



Volume I

Strategic Evolution of Earth Science
Enterprise Data Systems (SEEDS)
Formulation Team Final
Recommendations Report

July 2003

Submitted by the SEEDS Formulation Team

Dr. Stephen Wharton, Formulation Manager

Vanessa Griffin, Deputy Formulation Manager

Catherine Corlan

Kathy Fontaine

Gail McConaughy

Ken McDonald

Karen Moe

H. Ramapriyan

Richard Ullman

Abstract

The purpose of SEEDS is to establish a framework for distributed data management to maximize availability and utility of ESE products; leverage community expertise, ideas, and capabilities; and improve overall effectiveness of ESE-funded systems and services. Based on recommendations and guidance provided in the NewDISS document, SEEDS study teams, at a very high level, intend to recommend the following to the ESE:

- Adopt the levels of service recommended within and establish a working group to keep them up to date (Chapter 1)
- Adopt the standards process outlined within and establish a working group to enable process implementation and evolution (Chapter 2)
- Adopt the cost estimation model and require all awardees to contribute to the model's database (Chapter 3)
- Adopt the model, roles and responsibilities, and policies outlined to handle long-term archive issues (Chapter 4)
- Adopt the proposed reuse recommendations and establish a working group to enable implementation and evolution (Chapter 5)
- Adopt the recommendation to address the gaps in data service technology development and implement a SEEDS-specific technology vision (Chapter 6)
- Adopt the proposed metrics and establish a working group to enable implementation and evolution (Chapter 7)

Of course, each of these recommendations has much more depth than can be addressed here, and each is more fully explained in subsequent chapters. The overarching recommendation that will give all of the above recommendations the greatest chance for success is to

- Establish a SEEDS Office to coordinate all of the above recommendations in detail, across all stakeholders. (Chapter 7)

SEEDS has, from the start, played a coordinating role in bringing together disparate opinions from the stakeholder community to identify the pertinent issues facing data and information dissemination and management, assign those issues a level of urgency, and identify realistic and innovative solutions. That community, comprised of data and information systems and services providers; interested science, education, and applications providers and users; and other agencies and organizations with an interest in Earth science data issues, will now be able to see the fruits of their work collected in one place. This single recommendations document provides the reader with a better idea of the scope of SEEDS, and a way to more accurately assess the impact of these recommendations.

Contents

ABSTRACT.....	3
INTRODUCTION.....	5
1.0 LEVELS OF SERVICE	16
2.0 STANDARDS FOR NEAR-TERM AND LONGER TERM MISSIONS	19
3.0 COST ESTIMATION.....	31
4.0 DATA LIFE CYCLE AND LONG-TERM ARCHIVE	35
5.0 REUSE AND REFERENCE ARCHITECTURES.....	42
6.0 TECHNOLOGY INFUSION.....	48
7.0 METRICS PLANNING AND REPORTING AND GOVERNANCE	54
8.0 REMARKS ON THE NEXT PHASE	66
APPENDIX FOR CHAPTER 1.....	70

Introduction

The Strategic Evolution of Earth Science Enterprise (ESE) Data Systems (SEEDS) formulation study was established to develop a strategy and coordinating program to evolve the ESE network of data systems and service providers in the 2004 to 2015 timeframe.

Over the past decade, National Aeronautics and Space Administration (NASA) ESE has made a substantial investment in the development of data and information systems. This is most evident in the Earth Observing System (EOS) Data and Information System (EOSDIS) Core System (ECS), but also includes unique components developed by the Distributed Active Archive Centers (DAACs), the data processing systems, and other capabilities developed by the instrument teams that are still actively used and maintained as a result of heritage missions and initiatives. SEEDS is not intended to be a replacement of these capabilities, but rather the evolution of existing systems, through improved effectiveness and efficiency of operation and services, to maximize the return on those previous investments.

The purpose of SEEDS is to establish a framework for distributed data management to accomplish the following goals and objectives:

1. Maximize availability and utility of ESE products:

- Facilitate discovery and interchange of independently developed products and services.
- Promote product stewardship across multiple missions and/or projects.
- Reduce the lifecycle cost of providing data and information products.
- Ensure that products meet norms for utility, accessibility, security, and survivability.

2. Leverage community expertise, ideas, and capabilities:

- Provide a focal point to engage the community in evolving ESE standards, interfaces, and services from today forward.
- Establish a more broadly-based set of product and service providers.
- Encourage reuse of successful providers' expertise and capabilities across multiple missions and/or projects.
- Give systems and service providers appropriate local control over data system design, implementation, and operation.

3. Improve overall effectiveness of ESE-funded systems and services:

- Identify guidelines for distribution of organizational responsibilities, requirements, and interfaces.
- Improve the collective ESE data systems responsiveness to changes.
- Refine capability to predict, manage, and monitor trends in development and operations costs.

- Provide capability to measure quantity, quality, and impact of ESE system investments.
- Leverage competition, technology infusion, and reuse to improve system effectiveness.
- Ensure that services meet norms for utility, accessibility, security, and survivability.
- Monitor collective performance in meeting ESE objectives and goals.
- Maintain sufficient organizational structure to allow effective resource management and implementation for NASA to carry out its science mission.

This introduction provides an overview of the approach, status, recommendations, issues, and next steps for the SEEDS study as a whole. Section I describes the background, origin of SEEDS, and ongoing evolution of ESE data systems. Section II provides an overview of the formulation approach and introduces the study teams, gives examples of community involvement, describes the schedule and status highlights, and describes the next steps. Section III provides a summary overview of the study teams' recommendations for levels of service, data and information standards, cost estimation, lifecycle data management, software reuse and reference architecture, technology infusion, metrics planning and reporting, and governance.

I. Background

The acquisition of increasing volumes of data from public and commercial systems--data with better spectral and spatial resolution than ever before--presents a challenge to government and commerce to make those data and data products readily available to the user community, to extract the information and knowledge content from these rich observations, and assimilate the data and knowledge into decision support systems.

The ESE recognizes its responsibility to ensure that all the information, knowledge, and capabilities derived from its research program achieve maximum usefulness in research, applications, and education. ESE is evolving its science data and information systems towards a more open, distributed set of data systems and service providers. This approach will capitalize on the expertise and resources of the community of providers and facilitate innovation. Implementation of this approach relies, in part, on leveraging information technologies from the commercial sector, such as web-based techniques for data discovery and access, and involving the end user community in technology assessment and evolution.

In addition to its investment in EOSDIS, ESE has a number of initiatives to advance its data systems and deliver data products and results to the nation. First, it initiated a multiyear Earth Science Information Partner (ESIP) experiment that formed a federation of competitively selected data centers to further explore and refine the issues associated with distributed, heterogeneous data and information system and service providers. Second, the EOSDIS architecture evolved to accommodate the generation of data products by processing systems external to ECS developed under the direction of the EOS instrument teams. Third, the ESE chartered a study team called New Data and Information Systems and Services, or NewDISS, to capture and consolidate the input from the community in a series of recommendations. SEEDS is an outcome of this study

and is chartered to work with the Earth science user and data provider communities to generate approaches and plans for future ESE data and information systems.

A. Origin of SEEDS

The guiding principles of SEEDS were defined in the NewDISS strategy document. SEEDS starts from the premise that systems and services must be informed by, and supportive of, key science concerns and questions. It is also recognized that individual scientists and disciplinary communities of scientists are both consumers and producers of data products and derived information, and therefore must be key partners. Other principles relate to the issue of immediate and long-term services for a highly distributed and heterogeneous user base in the face of rapid technological change. These principles are summarized as follows:

- Science questions and priorities must determine the design and function of systems and services.
- Future requirements will be driven by high data volume, a broader user base, or increasing demand for a variety of data and data products.
- Technological change will occur rapidly; systems and services must be able to take advantage of these changes.
- Competition is a key NASA tool for selecting NewDISS components and infrastructure.
- Principal investigator (PI) processing and PI data management will be a significant part of future missions and science.
- Long-term stewardship and archiving must occur.
- NewDISS evolutionary design must, over time, move beyond basic data and information and towards knowledge-building systems.

The SEEDS concept began with the formation of NewDISS strategy team. In July 1998, Dr. Ghassem Asrar, NASA's Associate Administrator for Earth Science, instituted a NewDISS strategy team to define a future framework and strategy for NASA's ESE data and information processing, near-term archiving, and distribution. This strategy was intended to answer the question: Based on lessons learned, what is the recommended viable and evolvable way of building a set of data and information systems and services to meet the ESE program needs?

The strategy team issued a report, entitled NewDISS: A 6- to 10-year Approach to Data Systems and Services for NASA's Earth Science Enterprise, as the pre-formulation concept, to help guide NASA in the formulation and development of NewDISS. The report recommended that NASA should "Charter, without delay, a transition team with the objective of developing a transition plan, based on the findings and recommendations of this document, that would lead to the initiation of a NewDISS starting in 2001."

In response to this recommendation, NASA headquarters asked Goddard Space Flight Center (GSFC) to lead the formulation of a NewDISS program, including planning the transition to the new data management paradigm. The name was changed from NewDISS to SEEDS to clarify our intentions of evolving towards a more distributed and

heterogeneous network of system and service providers as opposed to implementing the next version of EOSDIS.

B. ESE Data System Ongoing Evolution

In addition to SEEDS, the overall ESE data systems are evolving.

- EOSDIS has evolved over the past five to six years in a path aligned with SEEDS principles--science data processing and end-to-end data management responsibility was given to EOS instrument teams where possible; and responsibility for operations moved to DAAC institutions.
- The ESE science data systems are becoming increasingly decentralized--instrument teams produce Terra and Aqua data from CERES, MODIS, MOPITT, and AMSR-E using Science Investigator-Led Processing System (SIPS), and instrument teams will produce all data from Aura instruments.
- There are increasing opportunities for distributed service providers--ESIP federation prototypes have been funded since 1999 and a Research, Education, and Applications Solutions Network Cooperative Agreement Notice (REASoN CAN) was released to support new initiatives starting in 2003.
- There is greater emphasis on reuse--DAAC-developed software has been reused by SIPS and integrated into the ECS baseline and it is establishing theme-based Science Data System concept for Global Precipitation Mission (GPM) and National Polar Orbiting Environmental Satellite System (NPOESS) Preparatory Project (NPP).
- NASA is identifying technology needs and investments--The Earth Science Technology Office Advanced Information System Technology (ESTO AIST) program sponsored three annual workshops to identify and refine ESE needs for information systems, including categories of direct applicability to NewDISS principles. (data and information production; search, access, analysis and display; and transmission and dissemination).

II. Formulation Study

A. Formulation Approach

The study team approach has been to define “what” the SEEDS office should do as distinct from the “how” of its organization or governance. The rationale was that it would be pointless to attempt to address organizational questions without first having a clear idea of what it was the organization was supposed to do. In addition, it was essential to engage the community in identifying the functions and priorities for SEEDS and to build community support and ownership of the SEEDS goals and processes.

The SEEDS formulation effort has been and will continue to be outward looking and inclusive. Wherever appropriate, the SEEDS studies have addressed as wide a range of related activities as possible, within government, industry, and academia in the U.S. and abroad. By taking a broad view, it is expected that the recommendations in this document capitalize on the extended experience base and the best practices and latest

technical approaches available to achieve maximum effectiveness and efficiency in development and operation of NASA Earth science data and information management systems.

The overall study team approach can be summarized as follows. Community expertise is leveraged through consulting arrangements. Existing practices, capabilities, and lessons learned are surveyed. The broader community is engaged in clarifying technical areas / questions to be addressed, identifying science concerns / issues pertinent to the study, developing and reviewing options to address concerns / questions, and developing and refining SEEDS recommendations. The surveys, questions, and recommendations are then iteratively refined in response to community feedback. To achieve its objective, the SEEDS Formulation Team has set up study teams to investigate specific subjects of concern to SEEDS and make recommendations. The seven study teams, with members from government agencies, universities, and industry, and their tasks are summarized in Table I-1 below.

Table I-1: Summary of Formulation Study Efforts

Study Team	Study Task Summary
<i>Level Of Service and Cost Estimation</i>	Worked with the research and applications communities to develop the minimum and recommended levels of service for core data sets and services required from ESE data management service providers. Determined, from benchmarking, what data management services should cost, and are developing a capability to perform end-to-end cost estimates for ESE data management services.
<i>Near-Term Mission Standards</i>	Considered ESE's near-term systematic measurement missions; recommended science data, metadata, and interoperability standards for applications; and incorporated advice and experience of mission science community in making recommendations.
<i>Standards and Interfaces Processes</i>	Defined a process for SEEDS to develop, adopt, evolve, and maintain standards and interfaces for data and information systems and services across ESE. The process capitalized on the methods and experience of existing relevant data systems standards bodies (e.g., ISO, OGC) and NASA programs (e.g., EOSDIS, ESIP Federation).
<i>Data Life Cycle and Long Term Archive</i>	Outlined policies to ensure safe handling of SEEDS and previous-era products as they migrate from data providers to active archive and long-term archive even as numerous individuals and institutions take responsibility for the product during its life cycle.
<i>Metrics Planning and Reporting</i>	Defined appropriate metrics and reporting requirements for the participants in ESE data management activities and demonstrate that proposed SEEDS organization structure can provide adequate accountability.
<i>Reuse and Reference Architecture Assessment</i>	Defined an approach for investment in software reuse and in the development of a reference architecture. Examined what is the best method to assure effective and accountable community involvement, and the best technical approach.
<i>Technology Needs and Infusion Prototyping Needs.</i>	Determined processes by which technology needs are identified and technology investments are infused into the evolving NewDISS. Recommended ways for SEEDS to leverage the processes of NASA ESTO's AIST program, involve ESE user community, and designate roles of ESTO AIST and SEEDS with regard to prototyping needs.

B. Community Involvement

The community is involved at many levels in SEEDS formulation as participants and consultants on the various study teams; as contributors of white papers, workshop attendees, and survey respondents; and as advisory panels in reviewing SEEDS plans and recommendations. Community involvement within the study team efforts is summarized in the following chapters.

C. Formulation Study Status and Schedule Highlights

The study teams have surveyed current practices, discussed the formulation approach and preliminary findings at workshops, and prepared an integrated set of recommendations.

The first public workshop was held at University of Maryland, February 5-7, 2002. The workshop had significant participation from the data provider community. The Formulation Team received feedback on additional process elements to be considered and had 15 white papers and 40 cost team recommendations.

The second public workshop was hosted by the San Diego Supercomputing Center, June 18-20, 2002. Presentations were held on best practices in other environments (NASA and non-NASA). The Formulation Team solicited community feedback and ideas in response to preliminary recommendations.

The Formulation Team defined objectives, guidelines, and criteria to reflect SEEDS principles in the REASoN CAN.

The preliminary findings were presented to Dr. Asrar and the Earth System Science and Applications Advisory Committee (ESSAAC) Subcommittee for Information Systems and Services (EISS). EISS endorsed the findings in their presentation to the full ESSAAC. The draft version of this document was posted for comment from January through April 2003, and many thoughtful, insightful comments were received and incorporated into this final version.

The third workshop was held March 18-20, 2003, in Annapolis to refine study findings and establish working groups charters. The response to the draft version of this document and to presentations at the workshop was very positive and constructive.

D. Next Steps

As indicated above, the activities leading to this report have focused on the “what” of SEEDS as opposed to the “how.” In addition, the Formulation Team also developed options for a SEEDS implementation organization. An initial presentation has been made to the ESE associate administrator and senior staff, and the associate administrator has requested further examination of how the present EOSDIS will relate to the SEEDS framework recommended in this report. It is expected that the associate administrator will select an option for implementation based on this additional examination.

While this document represents the end of the first phase of SEEDS development, and no decision has been made to formally endorse the formulative recommendations to date, the team will continue working to address the outstanding issues and questions surrounding SEEDS. To that end, the Formulation Team will be developing a transition plan to

implement the selected options, defining allocation of roles and responsibilities, further refining which elements should be centralized vs. distributed, defining organizational elements, working on assorted data system actions, preparing a draft of ESE guidelines and transition plan for associate administrator review and approval, and presenting to advisory panels for review.

III. Draft Recommendations

More detailed descriptions of the methodology, findings, and recommendations appear in the chapters indicated below. The reader is invited and strongly encouraged to comment on any aspect of the findings and recommendations. All comments will be addressed and incorporated into the final recommendations, to be delivered at the end of FY03.

A. Levels of Service (Chapter 1)

Findings – One level of service does not fit all situations, and users should reasonably expect different levels of service for different products. In addition, levels of service should not be linked to types of data providers (i.e., backbone data centers, applications data centers), but rather should be defined based on the provider’s capability, the user’s needs, and the types of data.

Recommendations – *The ESE should adopt the levels of service developed by this study as an initial working basis for definition of requirements for future ESE data activities. The requirements and levels of service should be subjected to ongoing review by ESE and the community via a working group, and should be updated as needed to reflect changes in ESE program needs, evolution of modes of operation that are driven by user needs, and advances in information technology.*

B. Data and Information Standards (Chapter 2)

Findings – Requirements for system interchange among ESE components are different from requirements for distribution to end-users. In the near term, the chief mode of delivering data remains the transfer of discrete files. Therefore, data format is the critical component of data packaging. The use of Web Service standards is still only emerging.

Multiple options should be provided for data packaging, especially for service to end users, even in the near term. Several missions have experienced success in data distribution to multiple user communities using different data format standards. Community-based standards, or profiles of standards, are more closely followed than standards imposed by outside forces.

Recommendations – *In the near term, ESE should maintain format translators to distribute products in multiple formats. Upgrade interoperability capabilities (catalog, inventory, distribution). Plan for evolution of packaging requirements. Support ESE-unique standards (development, maintenance, training, help desk). Support evolution of science data formats towards seamless operability.*

For the near term, ESE should require that standard products be file-based and use Hierarchical Data Format (HDF) or Network Common Data Form (netCDF) as an interchange format. Distribution formats should address user needs and convenience. Mission standard products should be further defined using profiles; use Global Change Master Directory (GCMD) Directory Interchange Format (DIF) for collection metadata, and the ECS or EOS Clearing House (ECHO) data model for inventory metadata (pending International Standards Organization [ISO] 19115 standard); be documented using EOSDIS guide standard; and use EOSDIS V0, Z39.50-based, or ECHO-compatible search and order protocols.

For ongoing refinement, adoption and possible development of standards, recommend that ESE adopt a process similar to the Internet Engineering Task Force (IETF) process and tailored to meet specific ESE needs. Develop a strategy for facilitating ESE standards compliance across the enterprise, including the performance of standards support services, e.g., user support, training, tool development. Encourage adoption of existing successful standards. Develop new standard if there are no existing viable candidates.

C. Cost Estimation (Chapter 3)

Findings – The cost-by-analogy method, which produces estimates based on comparable activities, is a valid approach to estimating cost. There are 4 entries at present; the database is expected to require a minimum of 24 to be operational. The first version of the model is to be completed in 2003.

Recommendations – *The ESE should adopt, as an aid to ESE program staff and principal investigators, the life cycle cost estimation model to estimate the cost of various types of data activities (e.g. DAACs, ESIPs, SIPs, RESACs, project data systems, etc.); The ESE should require, via appropriate language contained in each funding instrument, each funded science data and information service provider to provide actual life-cycle workload and effort information for the comparables database that is used as the primary basis for cost estimation.*

D. Life Cycle Data Management (Chapter 4)

Findings – A balance must be struck between "save it all no matter what" and "save only what we can afford right now." Considerations of data lifecycle issues should be built in to the entire process, but not be considered expendable when budgets decline. Complete documentation is vital to whatever is archived, no matter where it is archived.

Recommendations – *Define active and long-term archives (LTAs) for each data product. Data-buy terms must address question of eventual NASA ownership of data. All archive data collections should be complete, including required ancillary data, project and data set documentation, and the science production software. Data become the responsibility of LTA upon acceptance of data. Archived data should be available without loss or*

degradation in quality. Point of contact required that can answer questions about data or its use. Throughout a product's lifecycle a point-of-contact should be provided that can be utilized for questions about the data or use of the data.

Enter into agreement with archive provider for all LTA products. Keep pre-launch drawings, documentation and data and update the information to keep it accessible. Set up a liaison between the pre-mission team and the archive team. Transfer data essential for science data processing to the active archive as soon as possible. Transfer spacecraft and instruments pre-launch data to an active archive in a common format.

E. Software Reuse and Reference Architecture (Chapter 5)

Findings – Both the mission-critical and the mission-success communities are not satisfied with the status quo. The mission-critical community is strongly in support of extending and improving clone & own practices to enable developers to identify existing assets, subsequently copy those assets, and modify and integrate them more easily as needed for use in new systems. The mission-success community favors the open source--engaging developers across missions to collaboratively develop and update selected components or systems stored in a central repository--and feel that service encapsulation should be explored via technology development. Both communities strongly opposed attempting a product line approach, that is, reusing a set of core software components intentionally designed for a family of systems where the components are modified and maintained only by the organization responsible for the core components.

The opinions of the ESE community regarding reference architecture alternatives were not as strong as they were regarding reuse alternatives. It was decided that reference architectures will be a support function to the software reuse needs.

Recommendations – *Establish two working groups: one focused on the improved clone & own approach in mission-critical environments and one focused on the open source approach in mission-success environments.*

- *These working groups could be responsible for recommending specific reuse initiatives, and for working in the areas of outreach and education, support and enablement, and policy change to further enable reuse within their respective communities.*
- *Cross working group successes would be integrated and communicated as the processes evolve.*

Establish a separate body such as a reuse integration office whose functions would include: prioritizing and approving reuse initiatives; selecting and guiding community reuse projects; administering reuse incentives; conducting some reuse outreach and education activities; and including a small technical team to support all reuse-related activities, and evaluate cost savings and impacts on schedule of the reuse functions. Complement the reuse effort by an effective technology development and technology infusion effort to bring in new and increased functionality.

F. Technology Infusion (Chapter 6)

Findings – Technology development program funding does not cover tasks needed to make a technology suitable for operational deployment, nor to deploy it. Uncertainties surrounding the licensing of new technologies (especially those subject to commercial acquisition and those involving intellectual property rights) increase the risk of incorporating and becoming dependent on a new technology. New technologies can introduce performance and availability risks into operational systems, and there is no structured program to help evaluate and eliminate these risks. The technical infrastructure assumed by a new technology is often incompatible with the infrastructure of existing operational systems.

Recommendations – *Fund efforts to bridge current gaps between technology developers and the data service providers who are potential technology users. Develop a SEEDS capability vision that helps to capture, communicate, and refine the community’s understanding of the critical capabilities that will enable the next generation of ESE research and applications. Adopt and tailor processes for technology needs and gap analysis, while leveraging and tailoring community-based technology infusion processes.*

G. Metrics Planning and Reporting (Chapter 7)

Findings – The team (and survey) recognized two important metrics related to “outcome” – citations and customer “nuggets” (key success stories). There was, however, no consensus on other useful outcome metrics *per se*. Outcome metrics need to be developed to help measure the value of an activity’s data and services to the science or applications community, or measure the actual utilization of data by the community. Metrics derived from the user’s point of view, e.g., easy access to readily usable, well-documented data, products, and services, still need to be defined. “Output” metrics continue to be seen as a useful measure of the productivity of an activity. There was considerable interest in establishing an enterprise function for integrated reporting of metrics and successful accomplishments at the SEEDS program level.

Recommendations – *Because of the need to improve sponsor-required user satisfaction metrics or outcome metrics, it is recommended that this class of metrics be studied further. An extension of this study should be to identify metrics that are directly traceable to the objectives of the ESE science and applications program.*

A SEEDS Office should take on the responsibility of managing and collecting program-level metrics and accomplishments as an ESE function. It is recommended that metrics activity by the SEEDS Office be limited to those metrics that are required for program-level assessment and monitoring. The SEEDS Office would maintain and update the program-level metrics over time.

A SEEDS Metrics Planning and Reporting Working Group (MPAR WG) should be established for ongoing evaluation and evolution of appropriate metrics. Future solicitations for data systems and service providers should include a requirement for the

bidders to suggest a set of metrics that demonstrate how their proposed activities will address the goals of ESE's science and applications programs and require participation by the selected providers in the MPAR WG. The solicitations also must require that the providers gather and report on an agreed upon set of metrics.

H. Governance (Chapter 7)

Findings – There are aspects of each study team's recommendations that refer to requirements definition or implementation, policy definition or implementation, metrics collection or management, or international/interagency/university/etc. interaction. Such activities, in order to be successfully integrated across the ESE, require some kind of coordinating function.

Recommendation – *Establish a SEEDS Program Office to handle the coordination and integration of the various recommended functions across all stakeholders.*

1.0 Levels of Service

1.1 Purpose

The Levels of Service (LOS) and Cost Estimation (CE) team was established in order to: 1) Work with the science and applications communities to develop the minimum and recommended LOS for core data sets and services required from ESE data management service providers, 2) Determine, from cost by analogy methodology, what data management services should cost, and 3) Develop a capability to perform end-to-end cost estimates for ESE data management services. This section will only address element 1. Elements 2 and 3 are addressed in Section 3.0.

1.2 Members

Vanessa Griffin, GSFC, Team Lead; Kathy Fontaine, GSFC; Bruce Barkstrom, Langley Research Center (LaRC); Claude Freaner, NASA headquarters; Bud Booth, Stinger Ghaffarian Technologies (SGT); Greg Hunolt, SGT; David Torrealba, SGT; Mel Banks, SGT.

1.3 Goals and Objectives

In the SEEDS era, ESE data service providers will need to have as much flexibility to implement and operate as possible. However, the users of data from these providers will be expecting to receive a level of service similar to that received today from the EOSDIS. The goal of the LOS effort was to identify and recommend a range of service levels for the various activities to be carried out by SEEDS data service providers.

1.4 Approach and Community Involvement

To arrive at a minimum LOS, the study team first examined LOS requirements from the EOSDIS system (V0 and ECS), from which the team developed a high level set of LOS principles. Next, the team drafted a set of LOS requirements from those principles, grouped by data function. In determining the functions to be performed, the team drew from current experience with the DAACs, SIPs, and ESIPs. While the team had intended to group the functions into *physical* data service provider types (the original NewDISS concept), community feedback from the first workshop revealed that simply assigning LOS to functions and allowing the data providers to pick the functions they would need to implement was the optimum approach. This approach in effect permits the ESE community to define a *logical* data service provider type.

The study identified a range of services for the various functions that data service providers will need to perform. To ensure that investigators and project managers have

the greatest degree of flexibility to meet their requirements, the minimum LOS should be as non-constraining as possible. While the formulation team recommends minimum LOS, in many cases the minimum LOS will not be at the level desired by the user community. Thus, the study team has also defined a recommended LOS, as shown in the appendix. In all cases the LOS are defined for key functions to be performed by future data and service providers. In the future, proposers will pick and choose among the functions to be provided, as long as they meet the minimum LOS for each function. Note that the team has only recommended LOS for individual data providers and has not yet considered the service levels needed from cross-provider infrastructure components (e.g., networks, data access, metadata clearinghouses). These will be addressed in follow-up activities.

1.5 Potential Benefits / Impact

The need for a baseline of LOS for future NASA-funded data service providers is self-evident. Users of the current system have come to expect a minimum quality of service that should be present in the SEEDS era. In fact, in discussion with the ESE associate administrator, the SEEDS formulation team was asked to ensure that users would not see any decrease in service due to a transition from EOSDIS to SEEDS. Thus the potential benefit of the study is that the ESE will be able to solicit and monitor the LOS from all future data service providers.

1.6 Recommendations and Findings

The LOS study team recommends:

1. The ESE should adopt the LOS developed by this study as an initial working basis for definition of requirements for future ESE data activities.
2. The requirements and levels of service should be subjected to ongoing review by ESE and the community via a working group, and should be updated as needed to reflect changes in ESE program needs, evolution of modes of operation that are driven by user needs, and advances in information technology.

The findings of the LOS study are explained in detail in six working papers (WP). These papers, which continue to be works-in-progress, are provided in the appendices to this document and are available on the SEEDS web site, <http://lennier.gsfc.nasa.gov/seeds/>. Information relevant to this section is concentrated in WP 3 - Data Service Provider Reference Model - Functional Areas, WP 4 - Data Service Provider Reference Model - Model Parameters, and WP 5 - Data Service Provider Reference Model - Requirements / Levels of Service. Additional supporting functions such as facility maintenance and system engineering are described in detail in the Volume II Appendices.

1.7 Next Steps

Based on feedback received from the community in FY2003, the study team will refine the current set of data service provider functional areas, LOS, and requirements, and develop a community-based process that will evolve the current set of data service provider LOS.

2.0 Standards for Near-Term and Longer Term Missions

2.1 Purpose

Future ESE data systems will consist of a heterogeneous mix of interdependent components derived from the contributions of numerous individuals and institutions. These widely varying participants will be responsible for data management functions, including data acquisition and synthesis, access to data and services, and data stewardship.

“An important premise underlying the operation [of the ESE network of data systems and services] is that its various parts should have considerable freedom in the ways in which they implement their functions and capabilities. Implementation will not be centrally developed, nor will the pieces developed be centrally managed. However, every part [of the ESE network] should be configured in such a way that data and information can be readily transferred to any other. This will be achieved primarily through the adoption of common standards and practices [1].”

The SEEDS recommendations for standards rely on two principles that are in tension with one another. The first is that standards are best when developed by and for particular communities to meet specific, identified community needs. The second is that a standard must be widely followed in order for the ESE to receive benefits of standardization. These standards and standard interfaces will enable or facilitate the system interoperability and data interuse that is required to meet the overall objectives of the ESE. ESE must achieve a balance between the number of standards that must be supported and the specific requirements of particular communities of users. If each ESE mission science investigation, or distribution system, uses a self-defined standard then there is no standard. At the other extreme, if there is only one ESE standard, then it will be a bad fit for nearly all applications.

The SEEDS Formulation Team initiated two related studies to address the topics of data and information system and services standards. The Near-Term Mission Standards (NTMS) study advises the ESE on standards for use by the ESE near-term missions. The Long-Term Standards Process (LTSP) study defines a set of processes whereby SEEDS can adopt, evolve, and maintain appropriate standards through active engagement of the affected communities. The SEEDS recommendation is that the ESE develop a community-based process by which data systems standards for the ESE are chosen or developed with community input. The recommended approach, explained in Section 2.3 below, is to adapt a standards adoption, development, and approval process from that of the Internet Engineering Task Force (IETF). This process will guide the evolution of ESE standards. In the near term, however, this standards process is not in place, and yet, there are missions in the planning stages that may be impacted by changes in standards. The NTMS study recommends a first evolutionary step in adoption of standards by endorsing specific standards and practices. These are listed in Section 2.2 below.

2.2 Near-Term Missions Standards Study

2.2.1 Members

Richard Ullman, NASA/GSFC, Study Team Lead; Dr. Jingli Yang, Earth Resources Technology, Inc. (ERT); Cheryl Craig, National Center for Atmospheric Research (NCAR); Dr. John Evans, Global Science and Technology, Inc. (GST); Dr. Larry Klein, L-3 Communications Analytics Corporation; Dorian Shuford, ERT; Dr. Siri Jodha Singh Khalsa, L-3 Communications Analytics Corporation; and Matt Smith, University of Alabama at Huntsville (UAH).

2.2.2 Goals and Objectives

The goal of the SEEDS NTMS study is to provide specific, concrete recommendations on data format, metadata content, catalog interface, and documentation standards for the near-term missions. The recommended standards pertain to the data distribution to end-users and to the data interchange among the data systems and services components in the ESE network.

2.2.3 Process and Community Involvement

The study team began with the following list of near-term missions provided by ESE in October 2001:

Mission Name	Phase	Anticipated Launch Date
Landsat Data Continuity Mission (LDCM)	Formulation	2006
NPOESS Preparatory Project (NPP)	Formulation	2006
Ocean Surface Topography Measurement (OSTM)	Formulation	2006
Ocean Vector Winds	Formulation	2007
Global Precipitation Measurement (GPM)	Formulation	2007
Solar Irradiance	Formulation	2007
Carbon Cycle Initiative (CCI)	Pre- Formulation	2008-2012
Total Column Ozone	Pre- Formulation	N/A

We studied the published objectives of the assigned missions and interviewed some key planners in an attempt to understand the role data systems and data systems standards were expected to play in those missions and did play in their direct heritage. We discussed our progress at the first SEEDS public workshop. To verify our general understanding, we also asked for, and received, direct one-on-one feedback from the near-term missions on our draft survey.

We investigated each of the standards identified by the mission heritage survey and common standards used in other government agencies and industry. We researched their technical aspects and surveyed the opinions held by potential end users and producers. The survey, interview, and workshop opinions were consolidated. We developed a structured survey and individually interviewed many EOSDIS DAAC User Working Group members and data users and producers at the National Oceanic and Atmospheric Administration (NOAA). We also conducted a survey

of EOS data users and producers at the 2002 NASA Science Data Processing Workshop. Our report titled, “Near-Term Mission Standards Recommendations [2],” is in the appendix of this document. It describes the study methodology, findings, and the draft recommendations.

In balancing the different standards and their applications, we postulated findings and presented them at the second SEEDS public workshop. We discussed these findings with the workshop participants and explored potential recommendations. We also contacted each of the near-term mission planning groups and discussed our findings and the results of the workshop discussion to garner further feedback. Listed below are our major findings and recommendations to the ESE and to the near-term missions themselves. These recommendations should be considered as the nearest-term starting point for the standards evolution process.

2.2.4 Near-Term Standards Findings

2.2.4.1 Near-Term Missions Survey Findings

The following are general finding derived from the survey.

1. In early planning, mission planners were most concerned with maintenance of heritage data content rather than the particular standards by which content are transmitted.
2. Several missions have experienced success in data distribution to multiple communities using different format standards.
3. Heritage distribution to end users (i.e., those not on the mission science team) is most often through a DAAC.
4. The following heritage standards were identified:
 - Data format standards:
 - Hierarchical Data Format (HDF) (including the HDF-EOS profile of HDF), Network Common Data Format (netCDF), Geographic Tagged Image File Format (GeoTIFF), Fast Format, Custom Binary, Binary Universal Format for Representation (BUFR)
 - Metadata standards:
 - ECS data model, Federal Geographic Data Committee (FGDC) content standard, GCMD DIF
 - Documentation standards:
 - EOSDIS Guide
 - Catalog Interface Standards:
 - EOSDIS Version 0, American National Standards Institute (ANSI)/ISO Z39.50

2.2.4.2 Standards Analysis General Findings

The following findings address standards analysis in broad terms:

1. Requirements for system interchange among ESE components are different from requirements for distribution to end-users.
2. Figure 2.2-1 illustrates the different kinds of data flows that may exist for ESE data. System interchange packaging standards must focus on interface standardization,

completeness, and correctness of transfer over ease of use. The primary requirement for distribution to end-users is ease of use.

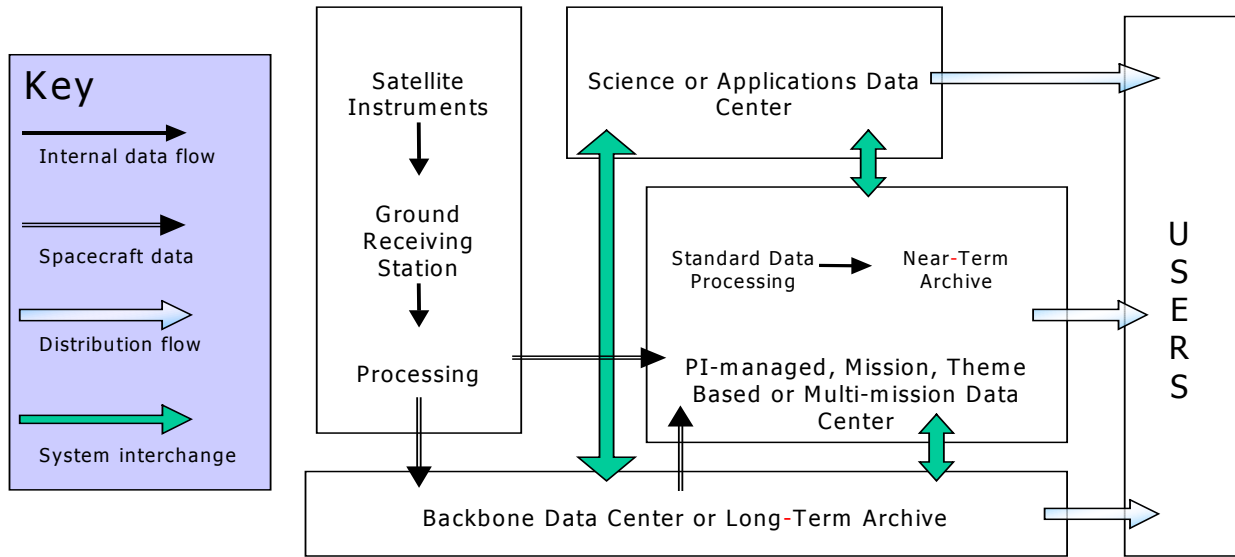


Figure 2.2-1 Notional ESE Data Flows

3. In the near term, the chief mode of delivering data remains the transfer of discrete files. Therefore, data format is the critical component of data packaging. Technologies such as content data standards are insufficient for transferring complex data between different user communities without information loss or corruption.
4. The use of a general standard (for example, HDF for data format or FGDC for metadata content) is insufficient for interoperability. The interfacing systems must also use a common profile of the standard. A profile is a specific convention of use of a standard for a specific user community.
5. Community-based standards, or profiles of standards, are more closely followed than standards imposed by outside forces. Community-based standards are standards developed by a community to meet cooperatively defined community needs.
6. The ESE, as a whole, or the systematic measurement missions independently, must plan for evolution of requirements for packaging of mission science data over the lifetime of the mission and beyond. These include standards for:
 - a. Data formats
 - b. Catalog interfaces
 - c. Associated metadata content and format
 - d. Documentation standards

2.2.4.3 Standards Analysis Specific Findings

The following are findings specific to format and protocol standards:

1. HDF (including the HDF-EOS profile of HDF) and NetCDF are appropriate format standards for system interchange.
2. GCMD DIF is widely accepted and appropriate for dataset catalogs.
3. Metadata content standards are converging on the ISO 19115 standard, because FGDC will adopt the ISO 19115 after it becomes final.
4. The EOSDIS Guide dataset documentation standard is successful and generally adequate for minimal description of standard data products.
5. For inventory interoperability, the EOSDIS Version 0 protocol is the only cross-enterprise standard in use since the early 1990s within ESE. More recent catalog protocols include the Geostationary Earth Observation (GEO) Profile of the ANSI/ISO Z39.50 by FGDC, the Catalogue Interoperability Protocol (CIP) Profile of Z39.50 by CEOS, and the OGC Catalog Interface Specification by OGC. The OGC Catalog Interface Specification is based on both CIP and GEO.

2.2.5 Recommendations

2.2.5.1 Near-Term ESE Missions Recommendations

Based on the survey findings, the following general recommendations were developed for standard data products:

1. ESE Standard Data Products must be file based and must be formatted for interchange among ESE data system components using HDF, HDF-EOS or netCDF.
2. ESE Standard Data Products must further be defined using a profile of HDF, HDF-EOS or netCDF. The profile must be chosen or developed with community input and in consultation with experts in the application of the base standard (i.e., HDF or netCDF).
3. ESE Standard Data Product dataset catalog metadata must be entered into the GCMD.
4. ESE Standard Data Product inventory metadata must be populated using either the ECS data model or the ECHO data model, since the ISO 19115 metadata standard is not finalized yet.
5. ESE Standard Data Products must be described using the EOSDIS Guide documentation standard.
6. ESE Standard Data Products must be made available for distribution by inventory using a system compatible with the EOSDIS V0 protocol, Z39.50 using CIP or GEO profiles, or any order and distribution mechanism compatible with ECHO.
7. ESE distribution components must enable packaging of standard products in formats and ways that emphasize end user needs and convenience.

2.2.5.2 ESE Data Systems Standards Evolution Recommendations

The following recommendations were developed to guide the development, maintenance, and monitoring of evolving data product standards:

1. Translators must be developed, maintained, and be made available to facilitate distribution of ESE Standard Data Products to end-users or applications users in multiple appropriate formats.
2. ESE data infrastructure components for catalog, inventory, and distribution must continually upgrade their interoperability capability to conform to the evolution of ESE, national, and international standards. Major infrastructure components include EDG client for universal search of ESE data holdings; the GCMD catalog of ESE datasets; Mercury, a web “harvester” approach for maintaining catalog metadata; the EOSDIS Guide system; and ECHO, a submissions-based inventory clearinghouse. Coordination among these infrastructure components is necessary.
3. Dataset documentation must be updated as documentation standards evolve. The EOSDIS Guide document standard should be maintained. Other documentation standards, especially Extensible Markup Language (XML)-based ones such as the Data Documentation Initiative (DDI) and Metadata Encoding and Transmission Standard (METS), should be monitored for applicability. ESE should also find techniques to make the guide more relevant to science data producers so that the documents are well written.
4. ESE must support ESE-unique standards through development and maintenance of standards software, user training, and help desk support to educate producers, consumers, and tool vendors.
5. ESE should invest resources in guiding the evolution of applicable science data formats through their respective governing processes with the goal of harmonizing them toward seamless interoperability. We recommend paying particular attention to the evolution of netCDF, HDF, GeoTIFF, and the World Meteorological Organization (WMO) BUFR and Gridded Binary (GRIB) formats.
6. Web service standards will have an impact on data, metadata, and interface standards. ESE should track developments in web service standards in the science and business communities.
7. Self-describing file formats remain a difficult problem with no adoptable solutions or standards available to date. ESE should continue to support the development of ESML, and should track more general information technology standards such as the Resource Description Framework (RDF) and the grid Storage Resource Broker/Metadata Catalog (SRB/MCAT) as potential solutions.

2.3 Long-Term Standards Process

2.3.1 Members

Study team:

Kenneth R. McDonald, NASA/GSFC (Study Team Lead); Jean-Jacques Bedet, Science Systems & Applications, Inc. (SSAI); Helen Conover, UAH; Allan Doyle, International Interfaces, Inc.;

Yonsook Enloe, SGT; Dr. John D. Evans, GST; Ramachandran Suresh, Mayur Technologies.

Consultants:

Prof. Liping Di, George Mason University (GMU); Prof. Jim Frew, University of California at Santa Barbara (UCSB); Douglas Nebert, FGDC; Prof. Silvia Nittel, University of Maine at Orono (UMO); George Percivall, GST; Lola Olsen, NASA/GSFC GSFC; Dr. Don Sawyer, NASA/GSFC GSFC; Dr. Chris Lynnes, NASA/GSFC GSFC.

2.3.2 Goals and Objectives

Standards and standard interfaces are important to the ESE for a number of reasons:

- To increase the use of the ESE data products for Earth science research and applications.
- To facilitate interoperability between different components of the ESE network of data systems.
- To make it easy for data and service providers to join the ESE network of data systems.
- To save costs for the ESE network of data systems as a whole.

The main objective of the study team is to have a fully developed set of standards processes and associated activities that have consensus support from the community to present to ESE management.

2.3.3 Analysis and Findings

The first task of the LTSP study was to compile a report on the standards activities of Earth science data systems projects and the processes, procedures, and results of relevant standards bodies and organizations. This report, titled, "Standards Organizations and Projects Survey Report [3]," was reviewed and analyzed to draw a set of general recommendations for SEEDS to follow and to develop candidate processes that the ESE could utilize to establish and support standards.

The LTSP study results include the work of the team members, reviews and suggestions of consultants, and community input.

From the study of previous and ongoing NASA programs and of existing standards bodies (e.g., ISO TC 211, OGC, World Wide Web Consortium [W3C]), the LTSP has identified a list of criteria that any ESE standards process should satisfy.

1. ESE should have a set of simple, open, well-defined processes to establish standards and standard interfaces for the ESE data systems. These processes must be evaluated using established performance metrics. The ESE standards must be documented and openly accessible.
2. ESE standards processes must support evolution of standards and standard interfaces (e.g., to respond to changing requirements or new technology).
3. ESE standards must be based on implementation experience and be supported by software tools.
4. ESE data systems standards processes must enable participation by the community and by external organizations. Active participation in the ESE data systems standards processes by

the community, including data users, missions, value-added providers, application users, and data centers, is essential. Active participation in the ESE data systems standards processes by external organizations, such as U.S. federal, state, and local agencies, international agencies, industry partners, commercial vendors, and international standards organizations, is also highly desirable and should be encouraged.

5. The ESE data systems standards process should be time bounded.
6. The ESE data systems standards process should have an appeal process to review contested decisions.
7. The ESE data systems standards process should encourage the use of existing successful standards and only develop new ones when deemed necessary. This will increase interoperability with existing systems and reduce ESE development costs.

2.3.4 Recommendations

2.3.4.1 Process Model

The LTSP study has identified a process or set of processes to develop or adopt and evolve and maintain standards and standard interfaces for data and information systems and services across the ESE. The notional process, described in our report "SEEDS Draft Standards Process Report [4]," is based on the process in use at the IETF. The IETF process provides technical excellence, prior implementation and testing, clear and concise documentation, openness and fairness, and potential for timelines.

- The ESE standards process should be based on the process in use at the IETF [5].
- The IETF standards process should be adopted but tailored to meet the specific needs of ESE, as detailed in Section 2.4.2 below.
- The findings, recommendations, and proposed process have been presented at the two SEEDS community workshops and have been refined to incorporate comments and ideas from the community.

2.3.4.2 Process Description

Process Initiation

Figure 2.3-1 describes the process to establish ESE standards. We anticipate many sources of standards and of requirements such as science users and applications, ESE project needs, existing standards, international or interagency agreements, HQ mandates, or vendor offered standards.

Once initiated, the process has two major pathways towards establishing an ESE standard. When an existing standard is applicable, that standard can be a candidate standard in an adoption process. When there is no suitable existing standard, a new one can be formulated following a development process. The ESE standards processes should encourage the adoption of existing successful standards, and only develop new ones when deemed necessary.

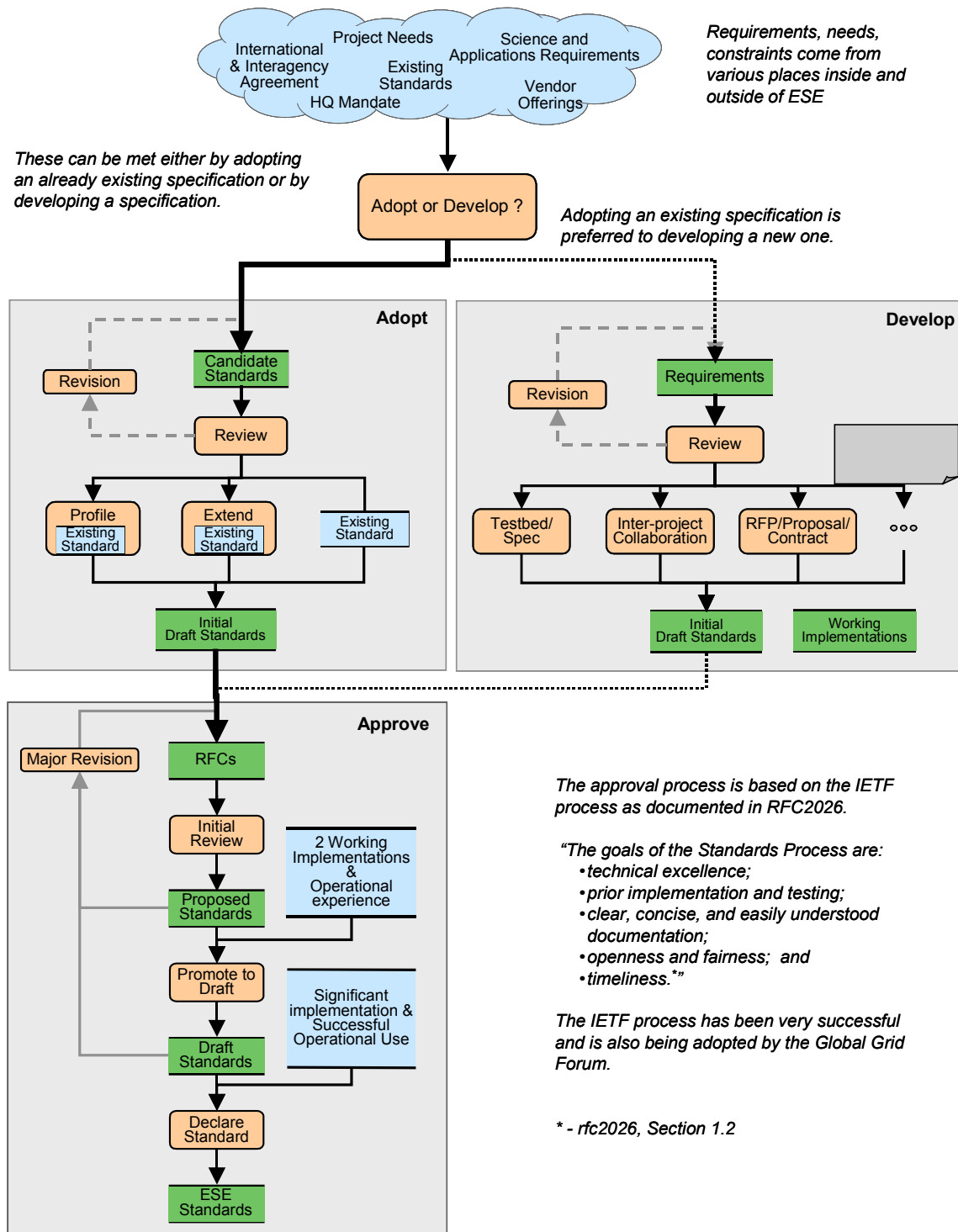


Figure 2.3-1 Process to establish ESE standards

Adopt

If a suitable standard already exists, an adoption process proposes and adopts candidate standards. The standards can be adopted with no modifications, adopted as a profile (i.e., with restrictions), or adopted with extensions. In all three cases, the candidate standards would first be reviewed to insure that there is sufficient need for the standard to be considered for adoption.

Develop

If no suitable standard exists to meet an identified need, a separate development process creates a new candidate standard. There are many possible approaches for SEEDS standards development:

- A multi-vendor testbed such as that employed by the Open GIS Consortium to develop some of its specifications.
- A collaborative process within ESE that would bring together several project teams that have a stake in the resulting standard.
- A traditional contracting approach where a vendor or research group is hired to help develop the standard.

Upon successful completion of the development phase, the draft standard should also be embodied in initial working implementations, which can be submitted into the standards approval process as part of the proposed standards.

Approve

In either case, for the developed or adopted candidate standard to become an ESE standard, it must undergo an approval process. The first step in the approval process is an “Initial Review” to select those standards that are likely to be of high quality and of widespread interest. The second step, promote to draft, requires two independent interoperable implementations. The last step, declare standard, requires many successful operational implementations. The second and last steps are subject to public review and comments. At each step, the documents generated are available for inspection by anyone in the community. Each step in the process is time bound, with a minimum and maximum review period (e.g., minimum of 6 months and maximum of 24 months). Submitting a proposed standard through these various steps ensures high quality standards, extensive testing, and minimizes both risk and cost associated with SEEDS standards.

Associated Activities – Standards Management

Furthermore there would be activities to maintain ESE standards whereby revisions or updates would be fed back into the standards process. In addition, support would be provided to users and potential users of the standard. This would include possible technical support to implementers, advice to potential users, and promotional activities advocating its use by many projects or communities. This assistance would facilitate and reduce development cost and increase standards acceptance and interoperability between systems.

2.4 Outstanding Issues/Implications

The ESE must take into consideration the following potential issues when implementing a SEEDS standards process. These issues derive from the unique circumstances of the ESE and the recommended IETF process as we presently understand it.

1. SEEDS scope

A clear understanding of the scope of the SEEDS-defined standards processes is needed. This will clarify the role and responsibilities of the various ESE offices and define which activities are supported, such as training, tool development, technical support, prototyping, implementation, and community participation in the standards process.

2. Authority issues

SEEDS needs to evaluate and recommend methods for the ESE to encourage and enforce compliance with ESE core standards.

3. Deep community involvement

SEEDS must define mechanisms for deep community involvement in refining and participating in the standards processes. The community is a source of valuable ideas, solutions, and requirements, and community participation will be vital to its acceptance by the community. The process must recognize and be responsive to a diverse community. A single standard may not apply to all missions, disciplines, and projects across the ESE. The process to establish core standards must have broad representation from the community.

4. Responsiveness to change in requirements or enabling technology

The process must be responsive to changes in requirements or enabling technology. We must recognize that standards need to evolve or risk obsolescence. At the same time the processes must be timely and efficient and produce high quality standards.

5. Requirements from multiple sources

The process to identify and vet requirements from multiple sources (headquarters, science interuse, applications, interagency and international agreements, etc.) for SEEDS needs to be developed. There is a need to foster opportunities for interagency communication and coordination, or we may not satisfy the needs of ESE. The SEEDS standards will not become widely accepted if requirements from multiple sources are not addressed.

6. Standards decision-making process

True consensus is difficult to reach in a broad and diverse community. Therefore, multiple options for decision making in the standards process need to be identified. Both true consensus and the IETF principle of "rough consensus" along with other decision methods

should be considered as candidate options for the decision making process. ESE, together with the affected communities, must determine the appropriate decision making process.

2.5 Next Steps

The separate NTMS and LTSP studies have completed their work, but the SEEDS standards process requires further definition. A merged SEEDS standards process support group composed of members of the two separate study teams will continue this work. Even broader input and deliberation is required. The REASoN CAN awardees and others will augment the group of process consultants and active study participants. Considerable work remains in order to refine and add detail to the process descriptions, address the identified issues, iterate the results of the LTSP, and support the recommendations of the NTMS.

The SEEDS standards process will direct its efforts in a number of areas. The review of data and information systems projects and formal standards organizations is complete, but the team will continue to maintain and update the LTSP report as required. As SEEDS begins transition into operation, the standards process must prepare to consider candidate ESE standards beginning with the recommendations of the NTMS study. In support of the overall transition, the standards process support group will work jointly with the REASoN CAN awardees on defining responsibilities and begin acting on these recommendations and integrating them with the broader recommendations of the SEEDS formulation.

2.6 References

- [1] A 6 to 10 Year Approach to Data Systems and Services for NASA's Earth Science Enterprise, Draft Version 1.0; February 2001; Section A.3.
- [2] Near-Term Missions Standards Recommendations, Draft Version 2.0; SEEDS Near-Term Mission Standards study team; July 30, 2002
- [3] Standards Organizations and Projects Survey Report, Draft Version 1.10; SEEDS Long-Term Standards Process study team; September 18, 2002
- [4] SEEDS Draft Standards Process Report, Version 1.7; SEEDS Long-Term Standards Process study team; October 7, 2002
- [5] Two IETF documents describe their process (RFC 2026, "Process Used By The Internet Community For The Standardization Of Protocols And Procedures", RFC 31060, "The Tao of IETF - A Novice's Guide to the Internet Engineering Task Force"1).

3.0 Cost Estimation

3.1 Purpose

The LOS and CE team was established in order to 1) Work with the science and applications communities to develop the minimum and recommended levels of service for core data sets and services required from ESE data management service providers, 2) Determine, from cost by analogy methodology, what data management services should cost, and 3) Develop a capability to perform end-to-end cost estimates for ESE data management services. This section only addresses elements 2 and 3. Element 1 appears in Section 1.0.

3.2 Members

Vanessa Griffin, GSFC, Team Lead; Kathy Fontaine, GSFC; Bruce Barkstrom, LaRC; Claude Frenner, HQ; Bud Booth, SGT; Greg Hunolt, SGT; David Torrealba, SGT; Mel Banks, SGT.

3.3 Goals and Objectives

The goal of the Cost Estimation Study was to use the recommended ranges of service levels for the various activities to be carried out by SEEDS data service providers, and to develop an approach for estimating the cost for future data service providers to deliver those services based on comparison with existing data service providers, while being only as minimally proscriptive as necessary.

The LOSCE study is an ongoing activity to develop a cost estimation capability that will enable cost trade studies by SEEDS program managers and future SEEDS data service providers. By the time this report was drafted, the primary work on identifying the range of LOS required from future service providers was complete while the effort to develop and operate a cost estimation model continued.

3.4 Approach and Community Involvement

To arrive at a minimum LOS, the study team began by examining LOS requirements from the EOSDIS system. LOS requirements for the V0 and ECS systems were analyzed and a high level set of LOS “principles” were developed. These principles are listed below. Next, the team drafted a set of LOS requirements from those principles, grouped by data function. While the group had intended to group the functions into physical data service provider types (*a la* the original NewDISS concept), community feedback from the first workshop revealed that simply assigning LOS to functions and allowing the data providers to pick the functions they would need to implement was the optimum approach to building a useful CE tool. Based on the feedback from the workshop along with feedback on the draft working papers, the study team established the LOS for the various functions. These functions and associated LOS provided the input for the cost estimation modeling.

Cost estimation for future services can best be estimated from current and recent-past costs to provide analogous functions and services. In brief, the projected workload for a new data service provider is compared to the workload performed by existing providers, and the effort required by the new provider is then estimated from the effort now required to perform a comparable workload. Estimated costs for the new provider are then obtained by projecting labor rates or commercial off-the-shelf (COTS) costs over the planned life cycle of the new provider, Therefore, the next step was to develop a database of information describing existing ESE and, as feasible, other similar data activities to establish a “comparables database” for data management services workload and effort. The database will contain baseline data for the eventual product of this study, a life cycle cost estimation tool that produces cost estimates for future ESE data activities based on the comparison with similar existing data activities.

The cost model development is an ongoing effort that will not provide initial operational capability until the end of FY 2003. As of the date this report was drafted, the study team had developed a prototype cost model and comparables database, based on a minimal case set of existing projects. The development of the cost estimation model is detailed in Section 3.5.

The findings of the CE study are explained in detail in working papers (WP) prepared by the study team. These papers, which continue to be works-in-progress, are provided in the appendices to this document and are available on the SEEDS web site, <http://lennier.gsfc.nasa.gov/seeds/>. Information relevant to this study is concentrated in WP 2 - Cost Estimation by Analogy Model and WP 4 - Data Service Provider Reference Model - Model Parameters. A future seventh white paper will provide an overview of the comparables database, comprising information obtained from existing ESE activities and other data centers.

And finally, it should be noted that the intent of this tool is to provide the ESE, and current and potential data system providers with estimation capability based on known cost drivers. We envision any interested party this tool during solicitations for data and information systems and services. We do not, however, expect that a single number, devoid of any association assumptions and trade-offs, will be used to support a given cost position. For instance, during a solicitation, the ESE would have one set of assumptions leading to a number, and the potential PI would have another. It is expected that, as is currently the case, any variances would be considered in the context of accompanying assumptions.

3.5 Cost Model Overview

The study team is in the development phase of the cost model effort. This section provides details regarding the cost model development and the comparables database. Figure 3.5-1 illustrates how the cost model is being constructed. It is important to recognize that the cost model part of the study is developing a “tool” suitable for use by the ESE and future investigators. Development of the tool parallels the SEEDS formulation study period, however, and the time needed to develop the tool extends beyond the formulation study period.

The CE tool being developed for ESE is based on a cost estimation by analogy approach, whereby the life cycle effort (staff, hardware capacity, software, facility, etc.) required to

implement and operate a future data activity, either stand-alone or as an increment to an existing data activity, is estimated based on the effort required to implement and operate similar existing data activities. The estimated effort is then turned into a cost estimate by application of expected labor rates, inflation, information technology cost curves, and other variables.

The information describing the effort required to implement and operate existing data activities is being compiled into a comparables database that will be used by the CE tool. As with all databases, the accuracy and quality of the cost estimates produced by the CE tool will be limited by the completeness and quality of the information contained in the comparables database.

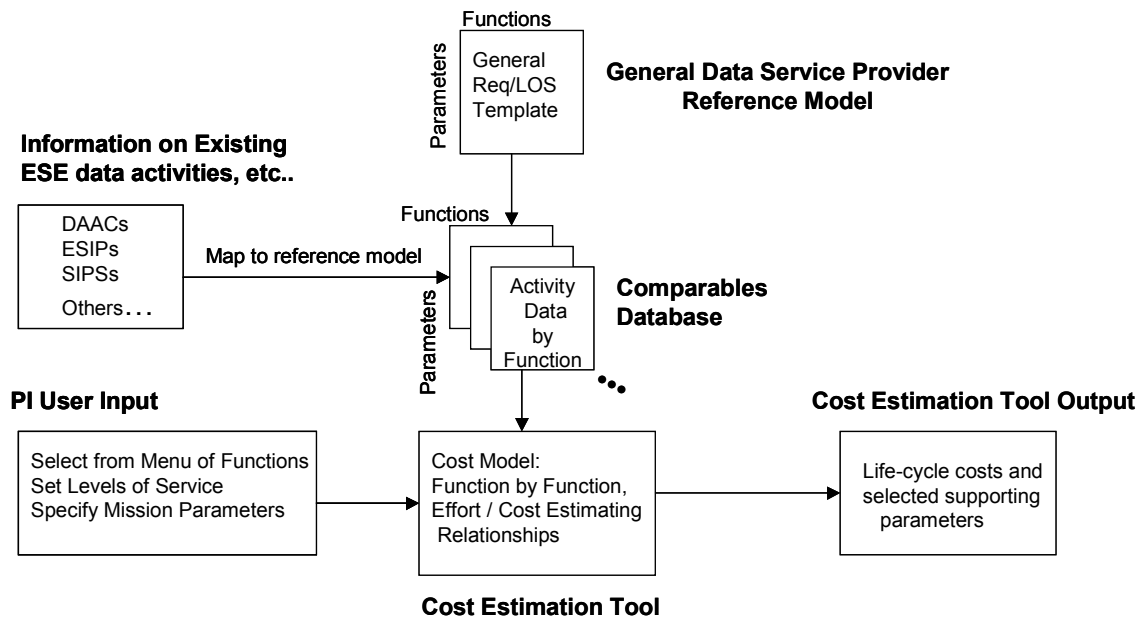


Figure 3.5-1 Cost Estimation Model Diagram

The number of available data points for the comparables database, i.e., existing data activities whose information is being collected, analyzed, and added to the comparables database, is projected to grow from 4 to 6 in October 2002, to approximately 24 by the time the CE tool is fully operational. By comparison, COCOMO II currently has 161 projects in its database. The LOSCE team will assemble the most comprehensive collection of information possible; errors of estimate will be included. The collection effort is beginning with Earth science data activities funded fully or partially by NASA, and will be extended to include other U.S. and possibly some international data activities as feasible. Some international data activities and some data activities funded wholly by other federal agencies cooperated with a previous (2001) data center operations costs benchmark study, but have not been surveyed for this effort. Commercial entities have not been surveyed as it is unlikely they would allow their proprietary information to be included in such a tool.

The study team has received feedback that basic approaches to implementing data activities are changing (e.g., from big systems supporting one or more large missions to small systems supporting single smaller missions, or from centrally developed systems deployed to multiple

sites, to locally developed systems that may share capabilities), and that this change could impact the cost-by-analogy methodology. In response, the CE study is, to the extent possible, concentrating on aspects of data activities that are, by virtue of their age or methodology, more similar to current and near-term future practice. The team is also continually updating the comparables database with the best possible information from current data activities.

As the CE tool development proceeds through a sequence of prototypes to an operational capability, the accuracy and reliability of the estimates it produces will gradually improve as the comparables database grows and as the effort and cost estimating relationships used by the model are refined.

3.6 Potential Benefits / Impact

The benefit of a useful CE tool is self-evident. The availability of a quality CE tool will allow future investigators to accurately predict their end-to-end costs for life cycle data management and service provision, and will allow the ESE and the SEEDS program to estimate the costs for future data management activities across the Enterprise.

3.7 Recommendations

1. The ESE should adopt, as an aid to ESE program staff and PIs, the life cycle cost estimation model that will be developed by the LOSCE team to estimate the cost of various types of data activities (e.g., DAACs, ESIPs, SIPSs, RESACs, project data systems). This would not be the only tool for cost estimation, but one that would be generally available.
2. The ESE should require, via appropriate language contained in each funding instrument, each funded science data and information service provider to provide actual life cycle workload and effort information for the comparables database that is used as the primary basis for cost estimation.

3.8 Next Steps

Continue development of the life cycle cost estimation by analogy tool and underlying model and increase the number of cases in the comparables database.

4.0 Data Life Cycle and Long-Term Archive

4.1 Purpose

The Data Life Cycle and Long-Term Archive (LTA) study group was established to develop a set of guidelines to manage ESE throughout the data life cycle (“cradle to grave”). These guidelines will provide the ESE mission science teams with a road map for the orderly transition of their data from production to an active archive and ultimately on to an LTA facility and hence preserve and protect the ESE investment in science objectives. The transfer of data to an LTA was perceived to be an "end of the mission" activity or the final phase in the completion or demise of a project. In the SEEDS era new science missions must plan up front for an orderly process that addresses data archiving, metadata collection, data access, and data delivery as the data progresses through its full life cycle.

4.2 Members

Previous Section Lead, Mathew Schwaller, NASA/GSFC; Current Section Lead, Ken McDonald, NASA/GSFC; Team Members: Richard McKinney, USGS/EROS/SAIC; Timothy Smith, USGS/EROS/SAIC.

Consultants:

Bruce Barkstrom, NASA/LaRC; Graham Bothwell, NASA/Jet Propulsion Laboratory (JPL); Jon Christopherson, USGS/EROS/Raytheon; Thomas Kalvelage, USGS/EROS; Steven Kempler, NASA/GSFC; Robert Wolfe, NASA/GSFC; Benjamin Watkins, NOAA/NCDC.

4.3 Goals and Objectives

The rationale behind and the motivation for the study of the Earth Science Data Life Cycle can be traced to a number of sources. NASA Policy Directives specifically call for the agency to “collect, announce, disseminate, and archive” all scientific and technical data resulting from NASA and NASA-funded research (NASA 1997). Various scientific and policy-making groups have reviewed and defined the requirements for essential data systems and services needed to ensure a long-term satellite data record in support of climate research (U.S. Global Change Research Program [USGCRP] 1999, National Academy of Sciences-Committee on Earth Sciences [NAS-CES] 2000). Recently, the Earth Observing System Science Working Group on Data (EOS SWGD, 2002) offered the following recommendations relevant to Earth Science Data Lifecycle:

- Science stewardship: NASA, in conjunction with NOAA and the U.S. Geological Survey (USGS), should determine what it is and how it works, at the various stages in the life of the LTA, who is responsible for it, and who funds it.
- The transfer of data sets to the LTA should begin as soon as feasible.

- Requirements of the individual instruments for the LTA should be communicated knowledgably to NOAA and USGS.
- Coordinated schedules and goals should be set up for working with NOAA and USGS to effect the initial LTA agreements, planning, and transfer.
- Instrument teams and DAACs should participate in advisory panels and committees within NOAA/USGS to specify and administer the LTA.
- EOS teams should identify and recommend LTA levels of service that should be provided by NOAA and USGS.
- Provision is required for NASA investigators to have ongoing access to the data sets under similar conditions to the present.

The data lifecycle approach provides the broader view of this lifecycle concept. Future mission-funding mechanisms should require more details regarding data management, data format, metadata content and collection, documentation, and other data archive transfer protocols deemed necessary and appropriate.

4.4 Recommendations

The Earth Science Data Lifecycle approach includes the identification of the data to be archived, the planning for the data acquisition and archiving, and the data ingest into the LTA. Recommendations have been developed for supporting and transitioning data through the various stages of its lifecycle and for policies that govern the entire process. The Earth Science Data Lifecycle also includes data production and reprocessing, data archiving, data distribution, user services, and finally, the disposition of data that is no longer of interest.

4.4.1 Policy Recommendations

- There should be an archive defined for each data product. Identify the active and LTA at the beginning of the mission, noting that for some products the active and LTA may be the same.
- "Data-buys" or any proprietary mission needs a time period to be defined after which the data becomes the property of NASA. Intellectual property rights issues need to be addressed.
- During the proposal process, the PI institution needs to commit/agree to a plan for the transition of its data as it progresses through its entire life cycle.
- A process should be established for the science assessment of the long-term potential of data and data products to assist in making decisions on data stewardship in a resource constrained environment.
- All archive data collections should be complete, including the archiving of the required ancillary data, project and data set documentation, and the science production software.

- Once a physical transfer of any data has occurred and been formally accepted by the archiving site, these data become the responsibility of the accepting party.
- Data may be transferred to other archives such as the National Archives and Record Administration (NARA) when deemed appropriate.
- Data should be available throughout its life cycle without loss or degradation in quality.
- Throughout a product's life cycle, a point-of-contact should be provided that can answer questions about the data or use of the data.
- Determine the relevance of Data Quality Act (67 FR 8452) and resolve issues related to watermark, provenance, reproducibility of data, peer review, integrity of data, and supporting information.

4.4.2 Mission Responsibilities

A mission is a project's time from conception, to launch, through data reception and production, to and including insertion into the active archive. It includes the planning, design, development, funding, and any other aspects of preparation. After the pre-launch and launch periods, the mission includes collecting and processing the raw data into a form usable by the science data users. Typically, a mission's life span is only a few years after which the mission ends with the materials being transferred to the LTA.

- The mission should enter into an agreement with active and/or long-term archive organizations for each of the products chosen for long-term archiving. The agreement should include detailed implementation plans, schedules, data costs, access policies, and budgets for each phase of the data life cycle, including LTA if it is a function distinct from the active archive. The agreement should contain the length of time the LTA must make the data readily available. The mission should ensure proper communication of requirements among the individual instrument teams, the active archive, and the long-term archive.
- Keep *all* pre-launch drawings, documentation, and data. Update the information and keep it accessible until transferring it to an archive.
- Set up a liaison between the pre-mission team and the archive team.
- Transfer data essential for science data processing to the active archive as soon as possible.
- Transfer spacecraft and instrument pre-launch data to an active archive in a *common* format (where *common* = a commonly used format or one that is jointly agreed). The pre-launch data must be capable of supporting the FGDC/ISO collection- and record-level standards.

4.4.3 Science Product Generation Responsibilities

The science product generation is the application of various algorithms to produce high-level products for the following scientific disciplines: land, oceans, atmosphere, hydrology, etc. The generated products are distributed primarily to the active archive centers but also can be distributed directly to users.

- A plan should be formulated to ensure that algorithms and related product information are adequately documented.
- Establish a substantive relationship with the archive for transfer of data and data stewardship.
- Establish a plan for all instrument data transfer that addresses unique aspects of individual instruments.
- Communicate requirements of the individual instrument teams for the LTA.
- Document the products, supporting data sets, production mechanisms, supporting processes, and calibration efforts.
- Set up coordinated schedules and goals to effect the initial LTA agreements, planning, and transfer.
- Ensure participation from instrument teams and advisory panels/committees in the negotiation of LTA requirements.
- Verify suitability from the instrument teams of the functions and services provided by archiving agencies:
 - Should be a standing role for “tire-kicking” and review purposes
 - Should review the prototype/early implementations to date
- The data shall be uncompressed, or, if the archive agrees, utilize a loss-less compression technique accompanied with the decompression algorithm and any additional software needed to read or decompress the data.
- The data, if transferred on media, shall reside on state-of-the-art media as negotiated with the Active Archive.
- The format in which the data is transferred should be non-proprietary and computer-compatible. If the data format is proprietary, a sunset date when the data would be considered public domain shall be negotiated and agreed upon prior to any transfer.
- The data’s file naming convention should be documented and provided to the archive.
- For all raster data sets, the transfer shall include a non-proprietary raw, raster-formatted browse for each record or granule of data along with the documentation of how the browse was created.
- The initial and any subsequent processing histories must accompany the data. The processing histories should list the hardware and software environment at the time the processing was performed.
- Complete metadata must be provided prior to any physical shipment of data and must be capable of supporting FGDC/ISO collection- and record-level standards.
- Additionally, the data shall be accompanied by a complete library of documentation, including, but not limited to, user information (e.g., guides, DIFs, fact sheets, FAQ sheets, user guides), instrument documentation (e.g., PDR and CDR information, lessons learned, hardware documentation, firmware documentation, engineering models, computer models), platform documentation (e.g., overview), algorithm documentation (e.g., ATBDs, ‘gray’ books, etc.).
- The data’s documentation shall be provided in hardcopy form where supported, and in a current softcopy format (e.g., Microsoft Word, PDF).
- To maximize the long-term utility of the data, the data provider should also provide a comprehensive training course in the science and prior management behind the instrument, data, processing histories, format(s), and algorithms.

4.4.4 Active Archive Responsibilities

The active archive is the system for processing support, archiving, documenting, and distributing the data and information for the life of the mission. The active archive's major role is to serve the day-to-day needs of the mission. At this stage in its life-cycle, the data are being regularly processed and reprocessed using on-going data validation and quality assessment results while also being provided to the broader community of science and applications users. The role of the Active Archive typically involves three components: scientific stewardship, customer service, and IT infrastructure requirements. The active archive supports the routine operations of data acquisition, data processing, data re-processing, and data staging for archive products requested for real-time and historical data from the mission. The active archive keeps track of the various processing algorithms that may be involved in the project and the appropriate metadata and browse links within the system. It is the system hub for the mission's product generation system and it is the fundamental source of information and data for LTA transfer activities.

- Establish a substantive relationship with the science product generators for transfer of data and data stewardship.
- Plan for archiving all instrument data transfer from the start.
- Capture and document archive requirements of the individual instrument teams.
- Review the documentation of the products, supporting data sets, production mechanisms, supporting processes, and calibration processes.
- Review and agree to the coordinated schedules and goals to effect the initial LTA agreements, planning, and transfer.
- Participate with the instrument teams and advisory panels/committees in the negotiation of LTA requirements.
- Work with the science production facilities to design the files to be archived to be in formats and documented in such a manner as to be easy and efficient to archive and distribute for years.
- Set up coordinated schedules and goals to effect the initial LTA agreements, planning, and transfer (avoid this being just another academic exercise).
- Ensure participation from instrument teams and advisory panels/committees in the negotiation of LTA requirements.
- Transfer data sets to the LTA beginning as soon as feasible.
- Transfers should begin early in existing/future missions with stable data sets, e.g., Level 0 and pre-launch calibration data.
- Transfer more dynamic data sets when nominally stable, and update as necessary, e.g., L1+, production software, supporting data sets, supporting software.
- Communicate requirements of the individual instrument teams to the LTA.
- Document the products, supporting data sets, production mechanisms, supporting processes, and calibration efforts.
- Allow for multiple versions of products in the LTA.
- Allow for production-on-demand systems.

4.4.5 *LTA Responsibilities*

The LTA is the archive in which stewardship of data, products, information, and documentation is held on a permanent basis or until the data is considered to be of no value. The stewardship entails preservation, maintenance, and access of the data to ensure integrity and quality of the data as the documentation indicates. The LTA is generally populated from the active archive.

- Establish a substantive relationship with the active archive for transfer of data and data stewardship.
- Plan for long-term archiving all instrument data transfer from the start.
- Communicate the long-term archiving requirements to the individual instrument teams.
- Determine where individual products should be archived (i.e., where an instrument team produces products nominally relevant to multiple archives).
- Make a plan for ensuring that algorithms and related product information are adequately documented.
- Determine how multiple versions of the LTA should be handled.
- Determine how reprocessing is handled in the longer term.
- Determine long-term handling for software that has source code maintained and built within configuration management systems such as ClearCase. Similar issues relate to migration of algorithms to different or future platforms and operating systems.
- Determine how sub-setting and other specialized services should be provided for LTA data requests.
- Determine how to handle records that need to be transferred to the LTA that are in paper form only.
- Determine how the scientific value of the data sets is maintained, and who has responsibility once the data are in the LTA.
- Confirm that LTA agencies should include all of the supporting data such as pre-launch calibration data and, if not, then alternative information may need to be provided.
- Investigate the LTA media criteria for each data set as follows:
 - Reliability: A second copy should reduce risk somewhat, but a reliable technology is mandatory. Reliability is proven over time.
 - Performance: High capacity saves significant space and high transfer rates speed up transcription.
 - Cost: The actual drive cost is fairly insignificant, but the media cost is quite important.
- Identify whether the LTA produces higher-level products or the higher-level products are provided.
- Determine how the product ordering is to function for the LTA.
- Determine requirements for metadata change once the data are moved to the LTA.
- Provide LTA access to product generation processing systems and schedules.
- Determine the need for an ongoing processing capability outside of the LTA, both for the near future and the very far future (i.e., the DAAC network of data and service providers).
- Provide science investigators with continual, ongoing access to the data sets, and without additional charges.

- Provide access to the LTA for metadata, spacecraft housekeeping and ancillary data information, engineering data, historical data, processing schedules, documentation, project policies and procedures, supported software, and documentation on data format and other standards.

4.5 Outstanding Issues/Implications

The Data Lifecycle Study has constructed a general set of recommended requirements based on input from and interactions with representatives from the community of users and providers of ESE data and associated stakeholders and from the review of relevant documents and workshop proceedings. These requirements will need a review and further iteration with a broader segment of the population of these interested parties. Therefore, the current recommendation must be considered preliminary.

The respective responsibilities of NOAA and the USGS for providing the LTA for NASA data are assumed in the requirements statements and currently the LTA requirements are to a large extent a continuation of the requirements for an active archive. However, the final set of requirements could and probably will be limited by the resources available to satisfy them. It is also safe to assume that new requirements particular to the access and use of the long-term data record may be added to the LTA requirements in the future.

4.6 Next Steps

The Data Life-Cycle Study Team will continue to interact with a broader segment of the community of ESE data users and providers. In addition, as resources allow, a Data Life-Cycle Working Group will be formed with representation from data providers, users and stakeholders to address the data life-cycle issues, expand and refine the recommendations and serve as an advisory board on data life-cycle topics. The current set of requirements and concepts will be presented and discussed at the SEEDS workshops and also at other user and provider conferences and meetings. In addition, particular effort will be directed at discussing and refining the requirements with the organizational entities that are responsible for the data at each step in its lifecycle. This will include representatives from missions, data processing facilities, active archives and LTAs. As the requirements are refined, the results of the SEEDS cost study will be incorporated to provide a better basis for the trade studies and negotiations that will be necessary for implementation. The study team will also construct data lifecycle language that can be incorporated into appropriate Announcements of Opportunity and Requests for Proposals.

4.7 References

NASA 1997. NASA Policy Directive 2220.5E Management of NASA Scientific and Technical Information (STI). August 5, 1997.

SWGd. <http://swgd.gsfc.nasa.gov>

5.0 Reuse and Reference Architectures

5.1 Purpose

This study will determine if software reuse and reference architectures can reduce the cost and improve the delivery of information services needed by future NASA Earth Science Enterprise missions, as well as increase effective and accountable involvement of the community. If so, the study will also begin to define the processes and mechanisms needed to achieve these benefits.

5.2 Members

Gail McConaughy, NASA/GSFC, Study Lead; Mark Nestler, GST; David Isaac, Business Performance Systems; Nadine Alameh, GST; Allan Doyle, Intelligent Interfaces, Inc.

The above listed team members were responsible for gathering, assembling, analyzing, synthesizing and presenting community expert opinion and interacting with the community in workshops. Identification of each individual of community member providing input is too numerous to cover here, but are mentioned by aggregate type in the full report (“SEEDS Reuse & Reference Architecture Study: Assessment of Approaches and Processes”), available in Volume II Appendices.

5.3 Reuse and Reference Architecture Definitions

Key software reuse and reference architecture approaches are defined below:

- *Clone & Own*: copying code and associated artifacts of one system for use in another system, where they may be independently modified and maintained.
- *Improved Clone & Own*: extending *Clone & Own* practices to enable developers to identify existing assets, subsequently copy those assets, and modify and integrate them more easily as needed for use in new systems.
- *Open Source*: engaging developers across missions to collaboratively develop and update selected components or systems stored in a central repository.
- *Service Encapsulation*: wrapping a complete, operational capability of a system with a network-accessible interface so that the capability can be used by others in-place, rather than re-implemented.
- *Product Lines*: reusing a set of core software components intentionally designed for a family of systems; the components are modified and maintained only by the organization responsible for the core components.

- *Notional Reference Architecture*: architecture that defines subsystems/subcomponents of a system and allocates requirements/functionality to each.
- *Specific Reference Architecture*: architecture that defines the services (including all parameters) and service invocation mechanisms of each component of a system in precise enough terms to build interfaces.
- *Coarse-Grained Reference Architecture*: architecture that defines external interfaces to major subsystems only.
- *Fine-Grained Reference Architecture*: architecture that defines internal interfaces within applications or functional components.

5.4 Community Involvement

The SEEDS formulation activity established a study to determine the opinion of the ESE community regarding the potential role of reuse in the development of future ESE data systems. The study included three steps.

- First, a range of reuse options was determined through discussions with community practitioners and industry experts. The community surveyed for this step included: NASA missions who have been successful in reuse, related NASA initiatives, related industry and government committees and consortia, other government agencies, and industry.

This broad examination developed five reuse options. These options included no change, that is status quo, and four of the options defined in section 5.2 improved clone and own, open source, service encapsulation, and product line approaches to reuse. Findings developed for each of these options are summarized in section 5.6.1.

Reference architecture options were developed. This finding is discussed in section 5.6.2.

- Second, community opinion regarding these options was solicited through workshops, surveys, and interviews. Survey responses were provided by representatives of SIPS, ESIPS, related university efforts, and DAACs.
- Finally, a process to enable reuse was characterized through community workshops.

Community opinion tended to divide along two main themes:

- Those involved in “mission-critical” activities, defined as those activities driven by launch schedules and a need for daily, highly reliable production, distribution and user services with operational promptness (e.g. SIPS, DAACs for standard products and high volume distribution), and,

- Those involved in “mission-success” activities, defined as those activities driven more by need for innovation in research in science, applications, or information systems by experimenting with differing products, approaches, and mechanisms to adapt to new understandings (ESIP-2s, -3s, scientific analysis, etc.).

Many individual community members participate strongly in both types of activities, but did “self-assign” to these groups in the workshops providing differing feedback depending on the driving activity type.

5.5 Goals and Objectives

ESE needs a more cost-effective DISS development approach for future missions because it is likely that legacy systems may well consume most of the projected ESE information systems budget. Future approaches would be more cost effective if they could leverage improvements in productivity achieved by smaller efforts developed as closely as possible to requisite expertise.

In addition, innovation needed by scientific and applications research requires a more flexible/responsive development approach. Very large development efforts require rigid requirements control to assure communication across very large staff sizes, while smaller efforts are able to respond more quickly. To leverage the community expertise and increase effective and accountable community participation, distribution of systems development and operations should be accommodated, while still assuring that such distribution retains the ability to optimize across all the efforts for the overall good (e.g., long-term data retention).

To address these issues, this study analyzes if and how reuse and reference architectures can reduce system development costs by leveraging the large base of existing ESE software, system assets, and expertise, including not just software but also its associated development artifacts (e.g., reusing test data and plans, design documents). In addition, reuse and reference architecture can enable an efficient market of components and services.

This study also analyzes if and how reuse and reference architectures can improve flexibility and responsiveness. Smaller development efforts can be effectively coordinated and integrated through the reference architecture, and assembly of new systems from reused or commodity components shortens schedules. Reference architectures can also increase community participation by enabling development to be performed wherever expert resources are available, by ensuring software interoperability of independently developed components and systems, and by providing a clear demarcation for delivered functionality.

5.6 Findings and Recommendations

The study's findings and recommendations are summarized below.

5.6.1 *Findings: Reuse*

- Both the mission-critical and the mission-success communities are not satisfied with the status quo.
- The mission-critical community is strongly in support of an Improved Clone & Own approach for mission-critical development efforts.
- The mission-success community favors the open source and service encapsulation approaches for mission-success development efforts.
- The options preferred by each community differ from the one(s) proposed to it by the other community. Community opinion was strongly against attempting a product line approach—despite the potential for higher levels of reuse—because past attempts resulted in serious cost and responsiveness problems and the requisite organizational and funding structures would be difficult or impossible to implement across the diverse ESE community.
- A cost savings sensitivity analysis performed by the study team indicates that a significant return on investment is possible if a reuse initiative could make even modest improvements in how often and how much reuse is employed in future system development efforts.

5.6.2 *Findings: Reference Architecture*

- The opinions of the ESE community regarding reference architecture alternatives were not as strong as they were regarding reuse alternatives.
- The community indicated that the ESE should develop a coarse-grained, notional reference architecture with concrete details in a limited set of functional areas.
- The community recommended against developing a fine-grained, specific reference architecture despite its theoretical benefits to component-level reuse and software interoperability because it would be too costly to develop and too constraining for use by the science community.

5.6.3 *Recommendations*

- Define and initiate community-owned, non-prescriptive, scalable and practical processes to implement community recommendations.

- Processes should start simply and evolve, emphasize directly enabling reuse over infrastructure activities, and rely on competition and peer review rather than mandates to drive reuse appropriately.
- SEEDS should competitively select and fund community reuse implementation projects and establish incentives that overcome artificial barriers to reuse.
- Establish two working groups: one focused on the improved clone & own approach in mission-critical environments, and one focused on the open source approach in mission-success environments.
 - These working groups could be responsible for recommending specific reuse initiatives, and for working in the areas of outreach and education, support/enablement, and policy change to further enable reuse within their respective communities.
 - The working groups should specifically be responsible for development of the recommended reference architecture as part of their support/enablement activities.
 - Cross working group successes would be integrated and communicated as the processes evolve.
- Establish a separate body, such as a Reuse Integration Office, whose responsibilities may include
 - prioritizing and approving reuse initiatives,
 - selecting and guiding community reuse projects,
 - administering reuse incentives,
 - conducting some reuse outreach and education activities,
 - including a small technical team to support all reuse-related activities and evaluate cost savings and impacts on schedule of the reuse functions.
- Complement the reuse effort by an effective technology development and technology infusion effort to bring in new and increased functionality:
 - Since reuse, by definition, utilizes legacy code, an isolated reuse effort may well lead to stagnation and is not recommended.

5.7 Outstanding Issues/Implications

The transition effort should determine whether intellectual property and contracting approaches may cause serious impediments to implementation. Initial analysis indicates that there are no show-stopping roadblocks, rather that current implementation of policy is slow, confusing and cumbersome to practitioners. Our approach would be to pursue these issues by working with a community prototype.

5.8 Next Steps

The following issues need further examination and elucidation in the transition development step:

- Approval for moving forward by headquarters and resolution and empowerment of an implementation organization with defined roles and responsibilities.

- Formation of the working groups with defined tasks and established metrics.
- Initiation of additional “enable and incentivize” strategies, involving
 - Policy change: examine impediments to reuse (particularly those related to intellectual property rights) and recommend workarounds.
 - Support and enablement: identify criteria and metrics for reuse, and define the structure and content of a reference architecture to enable reuse.
 - Outreach and education: facilitate the dissemination of information on the benefits of reuse and what reusable assets are available.
 - Reuse incentives: develop and institute incentives that reduce cultural and technical barriers to reuse.

Reuse projects: evaluate proposals for reuse projects and provide ongoing guidance to funded projects.

6.0 Technology Infusion

6.1 Purpose

Key to an evolving capability for Earth science data management and utilization is a continued infusion of state-of-the-practice technology advances. The initial NewDISS document repeatedly acknowledged the need to take advantage of new technologies to meet science and application demands for flexible, cost-effective data systems. The ability to discern key technology needs, based on a vision of needed capabilities as well as technological opportunities, is the first critical step toward using new technologies to help meet science and application goals. In addition, SEEDS must help ESE data system developers to incorporate new technologies into operational systems where the potential benefits can actually be realized.

A classic problem with traditional approaches to system development involves a critical gap between the development of technology and its use by implementing organizations. The SEEDS study team acknowledges this gap and proposes a focused approach to managing technology infusion for SEEDS implementers. The approach plans to leverage the contributions of the AIST component of ESTO. ESTO was established in 1998 to address technology needs for both acquiring measurements as well as using the resulting data. Therefore, the purpose of this study effort is to: 1) define processes to infuse new technologies into the evolving ESE data systems, 2) define and conduct community-based processes to identify needed capabilities and technologies, and 3) determine roles of ESTO AIST and SEEDS with regard to prototyping needs. The scope of the study also encompasses strategies for leveraging emerging technology development beyond ESTO such as technology programs at NASA for cross-enterprise use (e.g., Intelligent Systems), relevant federal programs at the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA), and industry-led endeavors such as OGC, the Global Grid Forum, etc.

6.2 Study Team Members

Karen Moe, GSFC/ESTO, Team Lead; David Isaac, BPS; Fred Brosi, GST; Vinil Patel, BPS.

6.3 Goals and Objectives

The goal of the technology infusion study effort is to enable future ESE data system evolution by leveraging technology advances. The key objective is to recommend a strategic process for SEEDS technology infusion that employs substantive community involvement to identify and prioritize technology needs, and implement a technology infusion process within the existing and planned ESE elements, i.e., flight projects, data system programs, ESTO, and SEEDS. The study team evaluated the ESTO AIST strategic planning process to assess its applicability to SEEDS, including its ability to identify technology needs and guide technology investments, and articulated a SEEDS technology planning process. Based on early recommendations from the

SEEDS community, the team plans to lead an effort to create a SEEDS vision based on community input to characterize needed capabilities, including scenarios for 2010 and beyond. Any new SEEDS results will be incorporated into the AIST technology needs database. Finally, the study team will continue to interact with the long term standards and interfaces and the reuse elements to develop a SEEDS technology infusion plan, and will research best practices and procurement options to flow from the identification of needed standards to the tools and approaches that will make incorporation cost effective. The infusion process itself will be refined and validated at a future SEEDS public workshop.

6.4 Findings and Recommendations

The technology infusion study group developed findings and recommendations based on community input primarily from the SEEDS public workshops. To initiate workshop discussions, the study group provided background material derived from current NASA technology development processes, the AIST capability needs database, and technology infusion literature and programs at other agencies. The community discussed and identified key technology infusion barriers, technology trends, capability needs, technology infusion processes and candidate strategies. Finally, the study group collected and consolidated the community input into a detailed study report, which was further distilled into the findings and recommendations below. These center on three topics: overcoming barriers to technology infusion, the role and beginnings of a capability vision, and strategic processes for infusing technology into ESE data systems.

6.4.1 Technology Infusion

Finding: Barriers to technology utilization could prevent SEEDS from realizing the benefits of technology development programs.

Recommendation: Fund efforts to bridge current gaps between technology developers and the data service providers who are potential technology users.

Community input suggests that a technology infusion process focused on information technology *utilization* is needed to enable the ESE community to better leverage the results of NASA's technology *development* programs. The community noted that substantial barriers to technology utilization exist, including the following:

- Technology development program funding does not cover tasks needed to make a technology suitable for operational deployment, nor to deploy it;
- Uncertainties surrounding the licensing of new technologies (especially those subject to commercial acquisition and those involving intellectual property rights) increases the risk of incorporating and becoming dependent on a new technology;
- New technologies can introduce performance and availability risks into operational systems, and there is no structured program to help “shake out” these risks;

- New technologies sometimes have been developed in isolation from real-world user and operational requirements; and
- The technical infrastructure assumed by a new technology is often incompatible with the infrastructure of existing operational systems.

A concerted technology infusion process could substantially reduce or eliminate these barriers. For example, systematically matching technology developers with potential early adopters and funding pilots that demonstrate the benefits of a technology in an operational environment could provide the small extra push needed to initiate widespread deployment of useful technologies. The SEEDS team recommends substantial investment in technology infusion to bridge the traditional gap between technology proof-of-concept and robust systems that are proven ready for operational deployment. A preliminary analysis suggests that the OGC's Interoperability Program and the Department of Defense (DOD) Advanced Concept Technology Demonstration (ACTD) Program could serve as models for a SEEDS technology infusion process. These processes place an emphasis on using open pilots and demonstrations to conduct collaborative technology identification, prioritization, and evaluation. The next step will be to solicit community input at the next SEEDS public workshop to understand the relevance of these process models and to further define a process specific to SEEDS.

At the core of this process should be a set of technology infusion initiatives covering a variety of small projects and activities, including operational deployment projects, deployment incentives, education/outreach activities, and support/enablement activities that are managed by a SEEDS program office and performed by the ESE community. Operational deployment projects would be focused directly on deploying new or underutilized technologies into an operational system, providing resources and incentives to enable teaming between technology providers and operational system developers. Matching technology developers with potential early adopters should not wait until a technology is fully developed, but instead should be done as early in the development process as possible to ensure technologies are aligned with real-world requirements. The ESIP Federation provides one innovative model for pairing technology providers with system developers to promote technology infusion. Other examples focus on NASA missions such as GPM SEEDS prototype and NPP. Education and outreach activities would be designed to increase awareness of select technologies to help data system users and implementers understand the potential benefits and costs. Examples might include demonstrations and workshops. Support/enablement activities would be designed to facilitate technology infusion. Examples include development of the SEEDS capability vision and efforts to address intellectual property issues.

6.4.2 SEEDS Capability Vision Development

Finding: Technology infusion initiatives cannot be effectively directed without a clearer understanding of the functional capabilities needed to enable the ESE research and application goals.

Recommendation: Develop a SEEDS capability vision that helps to capture, communicate, and refine the community's understanding of the critical capabilities that will enable the next generation of ESE research and applications.

Community input at the SEEDS public workshops indicates that the ESE should create a SEEDS capability vision to help guide technology infusion efforts. For example, some stakeholders noted that it was difficult to recommend appropriate and consistent strategies for meeting the SEEDS goals without additional consensus on what new functional capabilities are needed in order to achieve future ESE research and application goals. This will become increasingly true for technology infusion efforts, which, to be effective, must focus on those capabilities that will truly differentiate past and future systems. A capability vision would build upon and refine the basic SEEDS concepts, and begin to identify specific technical capabilities. The vision would then be used to help identify and infuse technologies that could provide the needed capabilities. In addition, the capability vision would help build the community consensus and stakeholder support needed to ensure the overall success of SEEDS. The vision could include the following elements:

- A strategic context that establishes the need from the perspective of the public (based on existing gaps and current trends) and identifies the related goals of the ESE;
- Capability themes that focus on a handful of critical capabilities needed to achieve ESE goals;
- Specific features (functional capabilities) that contribute to each capability theme, including a discussion of the current state, needed capabilities, enabling technologies, and future vision;
- Quantitative goals for each feature that enable the ESE to determine when the vision has been realized; and
- A high-level implementation plan that indicates how new capabilities will be provided, perhaps including specific recommendations from the SEEDS study teams, and areas needing research and development emphasis.

The first step is to develop a brief scenario illustrating how future data systems would contribute to achieving the ESE 2010+ vision. The FY03 recommendation is to tell a story (via animation or video) to convey SEEDS themes defined through community inputs on essential areas such as data search/access, data services, data distribution, tools and frameworks, automated operations, and security. General candidate capabilities identified by the community as important in the 2010 timeframe include the following:

- Data search/access-seamless/interactive catalog-to-data access, more direct data access, domain-specific portals, expert guided search, relevancy rankings, interactive/immersive interfaces, and content-based search;
- Data services-easier data fusion, online services (information not data), faster processing leveraging new technologies, improved formats/services for non-gridded data from nano-

sensors, better missing-value handling, improved knowledge extraction (especially for quality assurance), custom data products, and improved event detection (without reprocessing);

- Data distribution-more online data/information, faster subset processing (e.g., without decompression), direct distribution to mobile users, more direct broadcast from platforms, improved bulk data distribution, throttling and delivery prioritization, and flexible format delivery;
- Tools and frameworks- improved ease-of-use, including integration with COTS tools, improved quality labeling, custom (on-the-fly) service chaining, and high-level data manipulation (to reduce errors);
- Operations-automated operations for improved availability, improved billing, and accounting mechanisms; and
- Security-improved data access controls and policies (vs. all-or-none) and improved access to secured data.

The vision should speak to multiple audiences by illustrating from a variety of perspectives how future ESE data systems are envisioned to support the goals of ESE in the 2010 timeframe.

6.4.3 SEEDS Technology Strategic Planning Process

Finding: The ESTO process for AIST strategic planning basically fits the needs of the SEEDS objectives for technology identification and investment management; however, new strategies for infusion need to be explored.

Recommendation: Adopt and tailor the AIST processes for technology needs and gap analysis, while leveraging and tailoring community-based technology infusion processes such as those employed by the OGC Interoperability Program and the DOD ACTD Program.

The NewDISS document stressed the need to identify needed capabilities that are not now available and to facilitate the development of capabilities. In doing so, NewDISS was recognized to be “constantly changing, continually seeking the goal of optimizing performance and usability for the ever-changing aggregate of ESE data activities.” In 1999, ESTO developed the initial ESE information system needs through a public call for earth and computer scientists’ inputs for information technology of importance to ESE current and future programs. ESTO sponsors similar community workshops approximately every 15 months to keep the needs up to date. The SEEDS effort has consistently recognized the importance of cooperation and exchange between community interests and NASA management regarding evolving data systems and services. Interactive workshops, followed by analysis and feedback to the community, will continue to be used by both AIST and SEEDS to capture evolution requirements. Furthermore, ESTO has established mechanisms for monitoring technology to aid in gap analysis that can be used by

SEEDS to identify highest priority needs and enable technology infusion. These include a needs database, online technology progress reporting tools, and investment traceability to needs.

6.5 Outstanding Issues/Implications

The SEEDS vision challenge is to succinctly capture the goals of the SEEDS effort and convey the intent without getting mired in minute details and complex issues. A successful vision illustration will allow the community to unite behind the implied technology needs and work together on prioritizing and evaluating potential solutions. The technology infusion processes will enable the assessment and demonstration of successful technology approaches to achieving that vision. Examples exist for community-based standards and technology assessment, but more effort is required to tailor a process that meets the unique drivers for SEEDS, notably the differences between a mission focus vs. a science/application value-added focus. The process needs to address how to prioritize needs and innovative strategies to promote successful prototypes to operational readiness in each of these perspectives.

6.6 Next Steps

The technology infusion study team proposes that ESE adopt a policy of progressive technology infusion by funding opportunities for technology providers and system developers to team on prototype infusion efforts. A substantial budget is proposed to fund competitive solicitations, wherein the technical content is to be developed by community-based processes that identify and prioritize technology needs. A SEEDS program office would provide coordination and articulate direction and priorities, as well as manage technology infusion investments for ESE. Leveraging the ESTO AIST technology strategic planning process, the study team produced a draft SEEDS technology development and infusion plan. Findings from the first SEEDS workshop characterized the technology infusion challenge facing data system developers. Processes for identifying and tracking needs and technology exist, but the initial technology infusion processes still require considerable community input and backing.

Progress on identifying key technology needs for the near term (2010) will benefit by assessing the feedback from the second SEEDS workshop and evaluating the ESE vision video produced under the direction of Dr. M. Schoeberl, GSFC. The goal is to produce an easy-to-understand storyboard that illustrates how future data systems can evolve to support the future vision. The study team proposes to produce the video to help communicate SEEDS capabilities to the Earth science data/service users, NASA, and community managers responsible for future ESE data systems. A shared vision of how SEEDS contributes to future data system capabilities will help focus attention on the key needs for technology improvement.

7.0 Metrics Planning and Reporting and Governance

7.1 Purpose

The data systems and services supporting the ESE are distributed and heterogeneous today and are expected to be even more so in the future, given the variety of Earth science disciplines to be supported, diversity of applications' goals, and the need to foster innovation and take advantage of broadly distributed expertise. A mixture of types of providers of data systems and/or services (Data System and Service Providers, or DSPs) is needed for ESE to accomplish its goals. These DSPs perform what may be broadly classified as mission critical and mission success activities. The mission critical activities require disciplined adherence to schedules, have significant operational demands, and are required to support many "downstream" activities that depend on them. The mission success activities are important to the overall success of ESE's mission and are characterized by emphasis on innovation, permit/encourage experimentation, and have few, if any, downstream dependencies. Combinations of mission critical and mission success activities constitute the value chains required for the nation's investment in NASA and ESE to have maximally beneficial impact on society.

Regardless of the types of activities, NASA is responsible and accountable to the Office of Management and Budget (OMB), Congress, and taxpayers for their (activities') conduct and success and their contribution to the ESE's programs as a whole. In turn, each of the NASA-funded DSPs is responsible and accountable to NASA for its own success. The basic differences between the different kinds of activities imply differences in the manner in which the respective DSPs' accountability to NASA is ensured. NASA's responsibility accordingly includes ensuring that the work of the program is allocated to appropriate persons and groups, that the work of these persons and groups is *appropriately* monitored and coordinated to ensure that the collective results of all of the work fulfill the overall goals of the program. Appropriateness here implies permitting the degree of autonomy to enable the DSPs to perform at their maximum potential, and yet ensuring that the DSPs' obligations to NASA are met. Key to ensuring accountability, especially in a highly distributed, heterogeneous environment, is to employ a set of metrics to measure progress and the degree to which a DSP organization is meeting its obligations, and integrate such measures of individual accomplishments to ensure ESE's overall success. A related concern is to ensure that the governance structure in NASA permits the diverse set of DSPs to thrive and contribute maximally to the success of ESE's programs.

The purposes of this study are to:

1. Recommend a governance structure for ESE-funded data systems and services that meets ESE's needs, recognizing and accommodating the diversity of the required set of DSPs.
2. Ensure that the metrics planning and reporting processes are commensurate with the needs for accountability.
3. Recommend mechanisms to ensure that the metrics planning and reporting processes publicize accomplishments of individual DSP organizations as well as that of the ESE.

7.2 Members

H.K. “Rama” Ramapriyan, NASA/GSFC, Team Lead); Arthur (Bud) Booth, SGT, Inc.; Howard Burrows, IBM/JHU/AUSI ESIP; Bob Chen, SEDAC; Donald Collins, JPL PO.DAAC; Kathy Fontaine, NASA/GSFC; Greg Hunolt, SGT, Inc.; Frank Lindsay, GLCF ESIP, University of Maryland; Hank Wolf, SIESIP, GMU.

7.3 Goals and objectives

The specific goals and objectives of this study are:

1. Define appropriate levels of accountability for ESE-funded DSPs.
2. Recommend a governance structure accommodating the diversity of DSPs.
3. Identify appropriate solicitation opportunities and funding mechanisms.
4. Define appropriate metrics collection and monitoring mechanisms for reporting (publicizing) performance (accomplishments).
5. Recommend, to ESE, appropriate language for inclusion in various types of solicitations.

7.4 Approach and Community Involvement

The Metrics Planning and Reporting (MPAR) study team began its work in December, 2001, with H. K. Ramapriyan of GSFC and contractor support by SGT. The team began by studying the experience of ESE-funded DSPs. In order to collect and document this experience, the team began development of a questionnaire regarding the views of presently funded DSPs about funding mechanisms, accountability, and metrics collection and reporting. Questions were drafted asking DSPs about the specific funding mechanism(s) used by their sponsors, how they were held accountable by their sponsors for their performance, and the metrics they were required to provide to their sponsors. The questionnaire also asked DSPs for their evaluation of the appropriateness and effectiveness of both the funding mechanism and the metrics. In parallel, the team produced a report on the various funding mechanisms that NASA can use to fund and administer DSPs, including a summary of the conditions under which each type of funding mechanism is appropriate, given the procurement regulations under which NASA must operate.

The MPAR team presented a status report and hosted a breakout session on metrics at the first SEEDS community workshop in February 2002. After this workshop, three community representatives volunteered to join the MPAR team and were welcomed aboard. The MPAR team next completed the development of its questionnaire for the DSPs. In addition to reviewing the questionnaire, the three community members of the MPAR team provided responses for their own activities, and helped finalize the questionnaire. The questionnaire was then sent out to 30 DSPs and 18 responses were received.

Initial results from the questionnaires and progress on the MPAR study was discussed at a breakout session at the second SEEDS community workshop in June 2002. Two additional community members joined the MPAR team after the second workshop. The MPAR team

developed a white paper on governance and conducted a series of discussions by telecon on governance, with the active participation of its community members.

The governance principles and structure recommended in Section 7.5 resulted from the MPAR team discussions. The MPAR team sought a balance between the needs for flexibility and accountability, recognizing that the focus of governance must be distributed appropriately over the levels of the structure rather than concentrated at the top, and that the formal structure should not inhibit, but rather promote, spontaneous and informal collaboration by ESE elements on efforts to meet needs that may emerge.

The findings and recommendations in Section 7.6 below regarding metrics are based on the discussions at the workshops, surveys and regular telecon discussions among the MPAR team members.

7.5 Recommended Governance Structure

The basic principles used in developing the recommendation for a governance structure for ESE-funded data and information systems and services are that the structure should:

1. *Fit logically within the overall structure of the ESE science and applications program to accomplish its goals.*
2. *Provide effective monitoring and coordination of all elements of the ESE data and information services program, appropriate to the nature and assigned work of each.*
3. *Recognize and accommodate the distinct types of DSPs needed to meet the overall objectives of the program and tailor the management styles appropriately*
4. *Provide just the right amount of control, coordination and monitoring (e.g. requirements for reporting and metrics) of program elements required for overall program success.*
5. *Ensure that assignment of responsibility to program elements and relationships between program elements is clearly defined without contradictions, overlaps, ambiguities, or omissions.*
6. *Foster voluntary cooperative activities among members of the ESE data and information services community to freely and creatively respond to particular needs or circumstances, to help each other in the performance of their individual missions, and to enhance through cooperation their ability to meet ESE science and applications program needs.*

The team recommends that a SEEDS Program Office be established with a set of functions to enable bottom-up inputs from the DSPs, to coordinate reporting, to facilitate cross-DSP interactions/collaborations and to ensure that “infrastructure items” needed in common by the DSPs are developed/procured and maintained. It is expected that such an office will coexist and interact with other program offices in support of the ESE. **It is to be noted that no specific**

recommendations are made here regarding the locations (i.e., NASA headquarters or field centers and organizations within field centers) of either the SEEDS Program Office or the other program offices. These details and any adjustments to the recommended structure as a result of ongoing community feedback will need to be worked out as the transition plan is developed.

Some of the key functions of the recommended SEEDS Program Office are listed below:

- Maintain ESE data and information guidelines:
 - Recommend guidelines to be adopted by Enterprise.
 - Support the interpretation and application of the guidelines.
- Promote community involvement via working groups in each of these areas:
 - Data, information, and interface standards.
 - Metrics planning and reporting.
 - Software reuse and reference architecture.
 - Technology infusion.
 - Level of service guidelines.
- Assess cost realism of proposed products and services:
 - Develop cost estimation tool(s) (including cost by analogy).
 - Establish and maintain a database to facilitate comparisons between proposed activities and similar activities that have been completed.
- Ensure lifecycle planning for ESE funded products:
 - Ensure that lifecycle plans are maintained for ESE products.
 - Describe levels of services, identify providers and ascertain resource availability.
- Establish broadly based set of distributed service providers:
 - Support development of announcements of funding opportunities (e.g., REASoN CAN).
 - Assist the Enterprise in the selection and oversight of awardees.
- Enable technology infusion:
 - Establish vision, needs, and opportunities.
 - Promote infusion into operational systems.
- Metrics planning and reporting:
 - Gather needs for and develop/maintain/provide tools for gathering metrics that can be used in common by DSPs.
 - Report on quantity, utilization, and impact of ESE products and services.

While we are not recommending here exactly how these functions should be grouped, our initial thoughts on notional details appear in Chapter 9.

7.6 Findings and Recommendations Regarding Metrics

The findings and recommendations presented below are summarized from Section 6, “Summary of Results and Conclusions” of the “SEEDS Accountability Survey Report,” and Section 3, “Levels of Accountability” of the SEEDS MPAR Final Report (draft). These two sections are included in this team report in Volume II Appendices. The survey report is based on responses received from eighteen NASA-funded DSPs of data and information systems and services. The opinions expressed below are integrated from these responses.

7.6.1 Solicitation Opportunities and Funding Mechanisms

Finding: Current use of NASA solicitation opportunities – NASA Research Announcement (NRA), Announcement of Opportunity (AO), Cooperative Agreement Notice (CAN), and Request For Proposal (RFP) – are appropriate for funding various types of DSPs and successful in ensuring competition and fairness for the activities foreseen.

Finding: Current use of NASA funding mechanisms – grant, contract, cooperative agreement, and internal funding instrument – are appropriate for funding the various types of DSPs and successful in ensuring the necessary reporting and accountability.

The survey showed that activities with a primarily operational function supporting the ESE program (e.g., DAACs) were funded by contract, interagency agreement, or NASA’s internal processes (e.g., POP). Figure 7.6-1 depicts the process whereby the appropriate choice of award instrument is determined by the “Principal Purpose Test.” The test, when applied to a future (or even current) activity, is in itself a direct measure of the accountability expected of an activity / DSP, since the procurement mechanisms and funding vehicles are tailored to the degree to which NASA requirements are directly addressed. (Also see discussion of degrees of accountability in Section 7.6.2). It should be noted that both contracts and cooperative agreements have considerable flexibility in defining and mandating performance reporting requirements.

Recommendation #1: It is recommended that ESE not seek exceptions to the current set of NASA regulations and guidelines for solicitation opportunities and funding instruments.

Choice of Award Instrument Principal Purpose Test

If this principal purpose of a transaction is to ...

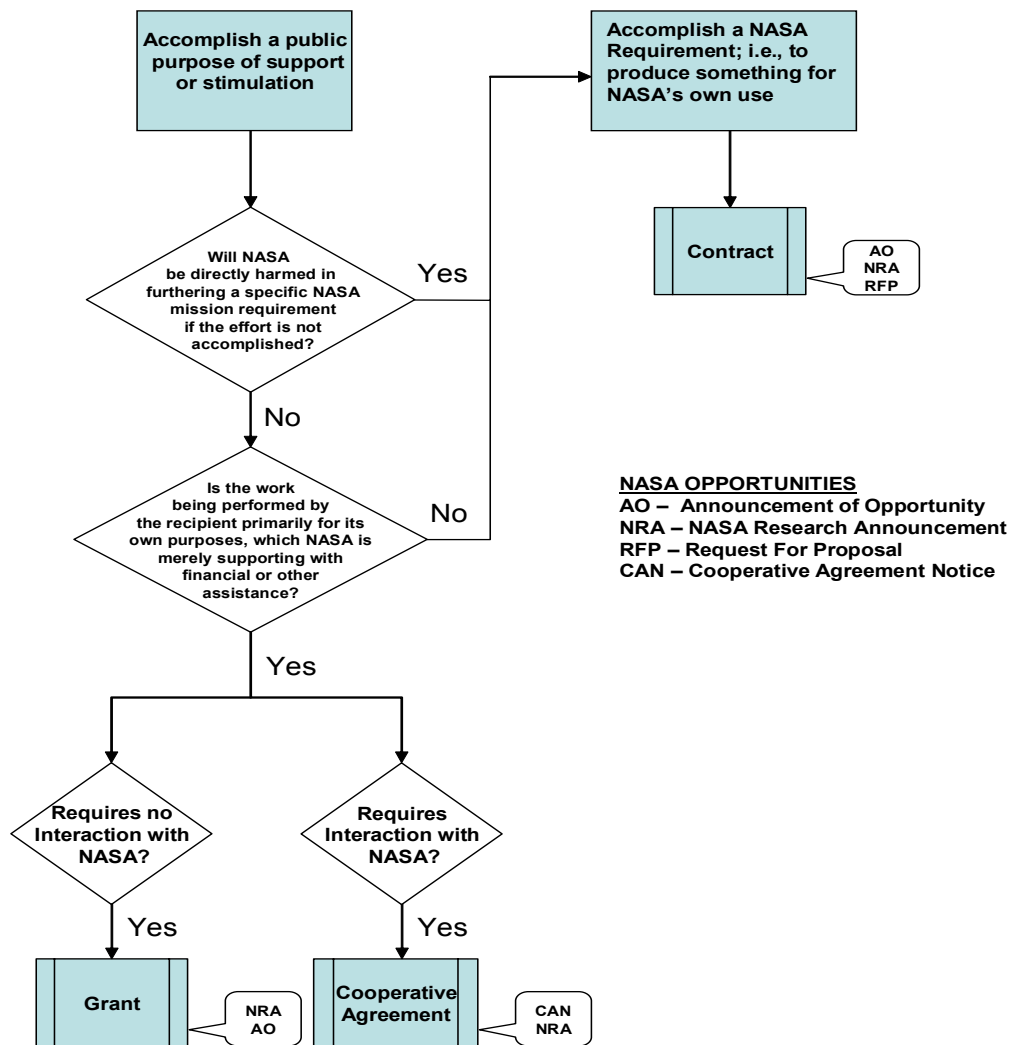


Figure 7.6-1. NASA's Principal Purpose Test

7.6.2 Levels of Accountability

Finding: A method that will define the appropriate level of accountability for an activity is the identification of the activity’s critical performance requirements. Three levels of accountability are considered, depending on five key (or core) attributes. These are shown in Tables 7.6.1, 7.6.2 and 7.6.3 for *high*, *medium* and *low* accountability levels, respectively. We could look upon high and medium levels of accountability as finer gradations of mission critical activities (sometimes subdivided as *mission critical* and *mission essential*).

Recommendation #2: It is recommended that the appropriate level of accountability for a DSP be defined by a combination of adherence to NASA’s “Principal Purpose Test,” as found in NASA Procedures and Guidelines (NPG) 58001, Part 1260.12, and implementation of the SEEDS accountability classification for DSPs as shown in Tables 7.6.1, 7.6.2, and 7.6.3. (The classification scheme is described in more detail in Appendix B, “Levels of Accountability.”). The levels of accountability required depend on the levels of service, and the metrics given in the following tables are examples of how the accountability and the levels of service could be ensured.

Both NASA funding instrument reporting requirements and a SEEDS level of accountability can be used to define appropriate metrics collection and reporting as a function of roles and responsibilities for potential DSPs.

Table 7.6.1 High Accountability for Five Data Service Provider Attributes

Timeliness Requirement	Description	Sample Metrics
Time-critical, schedule driven operations	All operations schedule-driven; near-real-time critical time constraints; all events scheduled. On-demand production with time constraints. Impact of an operational problem likely to be severe.	Percentage of ingest and production schedules met; Production backlogs; / monthly / trend
Accessibility Requirement	Description	
Search and order, data, products and services’ including user support, are public, open to all users	Services must support large, heterogeneous user community (on the order 10,000 - 100,000), high number of interactions. Problems have wide public exposure.	Profile of user base; Number of accesses; Volume data and products delivered; Volume delivered by request source; User Satisfaction metrics; / monthly / trend
Dependency Requirement	Description	
Requires ingest of satellite data streams for product processing; and creates and distributes products required by other DSPs	Ingest of Level 0, or similar satellite data streams; others depend critically on receiving your product(s) in order to perform their functions; performed on an scheduled, operational basis	Percentage of standard products delivered on time to another ESE DSP; Production backlogs; / monthly / trend
Product Quality Requirements	Description	
Products generated with peer-reviewed science algorithms; validated, provisional and beta data production supported; robust documentation, quality parameters flagged	Standard products used by users who require science-quality products in their processing and analyses.	Number/List of validated standard products (VSP) generated; Number of standard products cited in literature; Number of distinct users requesting VSP; / monthly / trend

Data Maintenance Requirements	Description	
Long-term data stewardship of Level 0 and higher data products received and generated at a DSP	Applicable to long-term data archival facilities where ongoing stewardship is critical to preserving science value of data	Volume of data and products archived by Level; Capacity analysis; Number of accesses of archival data and products > 1 year old; / monthly / trend / media type

Table 7.6.2. Medium Accountability for Five Data Service Provider Attributes

Timeliness Requirement	Description	Sample Metrics
Non-time-critical, scheduled operations	Operations nominally scheduled; time constraints are not critical; non-real-time events. While impact of a problem can be severe, there is more leeway for resolution before criticality.	Percentage of ingest and production schedules met; / monthly / trend
Accessibility Requirement	Description	
Search and order, data, products and services, including user support, are available to the science and applications community	Services focused on science and applications users (on the order of 1,000 - 10,000), can assume users have science background. Problems more contained.	Profile of user base; Number of accesses; Volume data and products delivered; User Satisfaction metrics; / monthly / trend
Dependency Requirement	Description	
Creates and distributes products for use by other DSPs	Others depend on receiving your product in order to perform their functions; could be operational or non-operational	Percentage of data or products delivered on time to another ESE DSP; / monthly / trend
Product Quality Requirements	Description	
Variable product quality; quality parameters flagged	Ad-hoc products used primarily by science team	Number/List of products provided; / monthly / trend
Data Maintenance Requirements	Description	
Pre-determined data sets and / or storage capacity limited by a specified threshold	Applicable to local working storage only, data sets may be separately archived or there may be a short-term urgency for stewardship until data sets go to archive.	Volume of data and products archived; Capacity analysis; Transfer to archive actions; / monthly / trend / media type

Table 7.6.3. Low Accountability for Five Data Service Provider Attributes

Timeliness Requirement	Description	Sample Metrics
Ad hoc, intermittent; schedule not critical	Unscheduled, non-real-time events. Impact of a problem is unlikely to be severe.	N/A
Accessibility Requirement	Description	
Search and order, data, products and services, including user support, are available to a limited team of scientists or applications specialists	Services can be customized to meet needs of small, homogeneous group of users (on the order of 20 - 100). Problems affect only this small group.	Profile of user base; Number of accesses; Volume data and products delivered; / monthly
Dependency Requirement	Description	
Creates products, but others do not depend on receiving them	Others do not depend on receiving products from you	Number/List of any ESE DSP who uses your data, products or services; /monthly
Product Quality Requirements	Description	
Quality unknown; documentation minimal or doesn't exist	Experimental products, use at own risk	Number/List of experimental products provided; / monthly
Data Maintenance Requirements	Description	
Temporary or local working storage	Interim data and products; not for archive	Volume of data and products stored; / monthly / media type

7.6.3 Metrics Collections and Reporting

Finding: The team (and survey) recognized several important metrics related to “outcome” – citations and customer “nuggets” (key success stories). There was, however, no consensus on other useful outcome metrics *per se*. Outcome metrics need to be developed that measure the value of an activity’s data and services to the science or applications community, or measure the actual utilization of data by the community. Metrics derived from the user’s point of view (i.e., easy access to readily usable, well-documented data, products, and services) still need to be defined. “Output” metrics continue to be seen as a useful measure of the productivity of an activity. There was considerable interest in establishing an enterprise function for integrated reporting of metrics and successful accomplishments at the SEEDS program level.

One DSP recommended: “Development of a systematic, cross-DAAC search for citations and data usage in the scientific, policy, and popular literature and in online information resources. Such an effort would be more cost effective and less subject to bias if conducted for all DAACs by a third party such as a SEEDS Program Office. The ‘hits’ from such a search could be tabulated quantitatively and be used as the basis for documenting significant uses of data, e.g., in an important scientific publication or significant policy decision. Such materials could then be used by the NASA Earth Observatory, the DAAC Alliance Yearbook, and other outreach efforts.” Also, metadata on source information for the publication could be required, identifying what data sets were used and from whom the data/information was obtained. This could be used for metrics development.

Another DSP noted that a SEEDS office could require ESE activities to identify papers that highlight or use their products and collect them periodically into special volumes. A SEEDS office could publish an annual report that includes a brief summary of the work of each ESE activity, plus the first page of key papers published that were based on the activity’s data.

One DSP noted that it would be useful if a SEEDS office could anticipate the metrics desired and/or required by policy makers, HQ management, and lead center technical management, and include them as part of the DSPs’ reporting requirements prior to establishing the funding agreements.

ESIPs suggested that a SEEDS office could do a number of things to help publicize accomplishments. These include:

- Advertise on NASA web sites.
- Coordinate / facilitate outreach activities, e.g. attendance at conferences or workshops by appropriate activities.
- Sponsor a special journal issue or articles in journals about ESE activities.
- Organize a conference focusing on research contributions made by activities.
- Organize workshops highlighting data tools and products developed by activities.
- Facilitate partnering with other NASA programs such as the Space Grant Education program and other educational activities.

- Proactively link NRA's, AO's, etc., with ESIP activities, emphasizing the data and services provided by or available from the ESIPs.

Recommendation #3: Because of the need to improve sponsor-required user satisfaction metrics or outcome metrics, it is recommended that this class of metrics be studied further. An extension of this study should be to identify metrics that are directly traceable to the objectives of the ESE science and applications program, so that the effectiveness of the support that ESE data management activities provide to the science and applications program can be documented, and thus the contribution of ESE data management to successful outcomes of the science and applications program can be shown. Some examples of outcome metrics (from a DSP's point of view) are given below as a starting point:

1. *Enabling faster utilization of data*

- Fraction of data sent by a DSP to the science/applications users that was actually used for addressing science questions/applications.
- Time saved in data subsetting and going to the analysis step.
- Reduction of search and access time to obtain data of interest.
- Time lag between data collection and its utilization for science/application.

2. *Size/growth of user community*

- Number of different applications communities supported.
- Number of significant new applications supported (e.g., helping with disaster relief).
- Number of value-added DSPs supported and the "user fan-out factor".
- Market share served (percentage of target user community).
- Unsolicited or solicited external requests for key services (e.g., request from a scientific group to archive and distribute its data).
- Number of new users in a year per FTE staff.

3. *Support for Publications/Education*

- Volume of data and number of datasets used to produce scientific publications.
- Number of peer-reviewed publications resulting from data.
- Number of datasets used in classroom and teaching environments.
- Number of graduate degrees resulting from the datasets sent out to educational institutions.

Finding: The Team recognizes two levels of metrics reporting, a SEEDS Program level and a DSP level.

Recommendation #4: It is recommended that the SEEDS Program Office in the governance structure discussed in Section 7.5 take on the responsibility of managing and collecting program level metrics and accomplishments as an enterprise function. It is recommended that metrics activity by the SEEDS Program Office be limited to those metrics that are required for program level assessment and monitoring, and the SEEDS Program Office not become involved with metrics that are used internally by data

management activities for their own management and monitoring. Thus the SEEDS Program Office would be involved with one set of defined metrics for ESE data and information management and services, and would obtain from each data management activity that subset of the metrics appropriate for it (e.g. metrics required from operating activities would not be the same as those appropriate for research activities). The SEEDS Program Office would maintain and update the program level metrics over time.

Recommendation #5: It is recommended that a MPAR working group (WG) be established for ongoing evaluation and evolution of appropriate metrics. The MPAR WG would also look into means of minimizing the impact of program metrics collection on DSPs. This may include exploring commonality among metrics to be reported by various DSPs and recommending/providing tools to assist in gathering, maintaining and reporting on metrics.

Recommendation #6: It is recommended that future solicitations for DSPs include a requirement for the bidders to suggest a set of metrics that demonstrate how their proposed activities will address the goals of ESE's science and applications programs and require participation by the selected DSPs in the MPAR WG. The solicitations also must require the DSPs to gather and report on an agreed upon set of metrics.

7.7 Outstanding Issues/Implications

1. The SEEDS Formulation Team believes that establishing a SEEDS Program Office addresses most of the concerns expressed in the NewDISS Strategy Document, but that there still needs to be further discussion with the community before implementation.
2. The specific allocation of responsibilities between NASA Headquarters, Field Center(s) and specific organizations within Field Center(s) is left unspecified, and needs to be determined by high-level management.
3. While the proposed approach is expected to be cost-efficient in the long run, there are expected to be initial costs for implementing the evolution to the proposed governance structure. These need to be quantified and prioritized relative to other on-going work funded by ESE.
4. The scope of the program level metrics needs to be determined.
5. An initial set of suggestions has been made for outcome metrics, but needs to be refined on an on-going basis.

7.8 Next Steps

The following next steps are recommended for the ESE:

1. Define the scope (a clear statement of what the program includes, and as needed for clarity, what it does not include), goals, and objectives of the ESE data and

information services program within the framework of the ESE science and applications program it supports.

2. Validate a recommended governance structure through science community and upper management.
3. Assign roles and responsibilities according to the agreed upon governance structure.
4. Allocate budget to support evolution to the new governance structure.

The following next steps are recommended for the SEEDS Formulation Team to pursue:

1. Develop transition plans commensurate with governance structure agreed upon by ESE.
2. Integrate the GPRA metrics considered by ESE for FY 04 with the initial set of suggestions considered here (Section 3.6.3), the metrics recommended by ESE-funded DSPs (e.g., through the REASoN CAN) and develop a set of program level metrics that measure how well the data activities support the science and applications program.
3. Organize an on-going MPAR Working Group to ensure evolution of metrics, identification of tools to simplify metrics collection and reporting.

Conduct a study of tools, especially COTS, that could support collection and analysis of program level metrics, and arrive at approaches that minimize the burden on the reporting activities.

8.0 Remarks on the Next Phase

Over the next few months, the Formulation Team will be working through transition issues, including a potential organizational structure for the SEEDS Program Office. While the discussion below does not get into a great level of detail, it is presented here to indicate our current thinking. The information below does not represent a recommendation, and will be elaborated upon in a subsequent document.

Figure 8-1 shows notional details of the SEEDS Program Office. The figure shows several working groups. Each of the working groups is to be populated by representatives from the DSPs and is coordinated by a representative from the SEEDS Program Office. Examples of international and interagency interfaces/coordination activities are participation in standards' organizations to influence evolving standards, definition of interfaces between ESE-funded DSPs and those from other agencies and countries to promote exchange of data, and development of working agreements and interface documents. Examples of infrastructure items are networks, security, catalog/directory, common user interfaces and/or capabilities to facilitate unique user interfaces, metrics gathering and integrated reporting, hosting special workshops to publicize accomplishments, and tools in support of the working groups.

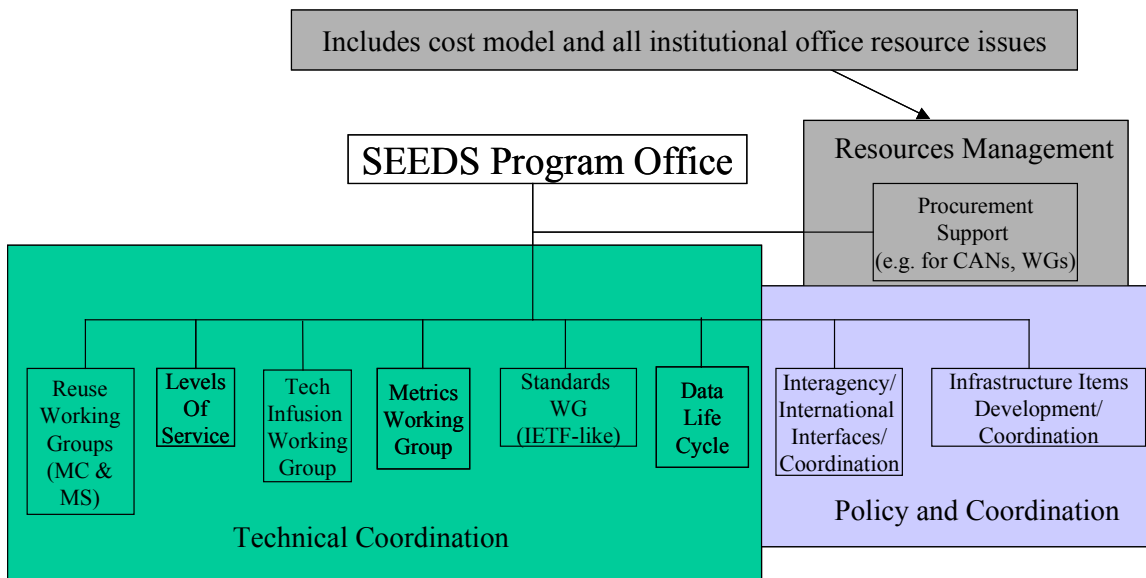


Figure 8-1 SEEDS Program Office – Notional Details

Figure 8-2 shows a few key notional inputs and outputs related to the SEEDS Program Office. A few comments about this figure are in order:

1. The term DSP is used for any entity responsible for meeting a set of requirements for providing data and information services under a funding arrangement with ESE. Some of the current examples of DSPs are DAACs, SIPs, GCMD, ESIPs (type 2 and type 3), RESACs, TSDIS, and SeaWinds/QuikScat, TOPEX/Poseidon, and SeaWiFS science data processing systems. The ECS contractor, EDOS contractor and other such entities funded by the ESE play an important support role to the DSPs by providing system capabilities, but are not covered by this definition of DSPs.
2. DSPs will be selected using competitive mechanisms appropriate for the dollar values and the type of activity (see above)
3. The mechanisms and organizational structure for the overall funding of DSPs is to be determined as a part of the development of the transition plan. It is anticipated that DSPs would receive funding from multiple sources (other program offices under ESE, NSF grants, other agencies such as NOAA, USGS and DOE) depending on the nature of their work.
4. The funding flow from the SEEDS Program Office indicated here covers **only** the activities needed to meet requirements uniquely arising from SEEDS. Examples of such activities are incremental work needed by the DSPs to: participate in working groups, facilitate reuse, infuse technology and comply with standards. It is also possible that requirements recommended through SEEDS activities will become part of ESE's program requirements that are costed and funded through other program offices and will not need incremental funds from the SEEDS Program Office. Such details are yet to be worked out as a part of the transition plan.
5. It is possible that Coordinating Entities (CEs – not explicitly shown in the figure) will be needed for:
 - Meeting a broader set of requirements than the DSPs defined above
 - Ensure that interfaces among such DSPs are defined, documented in appropriate “Inter-DSP Agreements (IDAs)”, implemented and tested
 - Ensure that data and information exchanges and other required communications among DSPs occur effectively
 - Collecting progress metrics and providing aggregated reports to the funding Program(s)
6. Two current examples of CEs are the ESDIS Project and the ESIP Federation. Of course, there are significant differences between these two examples, many caused by the nature of activities they coordinate. One significant difference is that the ESDIS Project currently has a mixture of implementing and coordinating roles. There could be several CEs within the ESE's purview. The details of this are to be determined as part of the transition planning process.
7. As indicated above, it is possible that some of the DSPs are funded by several sponsors, with NASA (ESE) being one of them. DSPs may come together as in the case of the ESIP Federation, be self-organizing and meet their overall objectives.

(This is generally represented by a “petri dish” diagram in Federation presentations). Figure 8-2 does not show this explicitly. However, the overall philosophies of competitive selection and empowerment of distributed, heterogeneous, interdependent entities, with emphasis on participatory processes for definition of interfaces and standards, are intrinsic to the SEEDS-era governance and essential for rapid response to expected, yet unpredictable changes in science, technology, and societal needs.

8. It is expected that the funding priorities and their allocations to programs will continue to be established by the ESE Associate Administrator in consultation with the program executives.
9. It is expected that ESE’s Program Executives will continue to be responsible for ensuring that the funding priorities and allocations match the strategic objectives of the ESE science and applications programs. The SEEDS Program Office ensures that, within its domain of responsibility, the performance metrics are mapped to these strategic objectives and progress is reported accordingly.

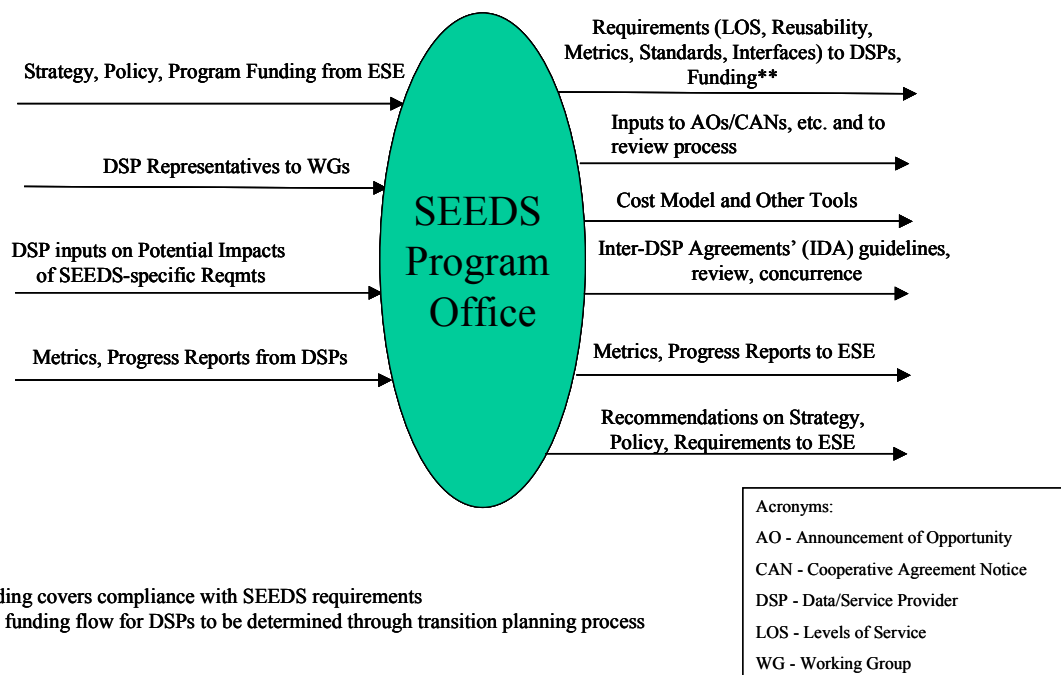


Figure 8-2 SEEDS Program Office – Notional Inputs and Outputs

It should be clear to the reader at this point that, while much has been accomplished to date, and a lot of thought has been put in to next steps, there is still a long way to go. Community involvement in this process is and remains vital to its success. SEEDS is an evolutionary process, and the state described throughout this document represents the best starting position that can be achieved to date. As the study teams migrate to

implementation working groups, and as they continue to incorporate community needs into the study processes, elements of SEEDS may stay the same, fall out completely, or dramatically change. This is and remains an evolving process, and will succeed if both NASA and the community continue the positive, productive journey that began with formulation.

Appendix for Chapter 1

1.0 Ingest Requirements / Levels of Service

- a. The data service provider shall ingest the following data [ingest data stream table, listing for each data stream: name, source, product types ingested, product type format (input and conversion after ingest if any) products ingested per day of each type, volume ingested per day]. The input data streams should cover all data to be received by the center, e.g. satellite data streams, ancillary data products, processed products generated by other data service providers, etc., based on its ESE mission, and accompanying metadata, documentation, retention plan (e.g. a part of a life cycle data management plan) etc.

Levels of Service:

- 1) operational (time-critical) ingest with immediate verification of data integrity and quality;
- 2) routine ingest and verification of data quality and integrity without tight time constraints;
- 3) ad hoc or intermittent ingest on a non-operational basis with verification of data quality and integrity;
- 4) ad hoc or intermittent ingest on a non-operational basis.

Levels of service can be mixed within a data service provider; i.e. different levels may be appropriate for different data streams.

- The data service provider shall provide standard metrics on ingest to the SEEDS Office.

1.2 Processing Requirements / Levels of Service

- a. The data service provider shall generate the following products (‘standard products’ characterized by a peer reviewed, validated, reasonably stable, ‘science quality’ processing algorithm), included required Level 1B products [standard product table, listing for each product type/series: name, format, retention plan, product instances produced per day, volume per day, required input data streams] on a highly reliable, operational basis, either on a routine schedule or on-demand, based on its ESE mission.

Levels of Service:

- 1) operational products shall be generated within 2 days of ingest/availability of required inputs;
- 2) operational products shall be generated within 7 days of ingest/availability of required inputs;
- 3) operational products shall be generated within 30 days of ingest/availability of required inputs.

- b. The data service provider shall generate the following products [product table, listing for each product type/series: name, format, retention plan, average product instances produced per day, average volume per day, required input data streams] on an ad hoc,

non-operational basis. (The product table can refer to known or expected products, or can be used to establish a capacity to support a level of ad hoc product generation (perhaps data mining or data integration) that will be used to support user needs as they arise.)

Levels of Service:

- 1) specific targets for processing adopted on a case by case basis;
 - 2) general goals for processing;
 - 3) no goals, purely ad hoc processing.
- c. The data service provider shall provide a capacity for reprocessing of standard products [standard product table] on an ad hoc basis in response to reprocessing requests.
- 1) the aggregate capacity for reprocessing shall be 9 times the original aggregate processing rate;
 - 2) the aggregate capacity for reprocessing shall be 6 times the original aggregate processing rate;
 - 3) the aggregate capacity for reprocessing shall be 3 times the original aggregate processing rate.
- d. The data service provider shall reprocess standard products [standard product table, listing for each product a reprocessing interval] according to a reprocessing schedule.

Levels of Service:

- 1) reprocessing shall be performed according to a negotiated reprocessing schedule;
 - 2) reprocessing shall be performed to meet the general goals of a nominal schedule;
 - 3) reprocessing shall be performed following a nominal schedule on a resource / time available basis.
- e. The data service provider shall accept science algorithm software from users for [product list], and perform integration and test of the software, and operational execution of the software to produce products.
- 1) the data service provider shall accept standard, research product generation software, and/or data integration and data mining software from users;
 - 2) the data service provider shall accept research product generation software and/or data integration and data mining software from users;
 - 3) the data service provider shall accept standard and/or research product generation software from users;
 - 4) the data service provider shall accept research product generation software from users;
 - 5) the data service provider shall accept standard product generation software from users.
- f. The data services provider shall be capable of cross-calibration of data from multiple sources to produce consistent product time series spanning multiple instruments / platforms.

- g. The data service provider shall provide standard metrics on production to the SEEDS Office.

1.3 Documentation Requirements / Levels of Service

- a. The data service provider shall generate and provide ESE/ SEEDS adopted standards compliant catalog information (metadata, including browse) and documentation describing all data and information produced and/or acquired and held by the data service provider.

Levels of Service:

- 1) data and product holdings (including multiple versions of products and corresponding documentation as needed) documented to the ESE / SEEDS adopted standard for long term archiving, including details of processing algorithms, processing history, many etc.;
- 2) documentation ensured to be sufficient for current use (e.g. product type descriptions, product instance (a.k.a. granule) descriptions including version information, FAQs, 'readme's, web pages with links to metadata, user guides, references to journal articles describing the production or use of the data or product);
- 3) documentation only as received from product provider.

- b. The data service provider shall update documentation of data and products with user comments, e.g. on parameter accuracy, product usability, data services available or needed for a product, etc.

Levels of Service:

- 1) data and products routinely updated with user comments;
- 2) data and products occasionally updated with user comments;
- 3) data and products rarely updated with user comments.

- c. The data service provider shall generate and provide DIF (Directory Interchange Format) documents to the Global Change Master Directory on all products available from the data service provider prior to their release for distribution.

1.4 Archive Requirements / Levels of Service

- a. The data service provider shall add to its archive or working storage the following data and products [archive product table, drawn from ingest data stream table, standard product, and ad hoc product tables and reprocessing volume] and related documentation / metadata.
- b. The data service provider shall provide for secure, permanent storage of data at the "raw" sensor level (NASA Level 0 plus appended calibration and geolocation information).
- c. The data service provider shall provide for secure storage of all standard or other science products it produces until the end of the science mission or until transfer to an approved permanent archive, per the applicable life-cycle data management plan (or separate retention plans).

- d. The data service provider shall have the capability to selectively replace archived product instances (single or large sets) with new versions, and to selectively update metadata and documentation (e.g. to update quality flags when a product is validated).
- e. The data service provider shall provide for an [archive] [working storage] capacity of [number] TB.

Levels of Service:

- 1) archive capacity is cumulative sum of all data ingested plus all products generated (including allowance for retaining multiple versions of the same product as required to provide needed support to the provider's science or applications community);
- 2) archive capacity is limited to a specified threshold.

- f. The data service provider shall perform quality screening on data entering the archive (e.g. read after write check when data is written to archive media) and exiting the archive (e.g. tracking of read failures and corrected errors or other indication of media degradation on all reads from archive media).

Levels of service:

- 1) exit and entry screening;
- 2) entry screening.

- g. The data service provider shall take steps to ensure the preservation of data in its archive.

Levels of service:

- 1) 10% per year random screening to detect and replace failing / degrading media;
- 2) 5% per year random screening;
- 3) 1% per year random screening.

- h. The data service provider shall provide a backup and restore capability for its [archive] [working storage].

Levels of service:

- 1) full off-site backup, with regular sampling and exercise of restore capability to verify integrity;
- 2) partial, [Backup Fraction - % of archive backed up], off-site backup, with sampling;
- 3) partial, [Backup Fraction - % of archive backed up], on-site backup, with sampling.

- i. The data service provider shall use robust archive media.

Levels of Service:

- 1) archive media consistent with best commercial practice;
- 2) archive media and system vendor independent;
- 3) archive media vendor independent.

- j. The data service provider shall plan and perform periodic migration of archive to new archive media / technology.

Levels of Service:

- 1) planned and budgeted for migration;
- 2) no planned migration, but ad hoc migration as need is seen to arise.
(Note - this requirement would not apply to a data service provider with a shorter lifetime than a migration cycle appropriate for its archive media / technology.)

- k. The data service provider shall provide standard metrics on archive to the SEEDS Office.

1.5 Search and Order Requirements / Levels of Service

- a. The data service provider shall provide users with access to all metadata and information holdings.

Levels of Service:

- 1) public access to all users;
- 2) access to the science and applications community;
- 3) access to a limited team of scientists or applications specialists.

- b. The data service provider shall provide a world wide web accessible search and order capability to [all users (including the general public) consistent with SEEDS standards and practices] [a limited set of science team members]. (Scope consistent with the level of service for requirement 2.5 a above.)

Levels of Service:

- 1) allow search for instances of multiple product types that pertain to a specified object or phenomenon (e.g. a named hurricane, a volcanic eruption, a field campaign, etc.)
- 2) allow search for instances of multiple product types by geophysical parameter(s), time, and space applied across multiple product types;
- 3) allow search for instances of multiple product types by common time and space criteria (coincident search);
- 4) allow search for instances of single product type by time and space criteria;
- 5) allow search for particular instances of a product type from a list of those available.

- c. The data service provider shall provide the user with the option of quickly viewing information describing any product returned as meeting search criteria.

Levels of Service:

- 1) descriptive information includes detailed algorithm and use explanations, references to a few published papers that describe the production or use of the product, standard guide and DIF metadata.

2) descriptive information includes references to a few published papers that describe the production or use of the product, standard guide and DIF metadata.

3) descriptive information includes standard guide and DIF metadata.

- d. The data service provider shall provide an interface for system-system search and order access as well as an interface for human users.
- e. The data service provider shall provide an interface to and support selected external catalog search capabilities (e.g. EDG, Mercury, Echo).

1.6 Access and Distribution Requirements / Levels of Service

- a. The data service provider shall provide users with access to all data and product holdings, including all standard science products (Level 1b, Level 2, and Level 3) produced by the data service provider.

Levels of Service:

- 1) public access to all users;
- 2) access to the science community or an applications community;
- 3) access to a limited team of scientists or applications specialists.

- b. The data service provider shall provide data and products to users in (at a minimum) one of the SEEDS core formats.
- c. The data service provider shall enhance its distribution capability with supporting data services such as subsetting, resampling, reformatting (e.g. to GIS formats), reprojecting and/or packaging to meet the needs of its users.

Levels of Service:

- 1) supporting data services available for most archived data and products;
 - 2) supporting data services available for less than half of archived data and products;
 - 3) supporting data services available for a few selected data and products only.
- The particular supporting services available would vary on a product by product basis, depending on the nature of the product and the needs of the user community.

- d. The data service provider shall provide data to users on an [operational, subscription (i.e. standing order), and/or in response to request] basis. (An operational basis means in part that a data service provider will formally commit in a level of service agreement or equivalent to terms of service.)
- e. The data service provider shall provide an interface for system to system network delivery of data and products.
- f. The data service provider shall perform timely distribution of data and products to users by network, providing an average distribution volume capacity of [number] TB per day.

Levels of service:

- 1) availability of a single product for access by user software within ten seconds;
 - 2) availability of a single product for network delivery (e.g. FTP pickup or push) within ten seconds;
 - 3) availability of a single product for network delivery within ten minutes;
 - 4) availability of a single product for network delivery within twenty four hours.
- g. The data service provider shall perform timely distribution of data and products to users on SEEDS standard media types in response to user requests, providing an average volume capacity of [number] TB per day.

Levels of Service:

- 1) shipping of media product within three days of receipt of request;
 - 2) shipping of media product within one week of receipt of request,
 - 3) shipping of media product within one month of receipt of request.
- h. The data service provider shall have the capacity to distribute products on an average of [number] media units per day.
- i. The data service provider with final ESE archive responsibility (i.e., a Backbone Data Center unless, for example, or a Science Data Service Provider which has held its products to the time for their transfer to the long term archive) shall transfer its data, products, and documentation (done to the long term archive standard) to the designated long term archive according to its Life Cycle Data Management Plan.
- j. The data service provider shall provide SEEDS standard metrics on distribution to the SEEDS Office.

1.7 User Support Requirements / Levels of Service

- a. The data service provider shall be capable of supporting [number] of distinct, active users per year who request and/or access and use data service provider products.
- 1) one user support staff member per 100 active users;
 - 2) one user support staff member per 500 active users;
 - 3) one user support staff member per 1,000 active users.
- (The number of active users is the number of distinct users who request, or through an automated means obtain, delivery of data and/or information products per year.)
- b. The data service provider shall provide a trained user support staff.

Levels of service:

- 1) below plus science expertise in data / product quality and their research uses.
 - 2) below plus technical expertise in data structures, use of tools for format conversions, subsetting, analysis, etc.
 - 3) below plus comprehensive knowledge of details of formats for most if not all products;
 - 4) user support staff are knowledgeable about the data service provider's holdings and ordering/delivery options.
- (Not all members of the user support staff would necessarily have the highest level of expertise.)

- c. The data service provider shall provide a help desk function (i.e., staff awaiting user contacts who can assist in ordering, track and status pending requests, resolve problems, etc.).

Levels of Service:

- 1) Help desk staffed seven days per week, twenty-four hours per day.
 - 2) Help desk staffed five days per week, twelve hours per day;
 - 3) Help desk staffed five days per week, eight hours per day;
- d. The data service provider shall provide on-line user support (FAQ, data, product and service descriptions, etc.).
 - e. The data service provider shall perform user outreach, education, and training.

Levels of Service:

- 1) Below plus provide user training sessions at universities, schools, etc.
- 2) Below plus expanded booth support including mini-workshops, user training sessions;
- 3) Below plus booth support at four conferences per year;
- 4) Produce and make available outreach material - pamphlets, brochures, posters, etc.

1.8 Instrument / Mission Operations Requirements / Levels of Service

- a. The data service provider shall monitor the status and performance of [name] instruments and in some cases also [name] spacecraft for which it is responsible, generating instrument commands and in some cases spacecraft commands as needed.
- b. The data service provider shall obtain the services of a NASA (or other spacecraft operator as appropriate) mission operations facility to provide instrument and spacecraft data and to receive, validate, and transmit instrument and/or spacecraft commands to the spacecraft.

1.9 Sustaining Engineering Requirements / Levels of Service

- a. The data service provider shall maintain and, as needed, enhance custom software it develops to meet its mission needs, and reused software it customizes and integrates, a total of [number] SLOC.

Levels of Service:

- 1) no or very infrequent interruptions of data service provider operations;
- 2) occasional interruptions in data service provider operations;
- 3) as needed, with interruptions in data service provider operations a secondary concern.

2.0 Engineering Support Requirements / Levels of Service

- a. The data service provider shall perform system administration, network administration, database administration, coordination of hardware maintenance by vendors, and other technical functions as required for performance of its mission.

Levels of Service:

- 1) no or very infrequent interruptions of data service provider operations;
- 2) occasional interruptions in data service provider operations;
- 3) as needed, with interruptions in data service provider operations a secondary concern.

- b. The data service provider shall perform systems engineering, test engineering, configuration management, COTS procurement, installation of COTS upgrades, network/communications engineering and other engineering functions as required for performance of its mission.

Levels of Service:

- 1) no or very infrequent interruptions of data service provider operations;
- 2) occasional interruptions in data service provider operations;
- 3) as needed, with interruptions in data service provider operations a secondary concern.

2.1 Technical Coordination Requirements / Levels of Service

- a. The data service provider shall provide staff required for participation in SEEDS processes, including ESE data services architecture refinement and evolution, and information technology planning.
- b. The data service provider shall provide staff required for participation in SEEDS processes to coordinate data stewardship standards and practices and development and maintenance of standards for content of life cycle data management plans.
- c. The data service provider shall provide staff required for participation in SEEDS processes to coordinate best practices among ESE data service providers, including quality assurance standards and practices for all phases of data services provider functions.
- d. The data service provider shall provide staff required for participation in SEEDS processes, and cooperating with other ESE data service providers in representing ESE / SEEDS in broader community processes, for developing and maintaining common standards and interface definitions, including those that enable interoperability within the ESE / SEEDS environment and with other systems and networks as needed to support the ESE program.
- e. The data services provider shall participate in SEEDS level and/or bilateral processes to coordinate production and delivery of products between ESE data service providers.
- f. The data services provider shall participate in SEEDS processes for coordinating user support guidelines and practices among ESE data services providers.
- g. The data services provider shall provide staff required for SEEDS coordination of security standards and practices to meet NASA or other established security requirements.
- h. The data service provider shall provide staff to coordinate standards for common metrics.

- i. The data service provider shall provide funding for travel to support technical coordination activities.

2.2 Implementation Requirements / Levels of Service

- a. The data service provider shall design a data and information system capable of meeting its mission requirements. The design shall address hardware configuration and interfaces and allocation of function to platform. The design shall address software configuration, including COTS, software re-use, and new custom software to be developed, including science software embodying product generation algorithms and/or software facilitating integration of science software provided by outside source(s).
- b. The data service provider shall develop a staffing plan that addresses staff required to implement and operate the data service provider over its planned lifetime. The staffing plan shall include a breakdown of positions and skill levels assigned to functions.
- c. The data service provider shall develop a facility plan, including planning for space, utilities, furnishings, etc., required to support its staff, data and information system, data storage, etc., and the environmental conditioning to be provided.
- d. The data service provider shall accomplish the implementation of its data and information system, including purchase and installation of hardware, purchase or licensing and installation and configuration of COTS software, modification, installation and configuration of re-use software, development of new custom software, and integration of all components into a tested system capable of meeting the data service provider's mission requirements.
- e. The data service provider shall perform ongoing applications software development.

Levels of Service:

- 1) Below plus implementation of applications software to perform a 'data mining' or data integration operation to meet a user need.
 - 2) Below plus implementation of product generation software embodying science algorithms, e.g. to produce a product to meet a particular user need;
 - 3) Implementation of software tools for use by users to unpack, subset, or otherwise manipulate products provided by the data service provider;
- f. The data service provider shall provide the staff needed to accomplish all needed in-house development and test activities.

2.3 Management Requirements / Levels of Service

- a. The data service provider shall provide management and administrative staff to perform supervisory, financial administration, and other administrative functions.
- b. The data service provider shall provide staff required for participation in SEEDS management processes, strategic planning, coordination with other data centers and activities beyond ESE/SEEDS.

- c. The data service provider shall provide staff with science expertise to coordinate the science activities within the data service provider and its interaction with the ESE and broader science community, including a visiting scientist program (or equivalent), collaboration among ESE data service providers to support science needs, annual Enterprise peer review, and support for its User Advisory Group and any other advisory activities appropriate given its ESE role and user community.
- d. The data service provider shall provide staff with system engineering expertise to plan information technology upgrades / technology refreshes, based on assessments of changing mission or user needs and availability of new technology. (Coordination with other ESE data service providers is included in technical coordination).
- e. The data service provider shall provide staff with data management expertise to develop data stewardship practices, perform data administration with science advice (via the User Advisory Group and other appropriate bodies), develop and maintain life cycle data management plans including data migrations. (Coordination with other ESE data service providers is included in technical coordination).

2.4 Facility / Infrastructure Requirements / Levels of Service

- a. The data service provider shall maintain site, system, and data security according to established NASA or other policies and practices while providing easiest possible access (consistent with required security) to its data and information services for its user community.
- b. The data service provider shall provide and maintain a fully furnished and equipped, environmentally controlled, physically secure facility to house its staff, systems, and data and information holdings.
- c. The data service provider shall provide a backup facility for its data and information holdings.

Levels of Service:

- 1) an environmentally controlled and physically secure off-site backup archive facility;
- 2) an on-site but separate environmentally controlled and physically secure off-site backup facility;
- 3) a backup capability within the data service provider's primary data system(s).
- d. The data service provider shall perform resource planning, logistics, supplies inventory and acquisition, and facility management.

Levels of Service:

- 1) no or very infrequent interruptions of data service provider operations;
- 2) occasional interruptions in data service provider operations;
- 3) as needed, with interruptions in data service provider operations a secondary concern.
- e. The data service provider shall provide network connections and services as needed to support its operations.