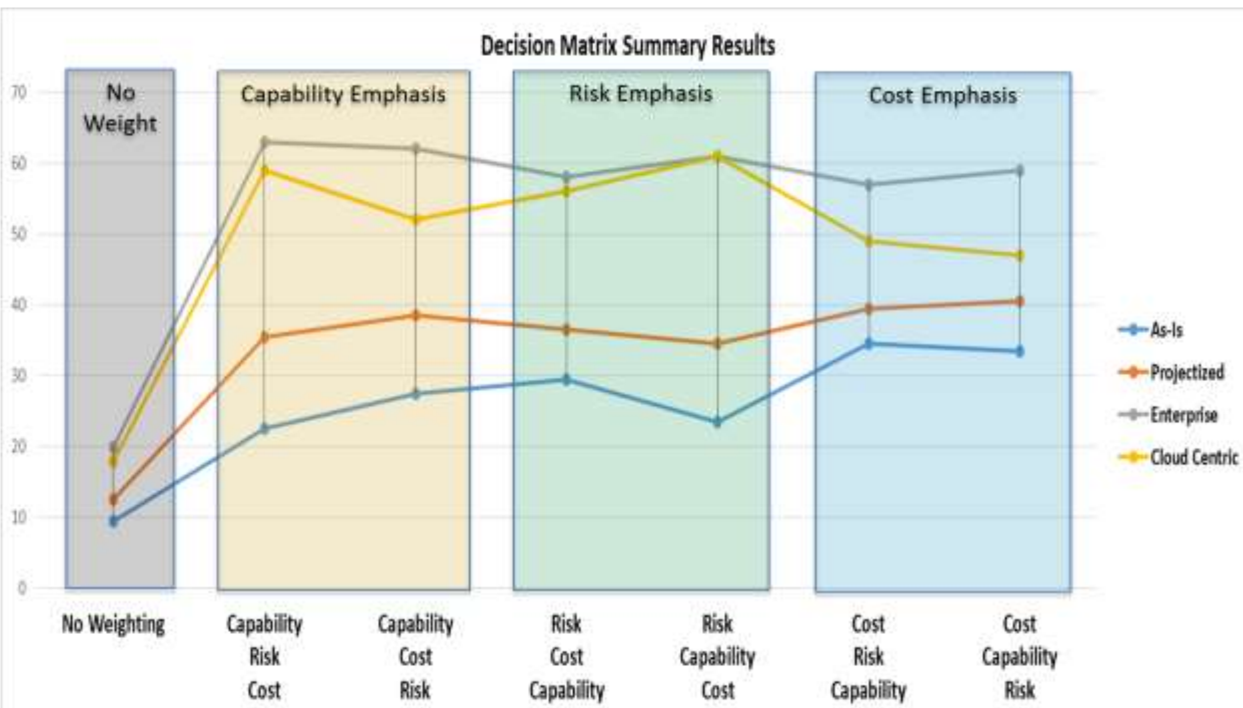


Figure 8–7. Decision Matrix Summary Results.

Figure 8–8 represents the study team’s key takeaways. Of note, the As-Is option is the lowest in all categories; the Projectized architecture is slightly improved, but not substantially; Cloud-Centric represents high performance, but comes with substantial risk and cost variability; and finally the Enterprise was ranked the highest in all categories.

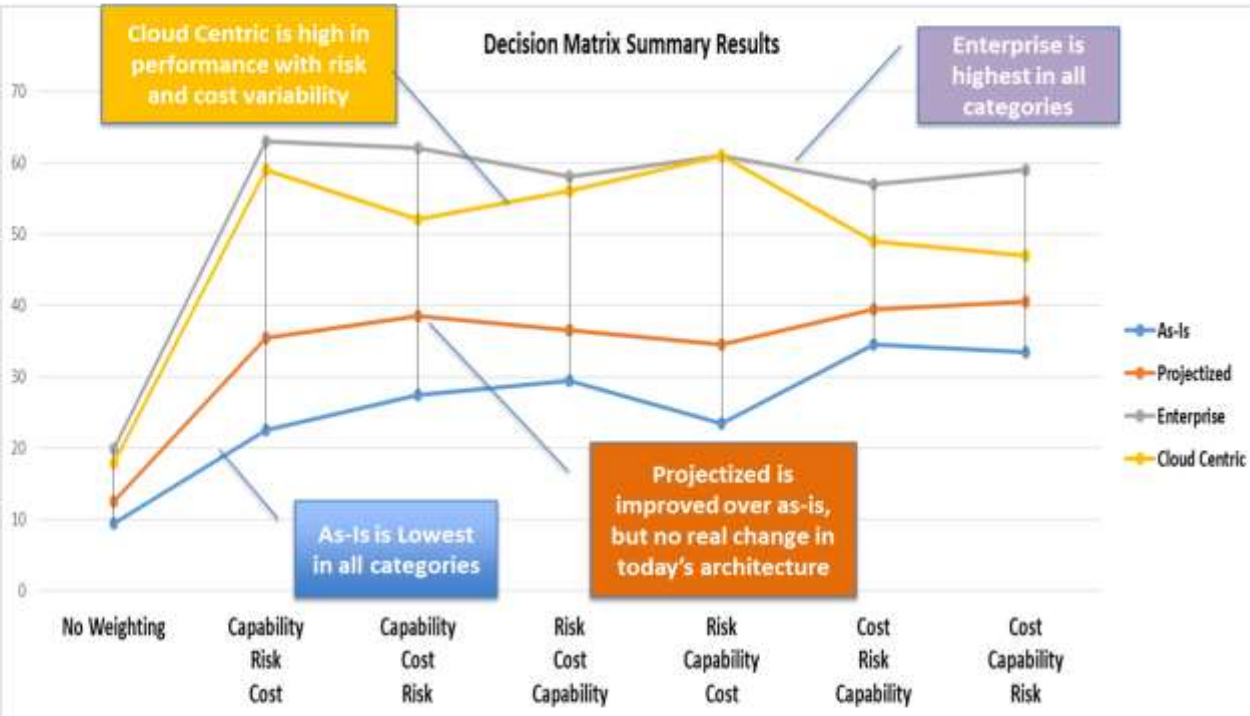
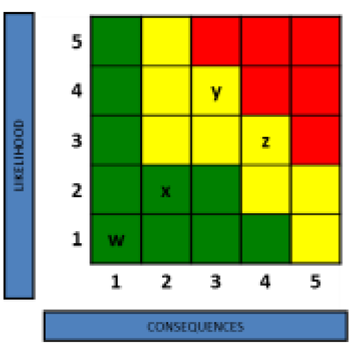


Figure 8–8. Decision Matrix Summary Observations.

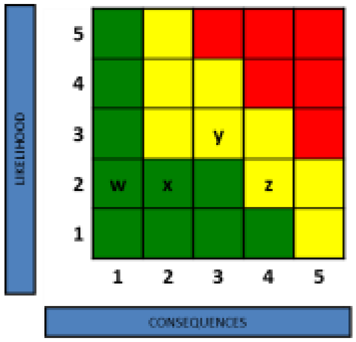
8.4.4. Risk Assessments

The resultant ranking was risk-informed by considering factors that were outside of the key measures of success, with particular emphasis on whether the architecture is acceptable to the EROS culture. The final risk assessments for each architecture alternative are shown in the following four graphics (figs. 8–9, 8–10, 8–11, and 8–12).



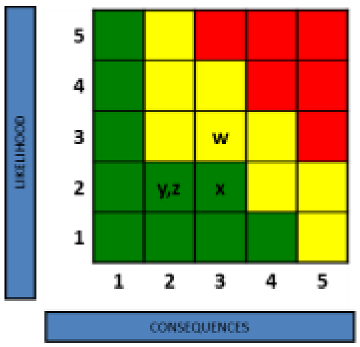
Trend	Rank	ID	Approach	Risk Title
	1	w		Culturally Implementable
	2	x		Stakeholder and User impact
	3	y		Enables New Science (e.g. LCMAP)
	4	z		Supports New Missions/Data Sets

Figure 8–9. As-Is architecture risk assessment.



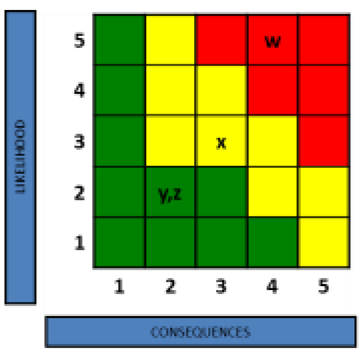
Trend	Rank	ID	Approach	Risk Title
	1	w		Culturally Implementable
	2	x		Stakeholder and User impact
	3	y		Enables New Science (e.g. LCMAP)
	4	z		Supports New Missions/Data Sets

Figure 8–10. Projectized Matrix architecture risk assessment.



Trend	Rank	ID	Approach	Risk Title
	1	w		Culturally Implementable
	2	x		Stakeholder and User impact
	3	y		Enables New Science (e.g. LCMAP)
	4	z		Supports New Missions/Data Sets

Figure 8–11. Enterprise architecture risk assessment.



Trend	Rank	ID	Approach	Risk Title
	1	w		Culturally Implementable
	2	x		Stakeholder and User impact
	3	y		Enables New Science (e.g. LCMAP)
	4	z		Supports New Missions/Data Sets

Figure 8–12. Cloud-Centric architecture risk assessment.

8.5. Summary Recommendation

At the conclusion of Phase 2, the study team recommended pursuit of the Enterprise Architecture alternative for Phase 3 assessment. This architecture option ranked the highest among the measures of success, was determined to be the lowest risk alternative, and best met the EAST challenge statement.

During checkpoint no. 2 with the steering committee and the study team's sponsor, all agreed with the team's recommendation. Direction was given at this time for the team to proceed to Phase 3 of the study for this architecture alternative.

9. EROS 2021 Vision – Enterprise Architecture

Following the selection of the Enterprise Architecture alternative at the end of Phase 2, the EAST was given a target date of 2021 for the completion of the transition to the target architecture. The following section describes the vision the study team constructed for the EROS enterprise architecture in 2021.

9.1. Information Model

The Information Model represents the structure of EROS's logical and physical data assets and data management resources. From the aspect of the architecture, a core component is understanding the data (information) lifecycle. Simply put, this is an essential perspective for developing a comprehensive architecture that ultimately meets the operational and performance requirements for the whole system. Additionally, an effective information model addresses the following:

- Aides in understanding the associated software, system, and service needs to support the entire lifecycle of the data.
- Includes existing and developing capabilities all the way to analysis and extracted understanding (for example, information).
- Describes the lifecycle of the data from collection to science use.
- Supports the required performance of the system from discovery to scalability and distribution.
- Ensures data provenance and integrity to support the business needs of the enterprise.
- Ensures that data are reliably managed, supportive of discovery, and fully used.

Done correctly, formulation of this type of information model takes a considerable amount of time and effort, and thus was not completed by the architecture study team. The team did, however, formulate a notional information model construct. The following components were identified by the study team:

- **Data Sources**: Encompasses the entire suite of Center data holdings along with attributed metadata and reference information.
- **Products**: Encompasses the entire suite of products and information types developed through the EROS architecture.
- **Query and Access**: Describes the discovery, distribution, and visualization services required.
- **Data Formats**: Describes the data format types supported by the enterprise architecture.
- **Data Management Planning**: Encompasses the policies, standards, and practices for backup, archive, security, and any other institutional services or requirements.

Given these component definitions, figure 9–1 represents a notional, high-level information model for EROS.

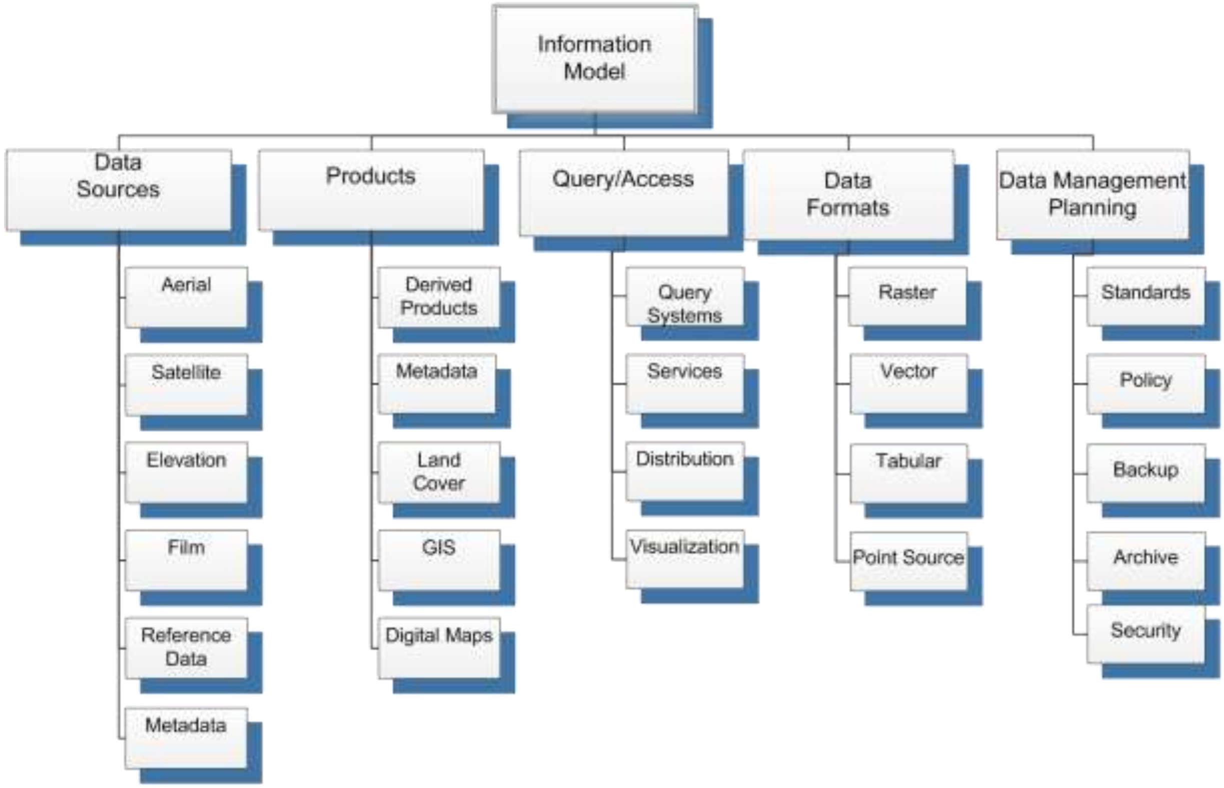


Figure 9–1. Notional EROS Information Model.

Build-out of the Information Model would be complex, especially for the existing and future data and information lifecycles within the EROS Center holdings. Figure 9–2 highlights just two further derivations of satellite and metadata sources.

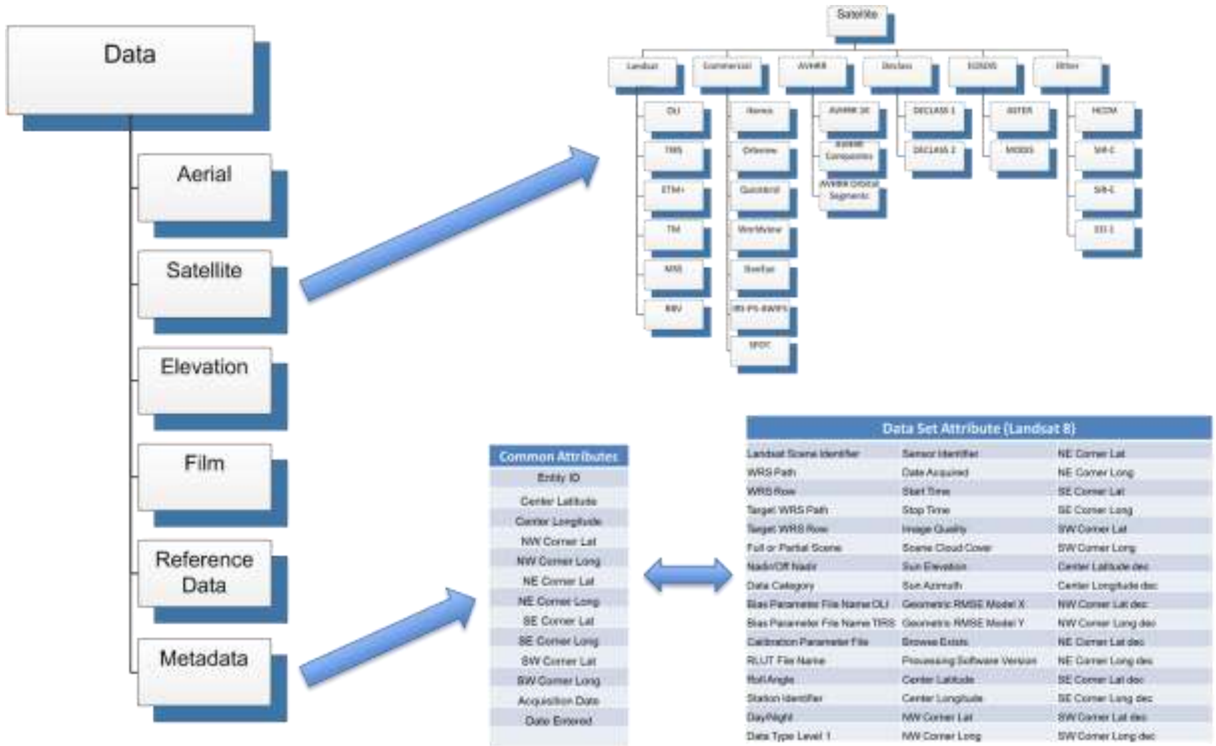


Figure 9–2. Example Information Model Build-out.

To fully capture, describe, and implement the EROS Information Model, a follow-on activity is required in association with enterprise architecture implementation. This model is integral to capturing the data and information lifecycles at EROS and represents an essential perspective and strategic planning tool for developing a comprehensive architecture, and ultimately operational system that meets performance requirements.

Done correctly, the effort required to complete this model can be substantial. The efforts by JPL to develop a Data (Information) Model for their Planetary Data System holdings spanned more than one year. The EAST recommends that a small tiger team be formed to develop the information model in parallel with Business Model refinement and implementation as a follow-on activity.

9.2. Technology and Applications Architectures

As defined by the study team, the Technology Architecture represents the infrastructure, computing and storage capabilities (for example, technology and system interfaces) of the system; and the Applications Architecture represents the applications applied in support of the EROS mission (for example, data flow and end use products).

The study team's 2021 vision for the technology architecture (see fig. 9–3) consists of shared infrastructure and enterprise services including storage, compute, and virtual desktop environments, along with ready access to trusted partner and commercial cloud capacity storage and on-demand, 'elastic' science computing resources.

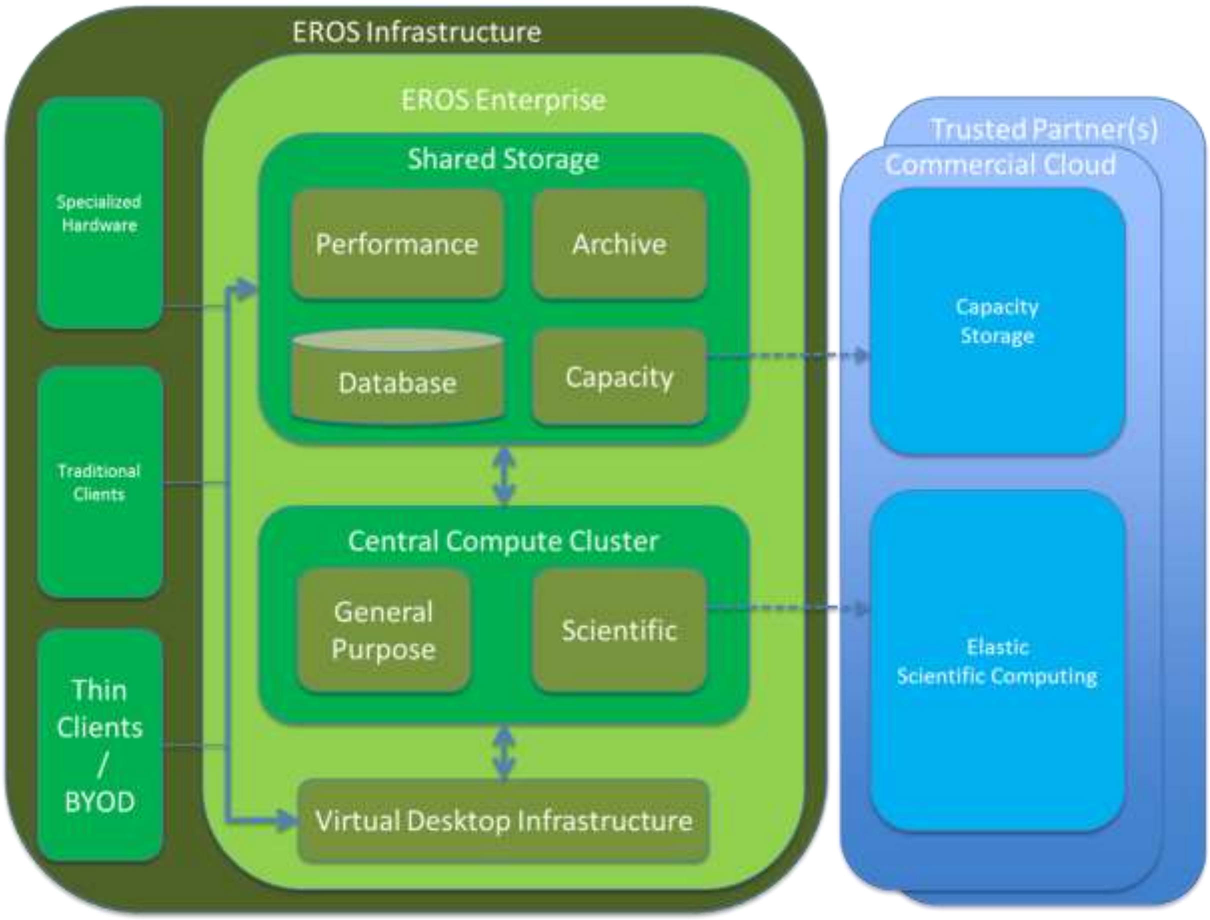


Figure 9–3 EROS 2021 Technology View.

The subsequent applications view (fig. 9–4) is represented by the following and shows the vision for shared resources spanning infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and database as a service (DBaaS).

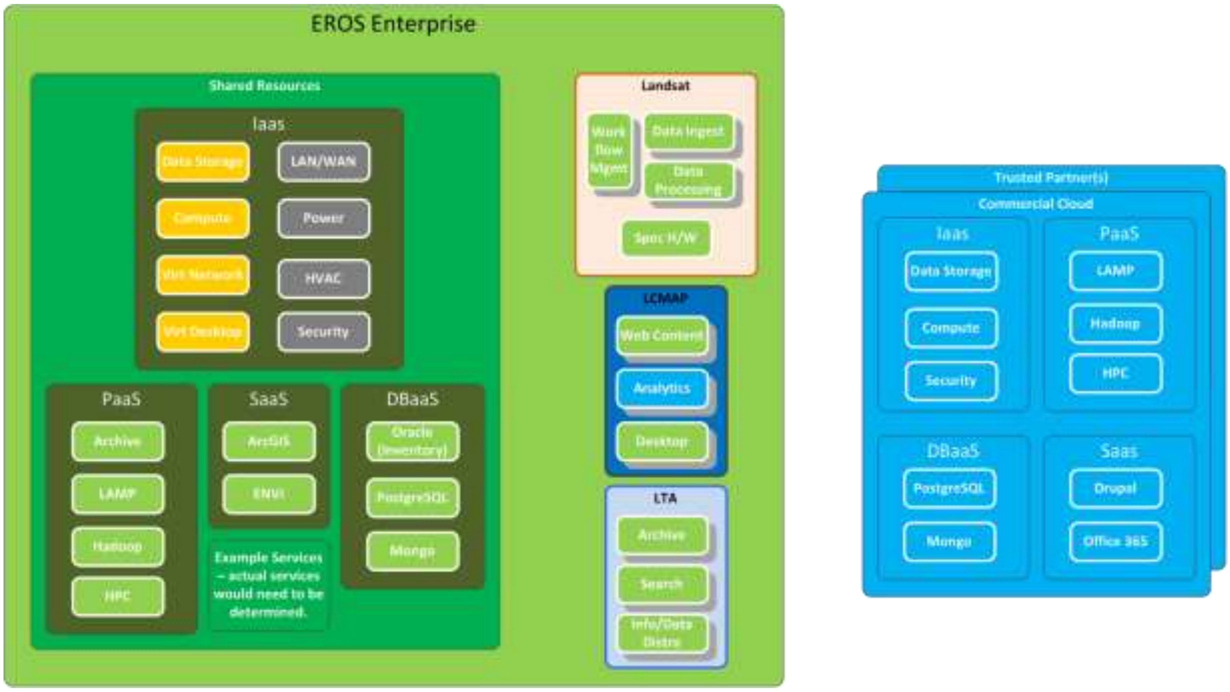


Figure 9–4. EROS 2021 Application View.

The use of service tiers should result in effective implementation of the technology and application aspects of the enterprise architecture. The study team defined each according to the following:

- IaaS (Infrastructure as a Service)
    - Refers to basic infrastructure (compute, storage, networking, and so on).
    - Virtualization, bare-metal provisioning, and other more sophisticated packages like OpenStack are examples of this.

- PaaS (Platform as a Service)
    - Refers to frameworks and toolsets that provide an environment for application development, deployment, and execution.
    - Examples of this would be Cloud Foundry, Hadoop, Amazon Electric Beanstalk.

- SaaS (Software as a Service)
    - Refers to the provision of specific hosted applications that are made available to users over the network (as opposed to users installing them locally).
    - There are several examples of this are relevant to the EROS architecture (for example, geographic information system (GIS) tools).

- IHAS (In-House Applications/Services)
    - Systems and services that have been developed at EROS.
    - In-house development can implement lower tiers of the service stack.

The business and information models govern how these tiers function, how services are provisioned, and how they relate to each other. The relation between the framework and the described service tiers is represented in figure 9–5.

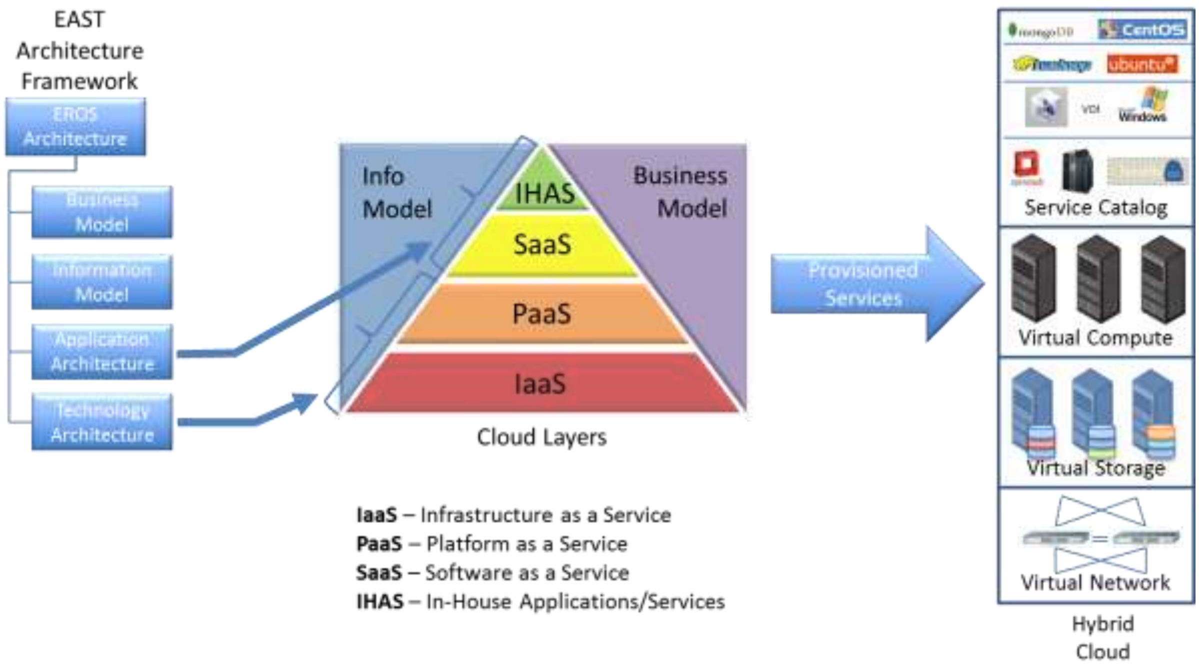


Figure 9–5. Enterprise Service Tiers.

Finally, it is necessary to consider both the physical and logical representation of the envisioned architecture (fig. 9–6). Importantly, fundamental components of the physical design of the EROS enterprise architecture are redundancy and reliability. To achieve the expectations of these success measures, data and computing capabilities are replicated across multiple physical locations (for example, computer rooms). This means that equipment failures in one location (or even the loss of an entire computer room) would not prevent normal EROS operations from proceeding normally, although perhaps with temporarily degraded performance. If done correctly, this replication is invisible to the projects and users of the services. Additionally, the planned deployment of a mesh network is absolutely necessary and will result in much higher aggregate network throughput and reliability (that is, no more single points of failure). Finally, secure connections are used to connect the EROS infrastructure with cloud vendors and external partner facilities.

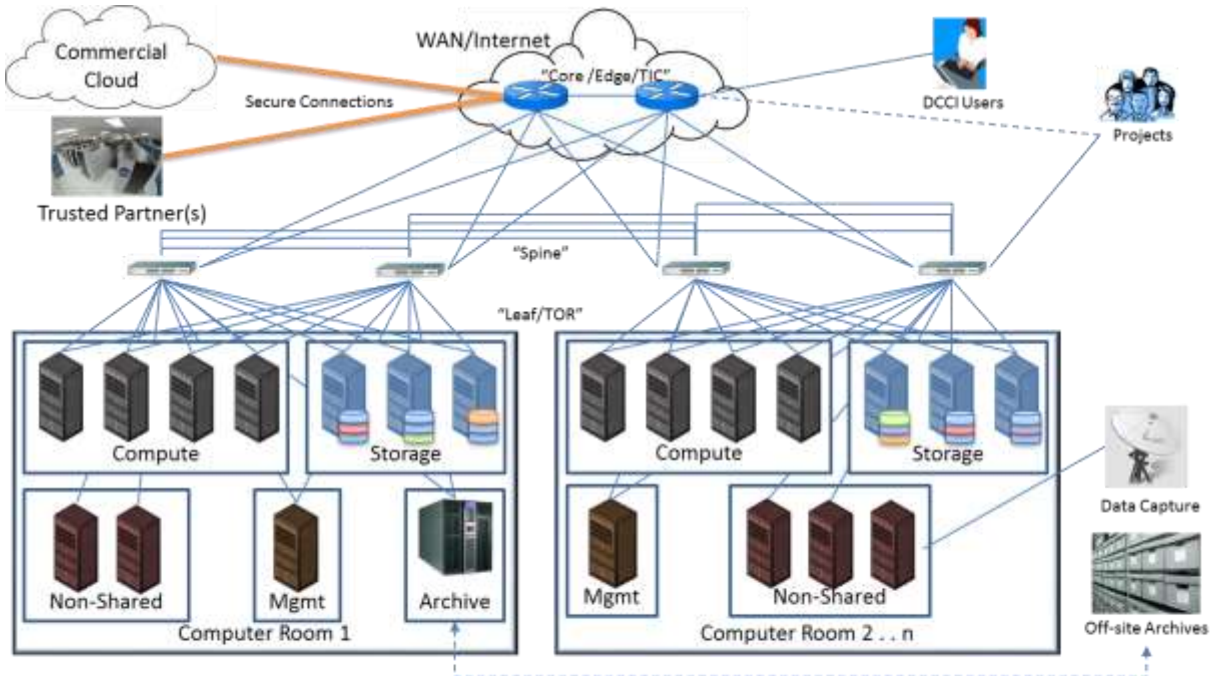


Figure 9–6. EROS 2021—Physical Architecture View.

With respect to the logical architecture view (fig. 9–7), the EROS enterprise architecture is implemented as a hybrid cloud. In this model, EROS will provide and manage many resources in-house, but have elastic storage and compute services as an example provided externally. Furthermore, most resources inside EROS are part of the shared compute and storage infrastructure, but there will always be nonshared or project-specific resources as circumstances demand based on project requirements. In this model, managed provisioning will allow projects to easily expand processing efforts using trusted partners or commercial cloud resources as available. Finally, the IT Business Office (described in section 9.3 Business Model) maintains control over the EROS infrastructure and tiers of service, whereas projects maintain control over the resources that they are provisioned.

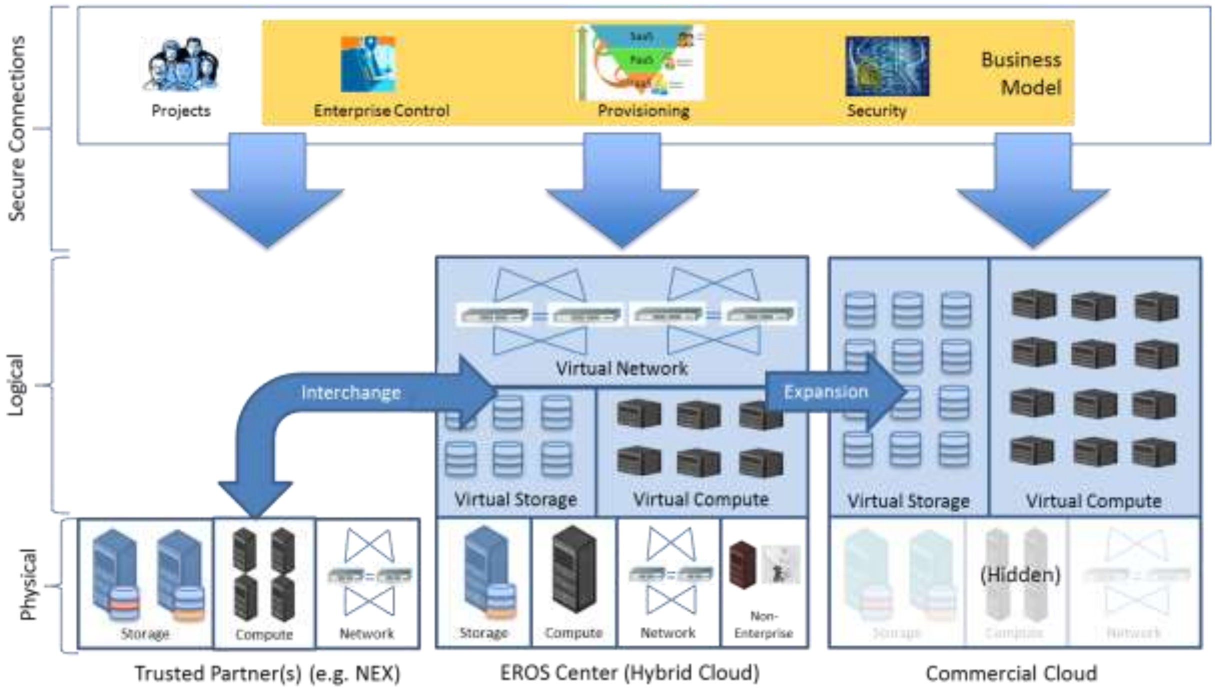


Figure 9–7. EROS 2021 – Logical Architecture View.

9.3. Business Model

The Business Model represents the business strategy, governance, organization, and key processes of the EROS Center. As such, the Business Model encompasses the activities needed to oversee and govern use of the proposed enterprise architecture. These activities include the following:

- **Governance:** encompasses roles and responsibilities, processes, standards, and exceptions.
- **Strategic Management:** encompasses public, private, and Federal partnerships, portfolio planning, investment management, and enterprise architecture strategic planning.
- **Service Management:** encompasses the service catalog, provisioning of shared resources, and charge back mechanisms for provisioned services.
- **Performance Management:** encompasses management and execution of service level agreements, application and measurement of metrics, and capacity management.
- **Change Management:** encompasses configuration and asset management.
- **Risk Management:** encompasses security and knowledge management.

The study team's vision for the EROS 2021 enterprise IT business model (fig. 9–8) comprises four key organizational components that include the following:

- Investment Review Board,
- Architecture Review Board,
- Projects, and
- Center IT Team.

Governance, strategic management, and risk management are addressed by these two governing boards; Investment Review Board (IRB) and Architecture Review Board (ARB).

The IRB is made up of EROS senior staff and is primarily responsible for oversight of the enterprise architecture's strategic management and risk management. As such, this board authorizes new business, provides strategic guidance, arbitrates ARB disagreements, and provides EROS Center portfolio management.

The ARB is made up of the project managers and CITT and is facilitated by the CITT Enterprise Architect. The ARB shares in risk management and is primarily responsible for the governance associated with the IT enterprise. As such, this board considers and approves new technology requests, provides architecture oversight, approves IT investment strategies and IT business office recommendations, and finally ensures the enterprise IT services implementation meets customer (project management) requirements.

Projects; Center IT Team (CITT)—Execution of the business model (fig. 9–9) is dependent on actual projects and the CITT.

Projects are directly responsible for their business requirements, role within the information architecture, software and in-house applications development, project risk management, stakeholder coordination, and finally identifying their requirements.

Other key execution tasks fall to CITT. These tasks include Center-wide risk management, management of Center-wide information technology, the IT business office, and overall operations and maintenance of the enterprise architecture.

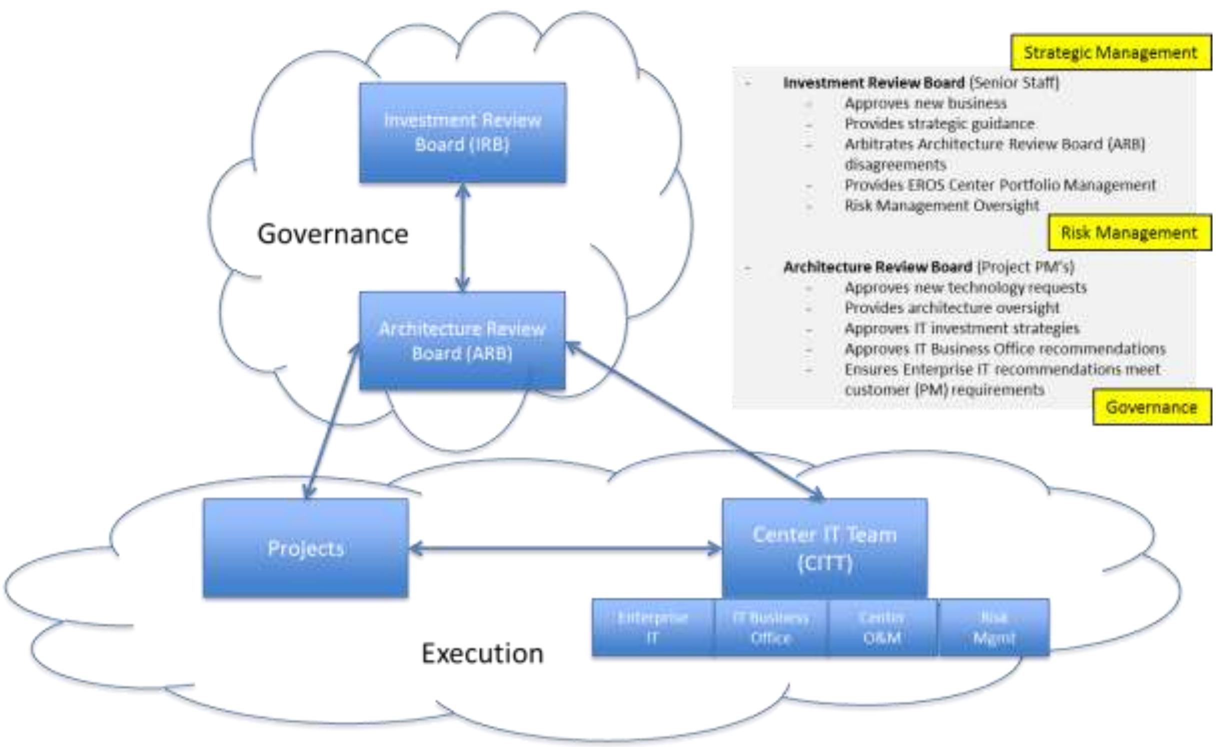


Figure 9–8. Enterprise IT Structure.

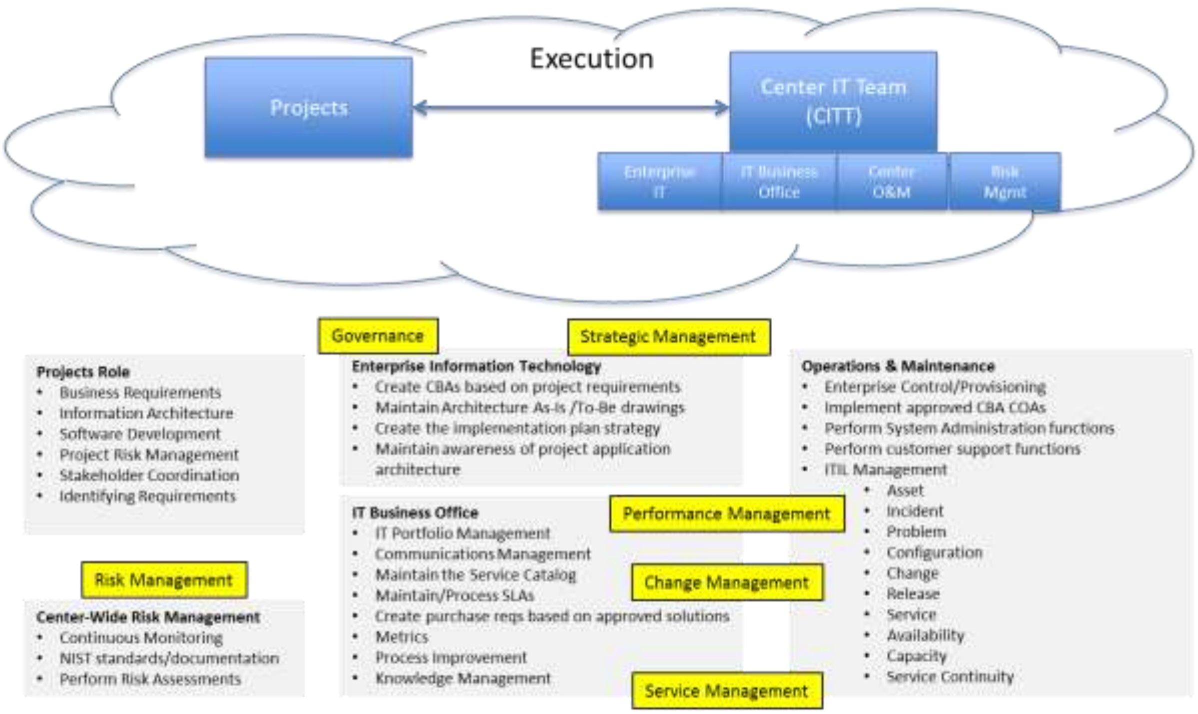


Figure 9–9. Enterprise IT Execution.

The key to a successful business model is agile and fast interactions (fig. 9–10) between the execution organizations (projects and CITT), along with communication with the investment and architecture review boards. To achieve this, CITT becomes the IT servicing organization for the Center. This includes serving as a focal point for provisioning shared services and facilitating potential waivers through the ARB. Key to this is the service catalog. The service catalog is what the projects use to determine and select needed capabilities. CITT must keep the service catalog up to date with available services and work with the ARB and IRB to incorporate new services or additional capacity. The central pivot point for this interactive process is the Service Level Agreement (SLA). All projects will have an SLA with CITT. Formulation of an SLA is envisioned to be nimble, efficient, and Web enabled, and will address needed capability and performance, along with the associated cost. Finally, CITT must interact with the ARB on a regular basis to advise the ARB on risks, update status on the IT business (for example, performance, capacity, and cost metrics, along with configuration and asset change management), technology trades, and operational status.

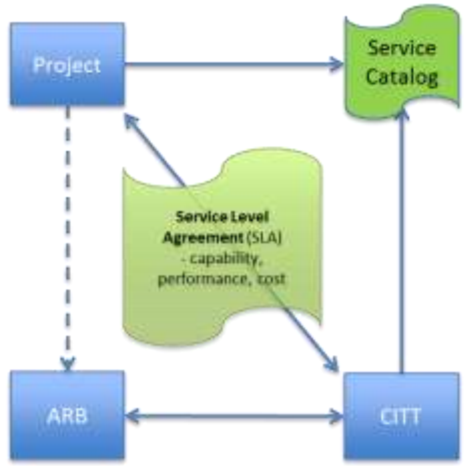


Figure 9–10. Enterprise IT Interactions.

The overall goal of this business model construct is to vastly reduce the number of activities today's project managers must concern themselves with. Today's project manager must consider the full gamut of IT business, including technology trends, IT security, capacity planning, procurements, system administration, technical support, and other IT investments. The project manager in 2021 will be much better positioned to focus on those activities that are most important to the project, such as science and business needs, requirements, stakeholder interactions, and development or operations.

9.4. Notional Implementation Approach

The study team weighed a number of the use cases and considered a few different examples of how the envisioned business model will work once implemented. The first example considered and assessed was that of project needs satisfied by the service catalog (see fig. 9–11). In this case, a given project communicates its needs that is, requirements) to the IT business office, whereby the IT business office confirms availability of services (that is, technology, applications, capacity, and performance). Once confirmed, the project selects the desired services; and an SLA is developed between the project and the IT business office capturing capability, performance, and cost. Once complete, the system is provisioned, and the capabilities are deployed for the project. Following completion of the SLA, this process should take a matter of minutes to a few hours to implement. Figure 9–11 provides additional detail.

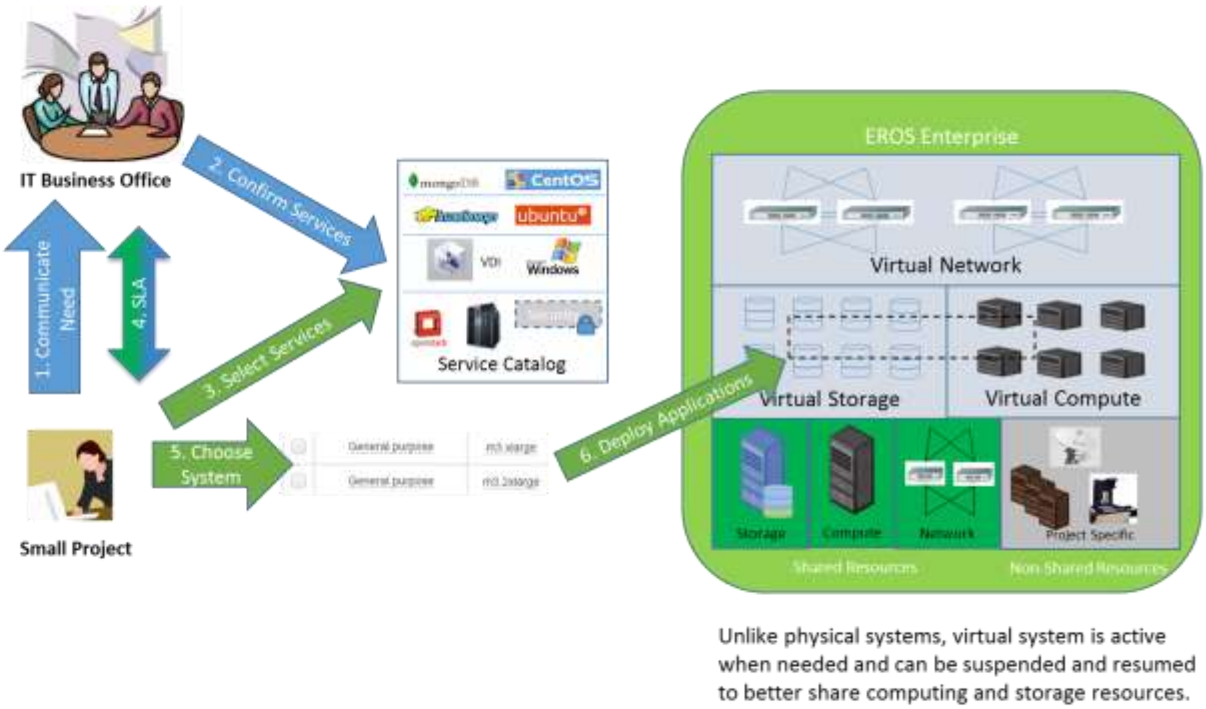


Figure 9–11. Example—Item in Service Catalog

The second example considered and assessed was that of project needs not available in the service catalog (fig. 9–12) and requiring ARB approval for implementation. In this case, needed services are added to the service catalog with ARB approval. In this case, the first two steps are the same. A given project communicates its needs (that is, requirements) to the IT business office, whereby the IT business office confirms availability of services. In this example, the needed technology, application, capacity, or performance is not available; therefore, the IT business office interacts with the ARB to seek approval for adding the needed service to the service catalog or pursues completion of a waiver for the project. Assuming approval is granted; the IT business office adds the needed service or procures the necessary technology and adds it to the service catalog. Once complete, the process is exactly the same as the first example. The project selects the desired services; and an SLA is developed between the project and the IT business office capturing capability, performance, and cost. Once complete, the system is provisioned, and the capabilities are deployed for the project.

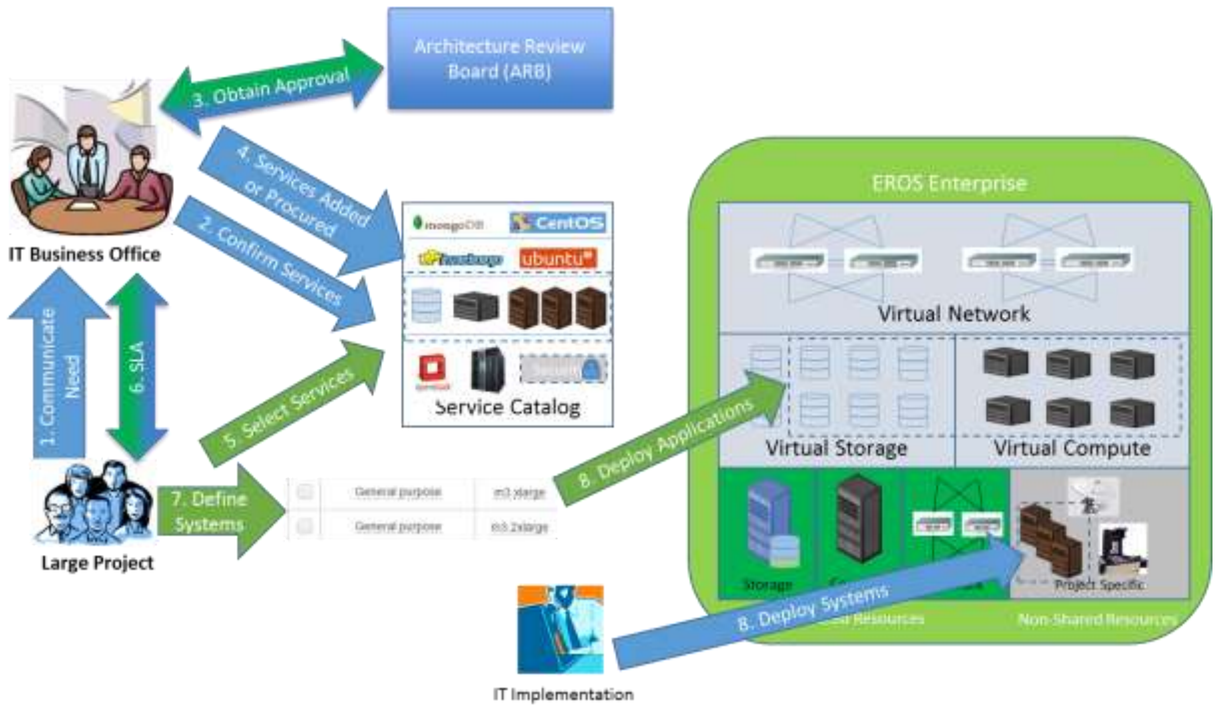


Figure 9–12. Example—Item not in Service Catalog.

9.5. Summary of Measures of Success Applied to 2021 Vision

Finally, the study team applied each of the prescribed success measures to the 2021 vision. The following improvements, findings, and attributes were noted by the team:

- **Effectiveness:** Enterprise computing and storage enable efficient use of shared resources; and availability of external resources allows for additional performance as needed.
- **Flexibility:** Total EROS infrastructure capacity can be seamlessly upgraded with no impact to projects; availability of external resources allows for additional performance as needed; and resource provisioning allows projects to scale up in days vs. months.
- **Sustainability:** Use of standards provides for increased sustainability of current and future data; service tier along with “living” catalog concept allows for introduction of new services with minimal impacts to existing projects; and standard configurations and refreshes/updates for infrastructure result in a more sustainable physical architecture.
- **Reliability:** Enhanced through access to highly available cloud resources and off-site partner resources; mesh network increases reliability; and design enables physical redundancy and failover between computer rooms.
- **Security:** Software-defined mesh network allows for more centralized policy-driven security rules, and standard configuration(s) for server images allows for more centralized security management.

9.6. EROS 2021 Vision Summary

The following key points were made by the study team in summary of the EROS 2021 enterprise architecture:

- An Information Model is integral to capturing the data and information lifecycles at EROS. It represents essential perspectives and strategic planning tools for developing a comprehensive architecture, and ultimately operational system that meets performance requirements. Finally, it uses common expectations and standards to provide increased sustainability of current and future data and information throughout the entire lifecycle.
- An effective Business Model that addresses each business element is crucial to successful implementation of a Center enterprise architecture to balance resources across projects, provide a living service catalog with simple and predictable cost model(s), enable architecture strategic planning and deployment, and allow project managers to focus on what's important—their stakeholders, requirements, development, and performance.

Projects interact with the EROS enterprise architecture, along with trusted partners and commercial cloud resources as required by their specific circumstances for effective, flexible, sustainable, reliable, and secure services and resources. Evolution to enterprise architecture enhances resource efficiencies and results in cost effective solutions across the Center's projects. This includes centralized security management, and enables effective use of external partners to augment Center capabilities as needed.

10. Roadmap

The road to an enterprise architecture state for the EROS Center evolution (fig. 10–1) will be evolutionary. The study team believes the combination of activities needed to achieve the 2021 vision will likely span 5 to 6 years of activity. Evolution to enterprise begins with implementation of centralized governance with early architecture activities focused on establishment of the needed infrastructure. Most existing projects will evolve from their As-Is state through the Projectized Matrix before fully becoming integrated into the Enterprise approach. Beginning in 2017, the study team anticipates new projects may begin within the Enterprise.

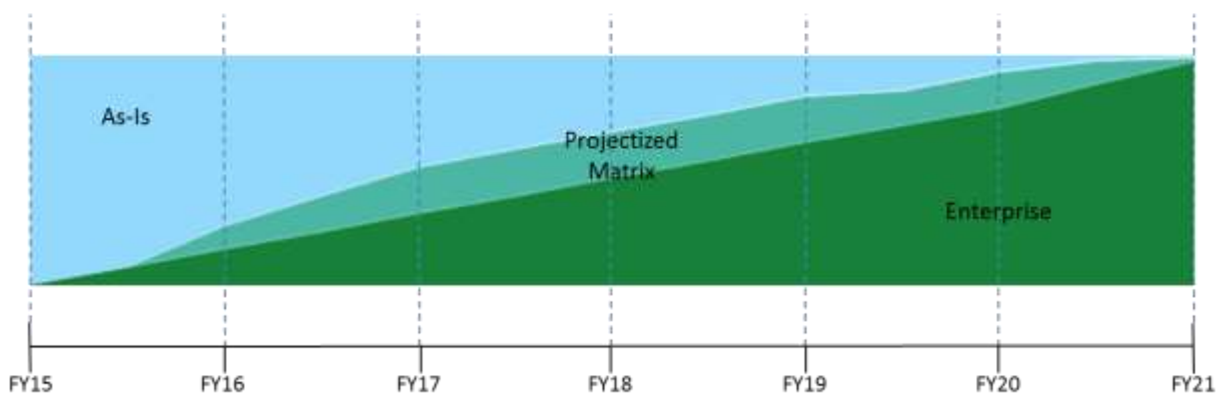


Figure 10–1. Enterprise Architecture Evolution.

Based on the EAST charter challenge statement and the direction to produce a roadmap that achieves an enterprise architecture implementation by 2021, the team produced a theme-based roadmap of the following fiscal year themes:

- **FY16: Enable the Enterprise:** Achieved through definition and implementation of business model and governance approaches.

- **FY17: Enable LCMAP:** Achieved through implementation of a business model, roll-out of initial services, and leveraging network infrastructure along with capacity, storage, and computing.
- **FY18: Enable Science Data Systems:** Achieved through metrics-based service expansion, LCMAP operating capability, and data format consolidation and standardization.
- **FY19: Enable LSDS (Landsat 9):** Achieved through expanded storage and computing services, access to partner organizations, and expanded network capability.
- **FY20: Seamless Integration with Trusted and External Partners:** Achieved through interface consolidation and standardization, near-autonomous provisioning of partner capabilities, and predictable cost models.

A high-level roadmap (fig. 10–2) was developed by the study team to incorporate and map Center milestones to enterprise architecture implementation activities. This high-level road map does not constitute implementation detail. Rather, it is intended to serve as a guideline toward implementation planning.

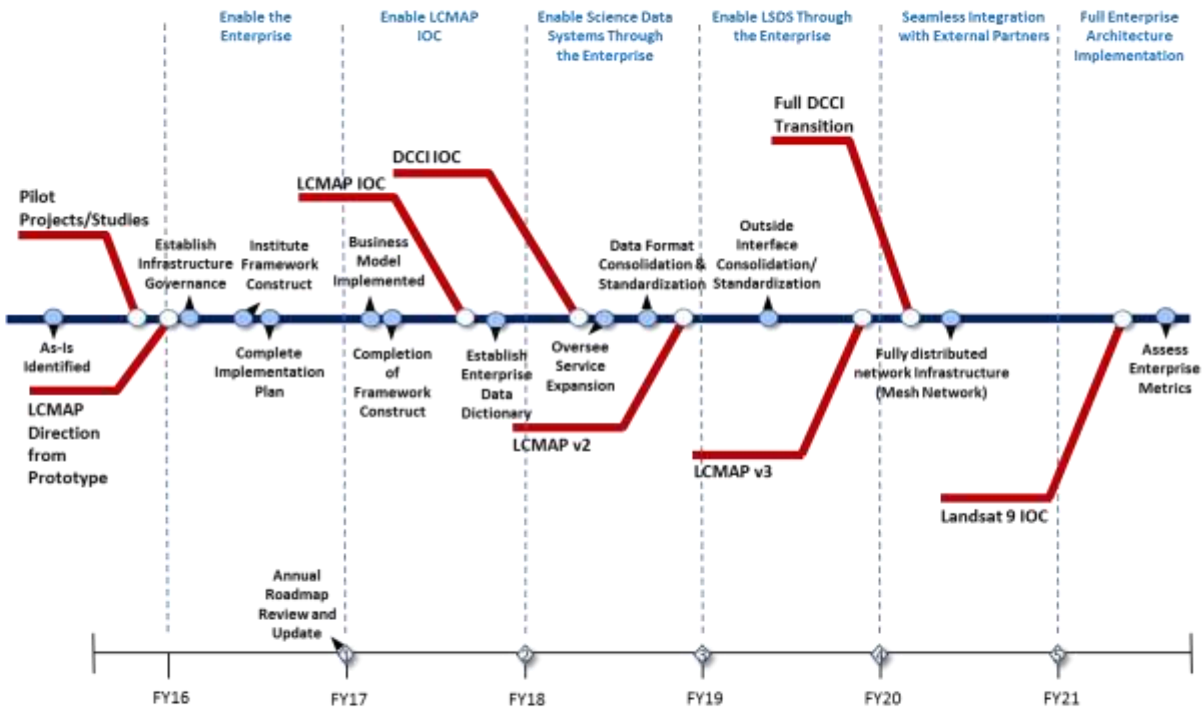


Figure 10–2. Enterprise Architecture Fishbone Diagram.

The study team also constructed a more detailed roadmap (fig. 10–3) that addresses the major activities within each architecture view. Consistent with the team's findings, the roadmap calls for early pursuit and implementation of an effective business model, along with definition of an information model and fielding of needed technology infrastructure in support of the enterprise architecture. Note that large Center projects will need to provide the initial backbone of the enterprise architecture, especially regarding technologies pertaining to capacity and performance storage and compute.

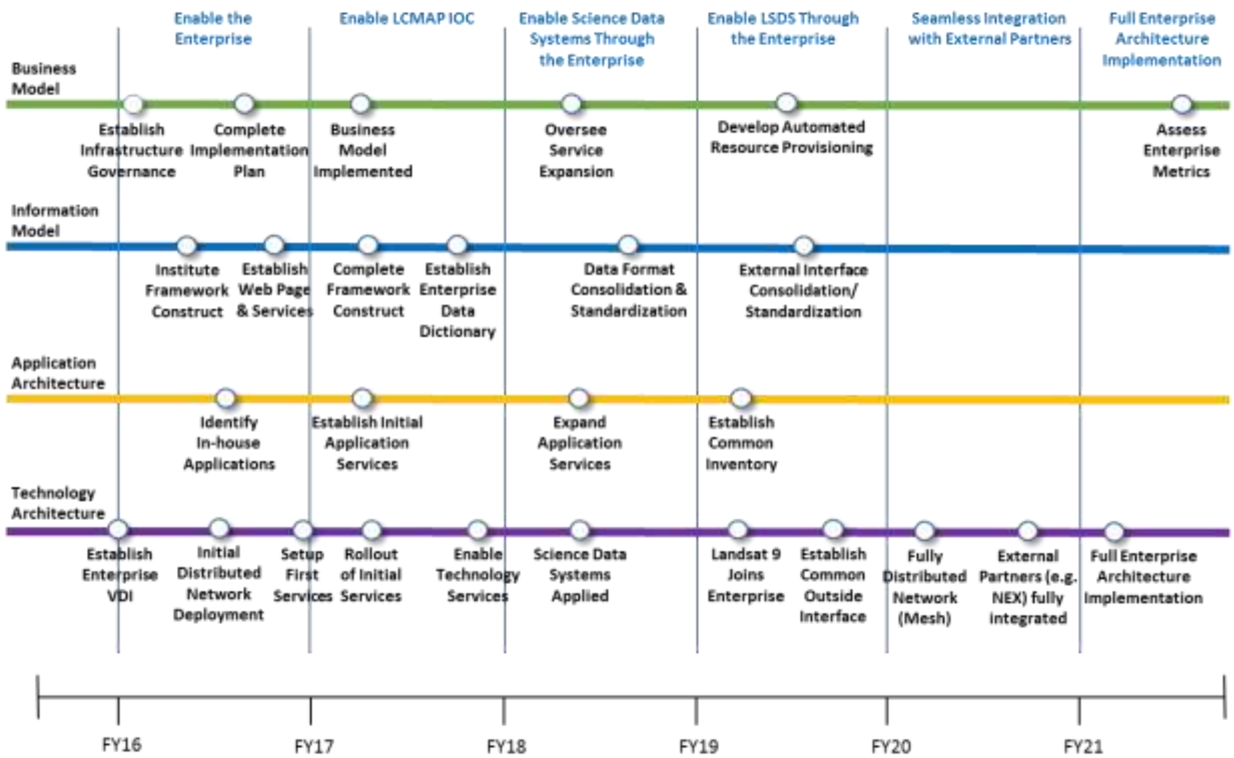


Figure 10–3. EROS Architecture Roadmap by View.

It is important to note that transition to an enterprise architecture must be evolutionary in nature. Even for this evolution to work in the timeframe noted, many activities will need to be run in parallel within each area of the architecture. Implementation planners will need to carefully consider the complementary nature of the approach along with dependencies to ensure all needed pieces are evolving at the correct pace for enterprise business, technology, applications, and information lifecycles to be used in an effective manner.

Additionally, regular (at least annual) review and update of the roadmap will be necessary for it to remain current with implementation plans and actual events. Also, the roadmap may need to be updated to align with EROS strategic plan milestones.

Finally, an implementation plan outlining the approach, roles and responsibilities, and methods for achieving the enterprise architecture will need to be generated as a follow-on activity to flesh out the details of the roadmap.

11. Overall Study Observations and Recommendations

The study team made three primary observations concerning the architecture study activities:

1) **EROS is currently effective as a collection of semi-independent projects, but inefficient at a Center level.** In many cases projects are very efficient within their project boundaries and in some cases effectively share experiences and capabilities with other projects. However, the way EROS operates today is inefficient from a Center perspective. There is limited overarching governance and strategic planning from a technology, applications, or shared services and capability perspective. In other words, sharing of information and resources across projects does exist, but it is ad-hoc. Over time,

EROS has become fragmented along project boundaries, with larger projects having greater capabilities and better technology than smaller projects. Furthermore, it took considerable effort for the EAST to determine, much less understand, the overall Center architecture, attributes of the information model, and user communities. There currently exists a substantial lack of documentation citing these important attributes. Finally, the As-Is Center business model is not conducive to supporting EROS strategic architecture evolution. Simply put, opportunities for strategic planning with the EROS As-Is architecture environment are limited.

2) **An EROS Enterprise architecture approach could provide many benefits.** Center-level system of systems view of EROS architecture enables effective strategic planning and a centralized security responsibility and capability improves the overall security posture. Additionally, virtualization and other enterprise services help accommodate various types of projects, from small science projects to large projects like LTA and Landsat, to new project endeavors like LCMAP. Finally, the proposed hybrid cloud approach allows EROS to more efficiently share internal resources and take advantage of rapidly evolving commercial and private partnership cloud capabilities.

3) **Transitioning from As-Is to a Target architecture is an evolution.** Business model, information model, and infrastructure changes are needed early on to accommodate eventual transition to an enterprise architecture. The business model must be agile, efficient, and cost-effective for projects and the Center to realize cost benefits. Finally, regular updates to an architecture roadmap and implementation plan are crucial to a successful transition.

The study team made a number of key recommendations for the follow-on architecture implementation planning and execution, including the following:

1) Develop an enterprise architecture Implementation Plan based on the roadmap.
2) Establish an EROS enterprise information model.
3) Fully develop and implement a Center-wide business model. In doing so, identify appropriate organizational changes and secure appropriate resources commensurate with what is needed to achieve the implementation objectives and timeline, establish needed enterprise governing bodies (IRB and ARB), and continue development of the service catalog and associated metrics as well as an agile and nimble SLA process.
4) Maintain As-Is Architecture and User Characterization documentation going forward.
5) Revisit architecture roadmap and implementation plan on an annual basis and update as needed.
6) Transition to a target architecture in an evolutionary compared to a revolutionary way. Build on existing strengths within projects to form the backbone and take into account EROS culture during the transition.
7) Build on interagency partnerships established during EAST in implementation. All partners that participated in the EAST are interested in continuing to work with EROS to learn from each other as implementation of an enterprise solution is pursued.

Appendix 1: Independent Comparative Cost Study Summary

The following sections describe the independent and parallel comparative cost study conducted by The Aerospace Corporation on behalf of the EAST. Use of the term “current” during this section refers to the time of the writing of this report (2015).

A.1 Background

The objectives of the comparative cost analysis were to assess EROS As-Is architecture costs, generate rough order of magnitude (ROM) costs for to-be recommendations, and identify cost risks and opportunities. The analysis focused first on the costs of the current EROS way of doing business, then broadened to examine costs of alternative ways of doing business using LCMAP as the use case.

A.2 Estimating Methodology and Philosophy

Firstly, work breakdown structure (WBS) was developed to divide Center expenditures into categories of labor and materials consistent with existing cost estimating relationships (CERs) and cost models.

Next, to establish the costs of the current way of doing business, four baselines were created. The baselines, in turn, were used to estimate future costs should the current business model be continued. The four baselines included the following:

1. Hardware baseline, which included the purchase prices of the current active storage, processing, computing, and network equipment inventory.
    - Purchase prices were normalized to the cost of today’s technology in today’s dollars using Federal Reserve Economic Database IT equipment indices.
2. Labor and contracts baseline, which parsed the 2014 Budget into the estimating WBS to categorize costs by function.
    - Labor costs were then inflated annually through 2021 (assumed no changes to the staffing plan).
    - Used to generate an annual total Center-level cost estimate for the As-Is architecture through 2021.
3. Volume baseline, which included the following:
    - The current actual total data storage capacity in-house (tape as compared to disk) was collected, then used to develop EROS metrics ($/terabyte (TB) based on the costs of tape (hardware and media) and disk in the hardware cost baseline.
    - The analyst then created a data-storage volume forecast for the Center through 2021, including Sentinel–2 data estimates.
    - The tape and disk $/TB metrics were then paired with the forecasted cumulative volume to form the estimated As-Is hardware cost through 2021, assuming no architecture changes are made.
    - These hardware estimates were then inserted into the estimating WBS to indicate areas where labor skill mix can shift as a result of reducing the cost and complexity of the architecture.

4. Processing baseline, this was used to compare the estimated processing cost of LCMAP in alternative architectural constructs.

To generate the baselines listed above and their constituent elements, a large amount of raw data was collected from official Government sources and databases at the Center. A listing of all data sources was provided in the cost analysis out-briefing delivered under separate cover at checkpoint no. 3 on July 10, 2015.

It is known that processing and storage costs are decreasing over time as a result of capability increases. The ROM As-Is hardware estimates incorporated industry trends, and the estimate outputs were compared with industry forecasts. Disk and tape were analyzed separately. Tape capacity is increasing at 30 percent per year, so when the cost of tapes is held steady, the cost per terabyte to add additional tape storage to accommodate the ever-increasing data volume at the Center drops annually. Similarly, disk capacity is increasing, resulting in a doubling or tripling of capacity with each new technology release, which historically occurs roughly every 3 years.

A.3 Baselining and As-Is Estimating Results

Baseline costs, ROM estimates, and metrics are not presented in the final report because of sensitivity and distribution limitations. These items were presented in the cost analysis out-briefing delivered under separate cover at checkpoint no. 3 on July 10, 2015.

As-Is Architecture Cost Findings

The current EROS disk storage ($/TB) metric is roughly 20 percent higher than the industry average. However, when the trend is examined over time through 2021, it becomes apparent that EROS metrics are in line with industry, just at a 1-year lag. The current EROS tape storage $/TB metric is roughly 30 percent lower than the industry average. EROS can expect to realize annualized capacity gains on par with industry forecasts, driving down the amount of hardware and media required.

As the quantity of hardware end items and tape media reduces over time because of capacity gains, the maintenance labor costs relative to hardware costs should be adjusted downward. Similarly, other direct costs, such as training and documentation, should also lessen as a percent of hardware expenditures. These are examples of cost opportunities that EROS can expect as a result of capacity improvements in storage systems and media, should EROS choose to meet increasing data ingest volume needs in-house.

Similarly, the layer of program management, systems engineering, and integration and test (PMSEIT) specifically associated with data center hardware (CITT function) will also grow as a percent of hardware costs as a result of needing less hardware to store more data. This can be viewed as a beneficial side-effect, as enterprise management will require a strengthening of this function.

To-Be Architecture Component Comparative Analysis

In support of the EAST's exploration of new ways of doing business, the cost analysis compared the cost of storing and processing LCMAP data in-house versus using the public cloud (for example, Amazon Web Services, Google, Microsoft Azure, and so on). Costs of processing using NASA Earth Exchange (NEX) were also explored. LCMAP was chosen as the use case for the comparative cost analysis in keeping with the EAST's goal of enabling it.

With the baseline costs and metrics already calculated in the As-Is cost estimate, determining the price of storing and processing LCMAP data in-house was as straightforward as applying those metrics to LCMAP data volumes. It is important to note that acquiring the disk storage and processing hardware to do LCMAP in-house is largely a nonrecurring cost. By contrast, use of the public cloud is an ongoing and cumulative cost. To illustrate this difference, the analyst chose not to answer the question, “Which method of storage/processing is cheaper?” but rather “At what point in time does cloud utilization become more costly than executing LCMAP in-house?”

LCMAP Data Storage: In-House or Cloud?

When it comes to storage, there are other cost considerations beyond hardware. Note that input/output (I/O) in-house is at no additional cost, but data egress is charged per instance in the case of the cloud. Additionally, cost savings associated with migrating to the cloud are realized by doing away with hardware, the staff that maintains it, and the facility space that houses it. Because EROS is bound by policy to always maintain a data archive, and USGS owns the facility, cost savings such as these are not realistic expectations. So, the cost analysis focused solely on the cost to acquire disk space to accommodate LCMAP in house versus leasing that space in the cloud.

When egress is accounted for, it becomes apparent that the cloud is a costly alternative to in-house storage. Increasing the frequency of data egress drives the breakeven point further to left, in favor of in-house storage acquisition. The break-even point can occur as quickly as within 3 months if egressed daily by one user, within 18 months if egressed weekly by one user, or within 38 months if egressed monthly by one user. Egress cost also multiply as a function of the number of users. Cloud cost estimates consider the average of multiple vendors’ currently advertised cloud storage costs, steady-stated throughout a 4-year LCMAP assumed duration.

LCMAP Data Processing: In-House, Cloud, or NEX?

The cost to process LCMAP data in the cloud (assuming it is possible) depends on the type of processing power leased. When comparing various public cloud options to EROS’s average processor purchase prices, the breakeven point against the average EROS processor cost occurs within 2 to 4 years for high memory or high central processing unit (CPU) jobs. Again, these estimates consider the average of multiple vendors’ currently advertised cloud processing costs, steady-stated throughout a 4-year LCMAP assumed duration.

The NEX offers a low cost-risk government trusted partner relationship for data processing. NEX pricing is formulaic and straightforward, and like hardware purchase prices and cloud lease costs, is also trending downward over time. Unfortunately, a ROM estimate could not be generated without an estimate of wall clock hours needed.

A.4 Recommendations

The comparative cost analysis did not find public cloud computing as being a lower cost option to EROS’s current way of doing business. Given that EROS’s current way of doing business is keeping pace with data center industry trends (despite a slight time lag), entering into the volatile cloud market at the present time seems to offer more cost risk than opportunity. Rather, the cost analyst recommends EROS include NEX as an option in follow-on studies and engage with NASA in detailed discussions. Lastly, as EROS moves ahead with its shift to an enterprise operations concept, a skill mix shift away from hardware maintenance in favor of strengthening asset acquisition and management is recommended.

Appendix 2: EAST Deliverables

The following listed materials were provided by the EAST to the study sponsor and steering committee:

- Final Report—To be released following delivery and approval by the study sponsor,
- Checkpoint no. 1 slides,
- Checkpoint no. 2 slides,
- Checkpoint no. 3 slides,
- Science User Characterization,
- Use Cases,
- EROS As-Is Architecture Drawing,
- Independent Comparative Cost Study—Comprehensive Slide Materials provided by The Aerospace Corporation,
- EAST Request for Information (RFI), and
- EAST RFI Responses along with respective one-on-one meeting materials.

Appendix 3: Acronym List

Acronym	**Definition**
AB	Applications Branch
API	Application Programming Interface
ARB	Architecture Review Board
ARC	Ames Research Center
ARD	Analysis Ready Data
BYOD	Bring Your Own Device
CBA	Cost-Benefit Analysis
CEOS	Committee on Earth Observation Satellites
CER	Cost Estimating Relationship
CITT	Center Information Technology Team
CLASS	Comprehensive Large Array-data Stewardship System
COBIT	Control Objectives for Information and Related Technology
CP	Checkpoint
CPU	Central Processing Unit
CRADA	Cooperative Research and Development Agreement
CRO	Coordination and Requirements Office
DAAC	Distributed Active Archive Center
DBaaS	Database as a Service
DCCI	Data Center Consolidation Initiative
DoDAF	Department of Defense Architecture Framework
EAST	EROS Architecture Study Team

EDRN	Early Detection Research Network
EO	Earth Observation
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
EROS	Earth Resources Observation and Science
ESD	Earth Science Division
ESDIS	Earth Science Data and Information System
ESRI	Environmental Systems Research Institute
ET	Evapotranspiration
FEA	Federal Enterprise Architecture
FEWS NET	Famine Early Warning Systems Network
FFRDC	Federally Funded Research and Development Center
GB	Gigabyte
GFOI	Global Forest Observations Initiative
GIS	Geospatial Information System
GOES	Geostationary Operational Environmental Satellite
GPS	Global Positioning System
GSFC	Goddard Space Flight Center
HPC	High Performance Computing
HVAC	Heating, Ventilation, and Air Conditioning
IaaS	Infrastructure as a Service
IEEE	Institute of Electrical and Electronics Engineers
iGETT	Integrated Geospatial Education and Technology Training
IHAS	In-House Applications/Services
IOC	Initial Operating Capability
IRB	Investment Review Board
IT	Information Technology
ITIL	Information Technology Infrastructure Library
JPG	Joint Photographic Experts Group
JPL	Jet Propulsion Laboratory
JPSS	Joint Polar Satellite System
LCMAP	Land Change Monitoring and Projection
LCMS	Landscape Change Monitoring System
LP DAAC	Land Processes Distributed Active Archive Center
LRS	Land Remote Sensing
LSDS	Land Satellites Data Systems
LSRD	LSDS Science Research and Development
LTA	Long Term Archive
MODIS	Moderate Resolution Imaging Spectroradiometer
MOE	Measure of Effectiveness
MOP	Measure of Performance

NASA	National Aeronautics and Space Administration
NESDIS	National Environmental Satellite, Data, and Information Service
NEX	NASA Earth Exchange
NIH	National Institute of Health
NOAA	National Oceanic and Atmospheric Administration
NPP	Suomi National Polar-orbiting Partnership
O&M	Operations and Maintenance
OMB	Office of Management and Budget
OODT	Object Oriented Data Technology
Ops Con	Operations Concept
PaaS	Platform as a Service
PDS	Planetary Data System
PM	Project Manager
PMSEIT	Program Management, Systems Engineering, and Integration and Test
POES	Polar Orbiting Environmental Satellites
RB	Research Branch
RCA-EO	Requirements, Capabilities and Analysis for Earth Observation
RFI	Request for Information
ROM	Rough Order of Magnitude
RS	Remote Sensing
SaaS	Software as a Service
SCI	Science
SDW	Spatial Data Warehouse
SGT	Stinger Ghaffarian Technologies
SIPS	Science Investigator-led Processing System
SLA	Service Level Agreement
STAR-FM	Spatial and Temporal Adaptive Reflectance Fusion Model
TB	Terabyte
TIM	Technical Interchange Meeting
TOGAF	The Open Group Architecture Framework
TSSC	Technical Support Services Contract
USDA	U.S. Department of Agriculture
USFS	U.S. Forest Service
USGS	U.S. Geological Survey
VDI	Virtual Desktop Infrastructure
VIIRS	Visible Infrared Imaging Radiometer Suite
WBS	Work Breakdown Structure
WELD	Web-Enabled Landsat Data