Chapter 7 Evolution of the Earth Observing System (EOS) Data and Information System (EOSDIS)

Hampapuram Ramapriyan, Jeanne Behnke, Edwin Sofinowski, Dawn Lowe, Mary Ann Esfandiari[1]

7.1 Introduction

One of the strategic goals of the U.S. National Aeronautics and Space Administration (NASA) is to "Develop a balanced overall program of science, exploration, and aeronautics consistent with the redirection of the human spaceflight program to focus on exploration" (NASA 2006). An important sub-goal of this goal is to "Study Earth from space to advance scientific understanding and meet societal needs." NASA meets this sub-goal in partnership with other U.S. agencies and international organizations through its Earth science program. A major component of NASA's Earth science program is the Earth Observing System (EOS). The EOS program was started in 1990 with the primary purpose of modeling global climate change. This program consists of a set of space-borne instruments, science teams, and a data system. The instruments are designed to obtain highly accurate, frequent and global measurements of geophysical properties of land, oceans and atmosphere. The science teams are responsible for designing the instruments as well as scientific algorithms to derive information from the instrument measurements. The data system, called the EOS Data and Information System (EOSDIS), produces data products using those algorithms as well as archives and distributes such products. The

---

first of the EOS instruments were launched in November 1997 on the Japanese satellite called the Tropical Rainfall Measuring Mission (TRMM) and the last, on the U.S. satellite Aura, were launched in July 2004. The instrument science teams have been active since the inception of the program in 1990 and have participation from Brazil, Canada, France, Japan, Netherlands, United Kingdom and U.S. The development of EOSDIS was initiated in 1990, and this data system has been serving the user community since 1994. The purpose of this chapter is to discuss the history and evolution of EOSDIS since its beginnings to the present and indicate how it continues to evolve into the future. See (Ramapriyan 2003) for a more detailed discussion of the history.

In the 1980s NASA's Earth science data were generally held by principal investigators or held at specialized data systems. Access to data and data products was limited to the individual scientist or small team responsible for generating the data. There were no policy-driven requirements for principal investigators to make their data available to other scientists or to a broader user community until the end of their missions. For the Upper Atmospheric Research Satellite (UARS) mission, NASA established a more open data policy whereby two years after the start of the mission the data were publicly available.

An even more open data policy was adopted by NASA for EOS. According to the EOS data policy, whose goal was to make the data available to a broad community, there was to be no exclusive access to data after an initial checkout period (EOS Project Science Office 1990). A set of "standard products" was defined for each of the instruments on the EOS spacecraft. The EOS instrument teams would develop these products using peer-reviewed algorithms and make them available to all users for research and applications. Recognizing the importance of data management, NASA started the Earth Science Data and Information System (ESDIS) Project separately from the projects responsible for the spacecraft and instruments. The purpose of the ESDIS Project was to develop and operate EOSDIS. Initially, there were to be two sites constituting EOSDIS – NASA's Goddard Space Flight Center (GSFC) and the United States Geological Survey's (USGS) Earth Resources Observation Systems (EROS) Data Center (EDC) for data processing, archiving and distribution. However, given the variety of Earth science disciplines covered by the EOS Program, it was not practical for two data centers to have the necessary expertise and understanding of the data needed to serve the scientific community effectively. This called for a larger number of data centers, distributed throughout the country to take advantage of existing scientific and data management expertise (Science Advisory Panel 1990).

In 1990, NASA selected several organizations in the U.S. based on scientific disciplines and heritage data management expertise. These were named Distributed Active Archive Centers (DAACs) since these data centers would provide a stable repository or archive of the EOS data products, actively manage the data in the repository, and would be distributed across the country. As the DAACs were established, it was also recognized that making heritage data more easily available to the community would be good preparation for managing the large data flows from EOS. The initial version of EOSDIS to accomplish this was called Version 0 (V0), a "working prototype with operating elements" (Ramapriyan and McConaughy 1991). Developed collaboratively by the DAACs and the ESDIS Project to improve access to existing data at the DAACs, V0 was operationally released to users in August 1994 and adopted a world-wide web (WWW) interface within three months thereafter. Also, tailored interfaces were developed by DAACs to serve their individual discipline communities.

In parallel with the V0 development, the ESDIS Project was preparing to satisfy the requirements for "big" data flows from the EOS missions through two major subsystems. The EOS Data and Operations System (EDOS) would be developed for data capture and initial (Level 0) processing. The EOSDIS Core System (ECS) would satisfy the remaining functions of flight operations (command and control of spacecraft and instruments) through its Flight Operations Segment (FOS) and the processing, archiving and distribution of the data from the EOS instruments through its Science Data Processing Segment (SDPS). The SDPS would perform all the functions past Level 0 processing of data from all EOS instruments (starting with those on the Tropical Rainfall Measuring Mission – TRMM – scheduled for launch in 1996) and would also support all the heritage data that were being managed using V0. As development of SDPS progressed it became clear that the system was too complex with too many requirements. Over the period 1995 through 1999, actions were taken to decentralize the development and simplify the system in order to meet the objectives of the EOS mission. Systems based on V0 at the DAACs were used to support TRMM. Generation of standard products from most of the instruments' data was moved to Science Investigator-led Processing Systems (SIPSs) that would be developed and operated by the respective instrument teams. An EOS Data Gateway (EDG) would be developed based on V0 IMS. The remaining functions in the SDPS were prioritized with inputs from the scientific user community, and releases of SDPS were scheduled to occur frequently with demonstrably increasing functionality with each release. This led to the successful completion of all subsystems needed to support the Terra mission (launched in December 1999) on time.

Since 1999, EOSDIS has been supporting ingest, processing, archiving and distribution of all the data from the EOS instruments and the products derived from them. The missions and instruments supported by EOS are shown in Fig. 7.1.1. The original DAAC concept has progressed to include a variety of data centers. Today, EOSDIS Data Centers hold over 2700 distinct datasets that include EOS and heritage (pre-EOS) products. A large and diverse community has become accustomed to data and information products from EOSDIS as evidenced by the number of users visiting EOSDIS web sites (over 450 thousand) and receiving more than one and a half petabytes of data in 2007. At the end of 2007, EOSDIS archives held about 3.75 petabytes of data, growing at a rate of ~1.7 terabytes per day.

An example of an EOS mission Earth science instrument is the Moderate Resolution Imaging Spectroradiometer (MODIS) flown on both the Terra and Aqua EOS mission satellites. This instrument provides data that improves our understanding of global dynamics and processes occurring on the land, in the oceans, and in the lower atmosphere. The two instances of this instrument contribute to a significant proportion of the EOSDIS data processing archive and distribution resources.

In 2005, after more than 10 years in operation it was time to re-examine lessons learned and seek significant improvements in a variety of areas. NASA established an EOSDIS Evolution Study to develop an approach and implementation plan that would begin to fulfill the objectives set forth in a vision for circa 2015.
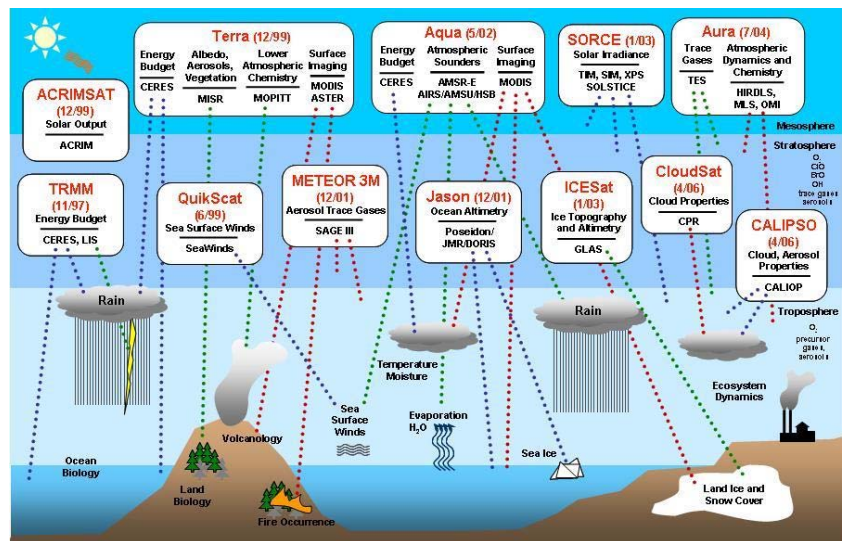
Fig. 7.1.1. Missions and Instruments Supported by EOSDIS

   The remainder of this chapter is organized as follows. Sect. 7.2 provides a discussion of EOSDIS, its elements and their functions. Sect. 7.3 provides details regarding the move towards more distributed systems for supporting both the core and community needs to be served by NASA Earth science data systems. Sect. 7.4 discusses the use of standards and interfaces and their importance in EOSDIS. Sect. 7.5 provides details about the EOSDIS Evolution Study. Sect. 7.6 presents the implementation of the EOSDIS Evolution plan. Sect. 7.7 briefly outlines the progress that the implementation has made towards the 2015 Vision, followed by a summary in Sect. 7.8.

7.2 EOSDIS and its Elements

The EOSDIS is a geographically distributed, end-to-end data system for command and control of EOS spacecraft and instruments; for receipt, capture, and Level 0 processing of telemetry data; and for production, archival, and distribution of science data. It includes the communications and administration infrastructure necessary to "glue" the system together and monitor its operation. The parts of the system that perform functions starting with command and control and ending in Level 0 processing constitute the EOSDIS Mission Systems. See Fig. 7.2.1. The remaining elements constitute the EOSDIS Science Systems maintained and operated by the NASA ESDIS Project. These science system elements were the focus of the EOSDIS Evolution study.

   EOSDIS Mission Systems monitor the EOS spacecraft and instruments and ensure that the science data reach the ground systems. The mission system Level 0 production facility, the EOS Data and Operations System (EDOS), is the primary interface to the EOSDIS Science Systems. Level 0 data reaches the science systems through the NASA Integrated Services Network.
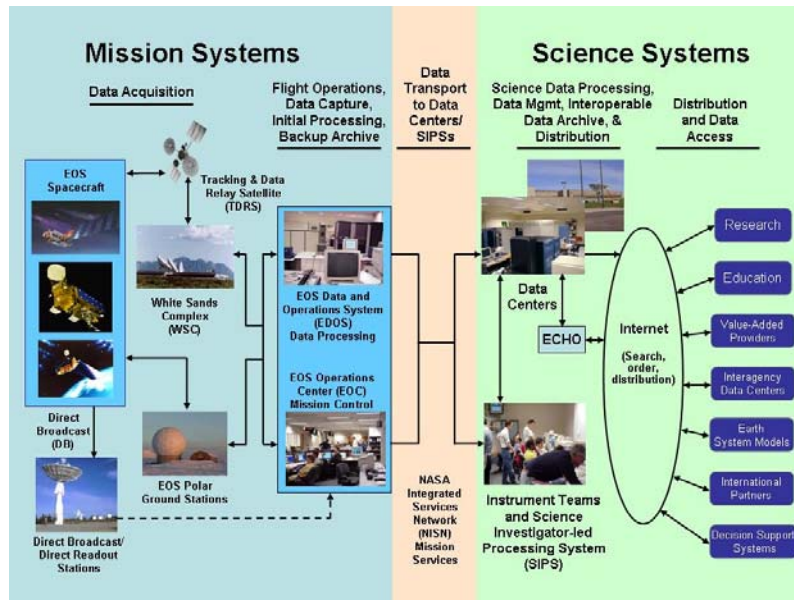
Fig. 7.2.1. EOSDIS Missions and Science Systems

7.2.1 EOSDIS Data Centers

The twelve geographically distributed EOSDIS Data Centers are collocated with other institutional facilities to achieve science synergy with the ongoing activities of those institutions. Each data center is responsible for EOSDIS data management and user services functions within a particular discipline area as presented in Table 7.2.1. These data centers are located throughout the U.S. (see Fig. 7.2.2):

The functions of the EOSDIS Data Centers are:

- Receiving EOS Level 0 data from the EOS Data and Operations System (EDOS)
- Receiving science software from EOS instrument teams and integrating it into an operational production environment
- Performing processing and reprocessing of standard data products following instrument teams' priorities
- Supporting science instrument teams as necessary in performing quality assurance of standard data products

- Ingesting standard data products produced at Science Investigator-led Processing Systems (SIPSs)
- Cataloging, archiving, and distributing EOS standard data products and other NASA Earth science data
- Providing data and information services and user support to the EOSDIS user community, and
- Preserving complete documentation of EOS data, instrument calibration, processing history, and processing source code

The EOSDIS Data Centers interface with other data centers and SIPSs for access to science data for production input, science data for archiving and distribution. The EOSDIS Data Centers' primary purpose is to interact with science data users from around the world, providing access to NASA data. Each of the EOSDIS Data Centers has a Users Working Group (UWG) consisting of representatives of the user community in its particular scientific disciplines. The UWG provides the data center with advice on the priorities for the data sets and services offered by the data center.
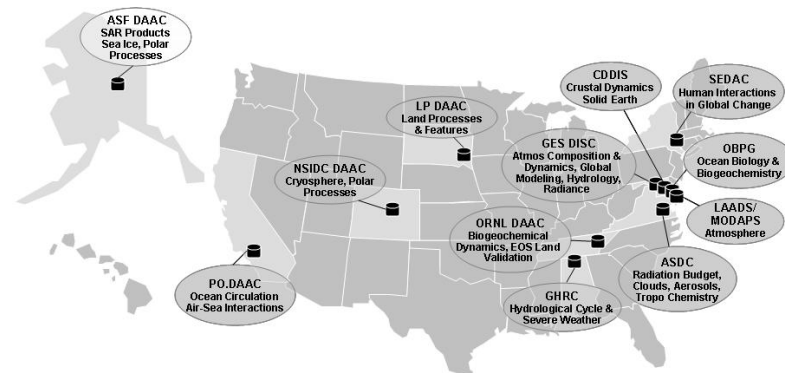
Table 7.2.1. EOSDIS Data Centers

Data Center	Location	Science Disciplines
Alaska Satellite Facility (ASF) Distributed Active Archive Center (DAAC)	Univ. of Alaska, Fairbanks, AK	Synthetic Aperture Radar (SAR) Products, Sea Ice, Polar Processes, and Geophysics
Crustal Dynamics Data and Information System (CDDIS)	NASA Goddard Space Flight Center Greenbelt, MD	Space Geodesy and Geodetics
Global Hydrology Resource Center (GHRC)	NASA Marshall Space Flight Center Huntsville, AL	Hydrologic Cycle, Severe Weather Interactions, Lightning, and Atmospheric Convection
GSFC Earth Sciences (GES) Data and Information Services Center (DISC)	NASA Goddard Space Flight Center Greenbelt, MD	Global Precipitation, Solar Irradiance, Atmospheric Composition, Atmospheric Dynamics, Global Modeling
Land Processes (LP) DAAC	USGS EROS Data Center Sioux Falls, SD	Land Processes, Land Imaging

Table 7.2.1. (cont.)

Data Center	Location	Science Disciplines
Langley Atmospheric Sciences Data Center (ASDC)	NASA Langley Research Center Hampton, VA	Radiation Budget, Clouds, Aerosols, and Tropospheric Chemistry
Level 1 and Atmospheres Archive and Distribution System (LAADS)/ MODIS Adaptive Processing System (MODAPS)	NASA Goddard Space Flight Center Greenbelt, MD	MODIS Level 1 and Atmospheric Data Products
National Snow and Ice Data Center (NSIDC) DAAC	Univ. of Colorado Boulder, CO	Snow and Ice, Cryosphere, Climate Interactions and Sea Ice
Oak Ridge National Laboratory (ORNL) DAAC	Department of Energy Nashville, TN	Biogeochemical Dynamics, Ecological Data, and Environmental Processes
Ocean Biology Processing Group (OBPG)	NASA Goddard Space Flight Center Greenbelt, MD	Ocean Biology, Sea Surface Temperature, and Biogeochemistry
Physical Oceanography (PO) DAAC	Jet Propulsion Laboratory (JPL) Pasadena, CA	Sea Surface Temperature, Ocean Winds, Circulation and Currents and Topography and Gravity
Socio-Economic Data Applications Center (SEDAC)	Columbia University Palisades, NY	Human Interactions, Land Use, Environmental Sustainability, Geospatial Data, Multilateral Environmental Agreements

Fig. 7.2.2. EOSDIS Data Centers

7.2.2 Science Data Processing Segment

The Science Data Processing Segment (SDPS) performs information management and data archiving and distribution at each data center location. Each data center performs these functions using a combination of standard capabilities provided by the ESDIS Project and hardware and software specific to the data center. Special SDPS hardware and software, known as the EOSDIS Core System (ECS), was developed to support the high ingest rates of the EOS instruments. ECS currently resides and operates at three data centers: the Langley Atmospheric Science Data Center (ASDC), the Land Processes Distributed Active Archive Center (LP DAAC) and the National Snow and Ice Data Center (NSIDC). Data products are processed by the SIPSs or, in a few cases, by systems interfacing with the SDPS at the data centers. The SDPS at the data centers ingests the data from the processing systems and archives them. The SDPS has interfaces with the EOS Clearing House (ECHO) to provide search and access through ECHO clients, such as the Warehouse Inventory Search Tool (WIST). The SDPS also provides software toolkits to assist instrument teams in their development of product generation software at their Science Computing Facilities to facilitate ingest of the resulting products into SDPS or into data center-specific archiving and distribution systems.

7.2.3 Science Investigator-led Processing Systems (SIPSs)

Most of the EOS standard products are produced at facilities under the direct control of the instrument Principal Investigators/Team Leaders (PIs/TLs) or their designees. These facilities are referred to as Science Investigator-led Processing Systems (SIPSs). The SIPSs are geographically distributed across the United States and are generally, but not necessarily, collocated with the PIs/TLs' Scientific Computing Facilities. Products produced at the SIPSs using investigator-provided systems and software are sent to appropriate EOSDIS Data Centers for archiving and distribution. Level 0 data products and ancillary data that begin the processing sequence are stored at the data centers and retrieved by the SIPSs. The geographic distribution of SIPSs is shown in Fig. 7.2.3.
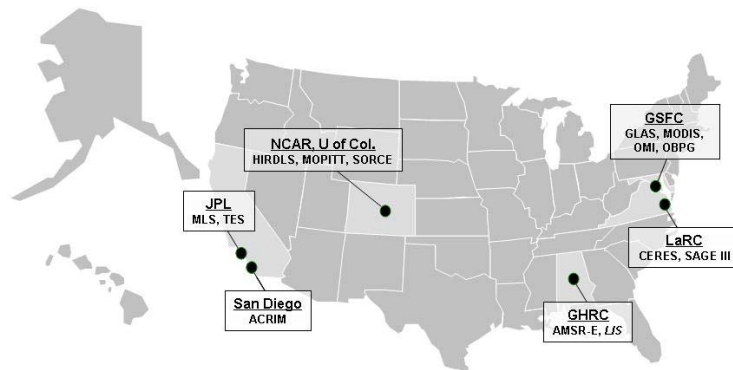

Fig. 7.2.3. EOSDIS Science Investigator-led Processing Systems

7.2.4 EOS Clearing House (ECHO)

EOSDIS provides convenient mechanisms for locating and accessing products of interest. The "look and feel" of the system is intuitive and uniform across the multiple nodes from which EOSDIS can be accessed. EOSDIS facilitates collaborative science by providing extensible sets of tools and capabilities that allow investigators to provide access to special products (or research products) from their own computing facilities. The EOS Clearing House (ECHO) is a system developed by the ESDIS Project to provide a centralized spatial and temporal metadata collection of

EOSDIS data. The fundamental principle of ECHO is to provide a central access path for any user interface developer, whether a NASA data system or any organization outside of NASA.

ECHO is the EOSDIS metadata framework by which EOSDIS keeps track of its vast data collection. ECHO is the middleware between EOS data and science data users via a service-oriented architecture. Data Partners provide metadata for their EOS data holdings and other Earth science-related data holdings. Client Partners develop software ("clients") to give science data users access to ECHO's registries using ECHO's open Application Programming Interfaces (APIs). Science data users search ECHO's registries and access data and services using an ECHO client. All of the EOSDIS Data Centers participate in ECHO by providing metadata to the ECHO database. One of the first user interfaces to be developed using ECHO is the Warehouse Inventory Search Tool (WIST), which provides web-based "one-stop shopping" for search and order capabilities within all of ECHO's data holdings. For more details about ECHO, see the ECHO web site (ECHO 2008).

7.3 Community Push towards Distributed Systems

Since the early years of EOSDIS there has been a push in the scientific community, with members from within and outside NASA, for a distributed implementation – from the points of view of both geography and responsibility. Two major influences from the community have introduced change in NASA's Earth Science data systems including EOSDIS over the last decade. These are recommendations from the National Research Council (NRC 1995), and the New Data and Information Systems and Services (NewDISS) Strategy Team (Maiden et al. 2000).

In the mid-1990s, there was significant community concern about the centralized nature of the development of EOSDIS and doubts about its being able to meet all the requirements to satisfy the broader user community beyond the scientific researchers. The National Research Council reviewed EOSDIS in 1995 and recommended that the science data processing, archiving and distribution should be performed by a "federation of competitively selected Earth Science Information Partners (ESIPs)" (NRC 1995). In response to this recommendation, NASA initiated an experiment in 1998 with a "self-governing" federation consisting initially of 24 competitively selected ESIPs, one-half (called Type 2 ESIPs) responsible for specialized research products and the other half (called Type 3 ESIPs) for products suitable for applications with commercial potential. The DAACs,

whose primary responsibility was schedule-driven operational production and support of large user communities, were later included in the federation as Type 1 ESIPs. Initially sponsored by NASA, the ESIP Federation now consists of more than 110 members including NASA and NOAA data centers, research universities and laboratories, educators, technology developers and commercial and non-profit organizations. The Foundation for Earth Science was established in 2001 as a coordinating organization that promotes the objectives of the ESIP Federation, namely, bringing the most current and reliable data products based on satellite data to a broad range of users and ensuring their utilization to address environmental, economic and social challenges of the world. Details about the ESIP Federation can be found on their web site (ESIPFED 2006).

NASA also commissioned, in 1998, the NewDISS Strategy Team with the charter to "define the future direction, framework, and strategy of NASA's Earth Science Enterprise (ESE) data and information processing, near-term archiving, and distribution." This team made a number of recommendations on how to proceed with ESE data and information systems and services over 6 – 10 years beyond the year 2000 (Maiden et al. 2000). The recommendations from the NewDISS Strategy Team are summarized below:

- Support a spectrum of heterogeneous technological approaches to NewDISS. This includes concentrating on integrating suitable existing data service capabilities, while also identifying and providing a means for delivering capabilities that do not yet exist.
- Clearly define the components of NewDISS, and ensure suitable management of the interfaces between them. This includes the definition of a set of "core" standards and practices, along with the means for selecting and maintaining them.
- Employ a NewDISS infrastructure that includes active liaison with service providers both within NASA and within the private sector for procurement of common operations activities.
- Employ competition and peer review in the process used for choosing NewDISS components.
- Empower science investigators with an appropriate degree of responsibility and authority for NewDISS data system development, processing, archiving and distribution.
- Use lessons learned from the current, experimental ESE federation as a step towards the NewDISS, and proceed with the Federation Experiment with this evolution in mind.
- 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.

Addressing these recommendations, NASA initiated a formulation study called Strategic Evolution of Earth Science Enterprise (ESE) Data Systems (SEEDS) during 2002-2003. NASA's GSFC conducted this study with significant involvement by the scientific user community. The focus of this study was on how a system of highly distributed providers of data and services could be put in place with community-based processes and be managed by NASA. The areas considered in this study were: levels of service and costs, near-term mission standards, standards and interfaces processes, data life cycle and longterm archive, reference architectures and software reuse, technology infusion, and metrics planning and reporting. As a result of the recommendations from this study, NASA established a set of four Earth Science Data System Working Groups (ESDSWG): Standards Processes, Reuse, Technology Infusion, and Metrics Planning and Reporting. The EDSWG continues to meet in groups and in plenary to promote integration of standards and capabilities into NASA's Earth Science Systems.

NASA views its data systems in terms of "Core" and "Community" capabilities. The core capabilities provide the basic infrastructure for robust and reliable data capture, processing, archiving and distributing a set of data products to a large and diverse user community. Examples of core capabilities are: 1. The Earth Observing Data and Information System (EOSDIS); 2. The Precipitation Processing System (Stocker 2003); 3. Ocean Data Processing System (Feldman 2007); and 4. The CloudSat Data Processing Center (NASA and CSU 2007). EOSDIS is a multi-mission data system that manages data from all of the EOS missions and most of the heritage (pre-EOS) missions. The Precipitation Processing System is recently evolving as a measurement-based system from the Tropical Rainfall Mapping Mission Science Data and Information System (TSDIS) and is planned to support data management for the Global Precipitation Mission (GPM). The Ocean Data Processing System, managed by the Ocean Biology Processing Group at NASA GSFC, is a measurement-based system that spans several missions ranging from Nimbus-7 to EOS. The CloudSat Data Processing Center is a system specific to CloudSat, one of the missions in the Earth System Science Pathfinder (ESSP) program. The latter three examples are "loosely coupled" with EOSDIS in that they exchange data with the EOSDIS Data Centers and are consistent with EOSDIS in the use of data format standards. In contrast to the core capabilities, community capabilities provide specialized and innovative services to data users and/or research products offering new scientific insight. Such systems are generally supported by NASA through peer reviewed

competition. Examples of community capabilities are projects under the Research, Education and Applications Solutions Network (REASoN), Advancing Collaborative Connections for Earth System Science (ACCESS), and Making Earth Science Data Records for Use in Research Environments (MEaSUREs) Programs.

Both core and community capabilities are required for NASA to meet its overall mission objectives. The focus of the ESDSWGs is on community capabilities. While the membership on the four working groups is open to all, the primary participation is by members of the REASoN, ACCESS and MEaSUREs projects. The working groups are a mechanism through which the community provides inputs for NASA to help with decisions relating to Earth science data systems. There is significant commonality in membership between the ESDSWG and ESIP Federation, thus bringing into the NASA Earth science data systems a broad community perspective.

7.4 Use of Standards and Interfaces in EOSDIS

The development and use of standards within the EOSDIS architecture has been one of the real success stories of the ESDIS Project. Standards play a critical role in how EOSDIS will serve to meet future needs. By adopting standards, we hope to foster inter-organizational data discovery and manipulation. To be useful and effective, standards must always be reviewed and modified. The ESDIS Project has always made a resource commitment to maintain and develop standards. The ESDIS Project has also opened the doors to the greater community by providing mechanisms to discuss and integrate standards into the EOSDIS. Early adoption of community standards by EOSDIS has proved to be a cost benefit to the ESDIS Project by allowing easier integration of new missions into the baseline, by reducing the complexity of the system, and by reusing existing software and processes and by enabling easier cross-training across EOSDIS.

EOSDIS has several ongoing standards activities. These include:

- Direct standards such as data format and metadata standards
- Standard usage of terms and documentation
- Standardized processes and metrics

7.4.1 Data Formats

EOSDIS has fostered the development of several standards used within science data processing systems. Historically, the format of data products

was picked by the principal investigators of each individual science instrument based on convenience and cost benefit to the processing teams. In order to facilitate the ability for diverse communities to use data in interdisciplinary studies, early in the development of EOSDIS the ESDIS Project conducted a collaborative study with the EOSDIS Data Centers of the then available standard formats for adoption in EOSDIS. None of these formats met all of the requirements. However, the Hierarchical Data Format (HDF), developed by the National Center for Supercomputing Applications (NCSA) at University of Illinois, satisfied most of the requirements. Therefore, the ESDIS Project selected this to be the data format (actually, a data formatting system with associated software tools) to be used for archiving and distributing data products for EOS instruments. The ESDIS Project has been supporting the maintenance and evolution of this formatting system, first at the NCSA and later at the HDF Group (THG). The EOSDIS Data Centers also maintain heritage data in other (native) formats, and provide format translations to users as needed.

The HDF is a multi-object file format that facilitates transfer and manipulation of scientific data across multiple systems. It supports a variety of data types. The HDF library provides a number of interfaces for storing and retrieving these data types in compressed or uncompressed formats. HDF files are self-describing and permit users to understand the file structures from information stored in the file itself. However, the traditional HDF file structure does not include geolocation information. Since it is critical for Earth observation data to be geolocated, the ESDIS Project developed the HDF-EOS format that included additional conventions and data types for HDF files. The three geospatial data types supported by HDF-EOS are: Point, Grid and Swath. Using the HDF format, the ESDIS Project took an additional step to identify three ways of looking at EOS instrument data products: point products, grid products and swath products. The standard HDF tools can also read HDF-EOS files. However, the HDF-EOS library provides software for easier access to geolocation data, time data and product metadata than the standard HDF library.

A key feature of these three product types in HDF-EOS is the identification of core metadata values that must accompany all products for inclusion into EOSDIS. Each of the EOS science data teams is required to submit data in the HDF-EOS format, with waivers provided only where justifiable. The Project provides many avenues of assistance to facilitate the acceptance of this standard by users including user guides, specialized software libraries, forums, websites and a yearly HDF Workshop held in areas across the U.S. Because the data format is widely published, the community is able to propose and develop tools to read and manipulate EOSDIS data. Two types of the most popular tools are subsetters and re-

projection tools. More details on HDF and HDF-EOS can be found on the HDF web site (HDF 2008).

7.4.2 Metadata Standards

EOSDIS has a strong commitment to metadata standards. Twenty years ago, the concept of deriving metadata from the actual data was considered burdensome to the science data producer community. Despite this initial resistance, the ESDIS Project created the EOSDIS Core data model which describes a standard set of metadata that are required for each data collection and products within the collection. Standard metadata required from the data providers include such basic information as product name, type, collection information, time of acquisition, and geographic coverage. The core data model was developed while the U.S. Federal Geographic Data Committee (FGDC) was developing the metadata content standard to be followed by U.S. agencies. The extensions of the FGDC standard for remote sensing metadata have been influenced by the EOSDIS Core data model (FGDC 2002).

This basic requirement has served to enable the development of a rich set of user interfaces and data discovery tools. All EOSDIS metadata are accessible through the EOSDIS Clearing House (ECHO) interface. ECHO has public application programming interfaces that allow access to the metadata, which are published in XML format. Use of the XML format is another standard adopted that allows for easy access by all types of world wide web interfaces.

EOSDIS Data Centers also use the Open GIS Consortium (OGC) web services. OGC standards have a particular affinity to geolocated data and are beneficial to users of many data products offered by EOSDIS. EOSDIS is starting to implement two particular web services: the Web Mapping Service (WMS) and the Web Coverage Service (WCS). As more data at the EOSDIS Data Centers are made available in WMS/WCS, users will be able to layer many types of NASA data on geospatial information systems. Use of Google KML files to layer EOS data on GoogleEarth is another standard that EOSDIS is adopting.

7.4.3 Terms and Documentation Standards

No discussion of standards within EOSDIS would be complete without a discussion of the usage of standardized terms and documentation. For example, the term "granule" was adopted early by the ESDIS Project to mean the smallest instance of a data product tracked in the data base for search-

ing, ordering and/or access. A granule can be one or more files. The concept of the granule is now universally understood within NASA Earth science communities.

The term "browse" is another standard fostered in EOSDIS. "Browse" data is now commonly understood to be small thumbnail images of the actual data. Access to browse data enables users to examine the dataset for desired features prior to the potential time-consuming step of downloading large datasets.

The ESDIS Project also focused on providing standard approaches to documentation. Every data collection in the system includes the Directory Interchange Format (DIF) registration which enables search from the NASA Global Change Master Directory (GCMD). Collections also include a standard guide document to the data set.

7.4.4 Process Standards

The ESDIS Project has developed several standardized processes to facilitate the configuration control and the management of the EOSDIS. Standardized processes are uniformly applied across EOSDIS elements. Data processing teams at the SIPSs and EOSDIS Data Centers participate in preparing and reviewing interface control documents and other related documentation. The ESDIS Project established the management tools and processes early in the ESDIS Project lifecycle to apply a routine approach to reviewing and changing documentation associated with EOSDIS. All project plans, requirements, and interface control documents are accessible on the ESDIS Project web pages.

Capturing system performance metrics is another example of a uniform process applied across the elements of the EOSDIS. Metrics such as product distribution, archive size and data center web activity are defined and reviewed at the ESDIS Project level and each data center provides a standard set of measurement inputs to the Project. Common metrics are then available not only to ESDIS management, but also to the metrics providers. Better project management is enabled by allowing the data centers access to their detailed metrics, at the same time allowing the ESDIS Project to have a system view across all of EOSDIS.

7.4.5 Standards to be Developed

While the EOSDIS has made great progress toward the introduction and common usage of standards, areas for improvement exist. We would like to see the development of "provenance" standards to provide the ESDIS

Project more complete information concerning the source and make-up of datasets. Provenance standards include the identification of information needed for the long term archive of datasets and associated material (e.g., documentation). The need for provenance standards is critical to establish both the heritage and quality assessment of the data.

Another area where standards still need attention is in the selection of dataset map projections. Despite efforts at coordination in the early years of the EOS mission design, each science instrument team was allowed by the EOS Program to determine the best projection and scale to be used for its data. Consequently, many differing projections are used for EOSDIS data. This makes it difficult for users to integrate and inter-use data from multiple instruments or disciplines.

7.5 Evolution of EOSDIS Elements Study

In late 2004, NASA Headquarters management initiated a review of the EOSDIS. NASA prepared a charter for the "Evolution of EOSDIS Elements (EEE) Study" (Cleave 2004) with the goal to "assess, by considering the future objectives, the current state of EOSDIS in order to identify the components that can/must evolve, those components that need to be replaced because of the rapid evolution of information technologies, and those components that require a phase-out strategy because they are no longer needed." The charter advanced objectives for the study as:

- Increase end-to-end data system efficiency and operability
- Increase data usability by the science research, application, and modeling communities
- Provide services and tools needed to enable ready use of NASA's Earth science data in the next-decadal models, research results, and decision support system benchmarking
- Improve support for end users

The EEE charter established two teams to accomplish these goals: a Study Team and a Technical Team. The Study Team received direction to provide recommendations consistent with the goal and objectives stated above and was charged with looking at the existing EOSDIS to determine the strategic evolution of its functions and elements in the broader context of the processes, goals, and objectives of the NASA Earth science strategy and plans for the next decade's data systems and architectures. The Study Team consisted of nationally recognized technical experts in Earth system science, applications and information technology. The Technical Team, led

by the ESDIS Project Manager, was made up of representatives from the ESDIS Project, DAACs and SIPSs, and selected consultants invited to provide independent perspectives on aspects of data system development from their experiences. The Study Team, along with the Technical Team, prepared a vision for the Evolution of EOSDIS Elements, (EEE Study Team 2005) projecting the system capabilities to the year 2015. The vision emphasized the need to ensure safe stewardship of the Earth science data while maintaining technological currency to further enable scientific research based on EOSDIS data holdings. This vision provided the guiding principles under which the Technical Team conducted its analytical work. The goals expressed in the vision and the tenets derived from them by the Technical Team are shown in Table 7.5.1. These goals and tenets were used in tracking the progress of evolution towards the vision.

The Technical Team performed a detailed analysis of the EOSDIS components and elements and developed an approach and implementation plan that would begin to fulfill the objectives set forth in the vision. The Technical Team sought inputs from the operators of each of the current system elements and encouraged their ideas and concepts for improving EOSDIS consistent with the vision. Selected consultants were invited to provide independent perspectives on aspects of data system development from their experiences.

The Technical Team analyzed the suggestions for: adherence to the vision and Study Team guidance; the investment costs, sustaining costs, and lifecycle costs; identification of the potential risks; implementation feasibility; timeframes and phasing opportunities; and for the affect on the user community. With this analysis, the element inputs were structured into the following set of alternative approaches:

- DAAC-focused – all DAACs develop their own archive management systems to reduce dependence on the core systems,
- SIPSs-focused – the SIPSs take on the archive, distribution and customer interface responsibilities in place of the DAACs, and
- Core System-focused – implement a re-architected ECS at all four DAAC sites where it was deployed at that time.

From these three alternatives a hybrid approach was defined, selecting the best aspects of each alternative that could be feasibly developed in concert. This fourth alternative, the Hybrid Approach, was advanced as the "best value" for cost containment, risk management and fulfillment of vision goals. NASA Headquarters approved this Hybrid Approach and directed the Technical Team to plan for its implementation.

Table 7.5.1. EOSDIS Evolution Vision – Tenets and Goals

Vision Tenet	Vision 2015 Goals
Archive Management	• NASA will ensure safe stewardship of the data through its life-time. • The EOS archive holdings are regularly peer reviewed for scien-tific merit.
EOS Data Interoperability	• Multiple data and metadata streams can be seamlessly combined. • Research and value added communities use EOS data interoper-ably with other relevant data and systems. • Processing and data are mobile.
Future Data Access and Processing	• Data access latency is no longer an impediment. • Physical location of data storage is irrelevant. • Finding data is based on common search engines. • Services invoked by machine-machine interfaces. • Custom processing provides only the data needed, the way needed. • Open interfaces and best practice standard protocols universally employed.
Data Pedigree	• Mechanisms to collect and preserve the pedigree of derived data products are readily available.
Cost Control	• Data systems evolve into components that allow a fine-grained control over cost drivers.
User Community Support	• Expert knowledge is readily accessible to enable researchers to understand and use the data. • Community feedback directly to those responsible for a given system element.
IT Currency	• Access to all EOS data through services at least as rich as any contemporary science information system.

7.6 Implementation of the Evolution Plan

The Hybrid Approach for implementation involved activities in five major areas of EOSDIS. These activities were carried out in parallel during the years 2006-2008. The first activity was re-architecting of ECS, consisting of simplifying the software and hardware architectures to reduce mainte-nance costs while improving service. This simplified ECS was deployed at three of the four data centers (Langley ASDC, LP DAAC and NSIDC). The second was the addition of the archiving and distribution functions for MODIS Level 1 and atmospheric data products to the MODIS Adaptive Processing System (MODAPS) SIPS, using on-line disks for the archive and reducing the size of the archive by processing Level 1 products on demand when deemed advantageous. The third was the deployment of the

Simple Scalable Script-based Science processor Archive (S4PA) system to replace the current ECS system at the GES DISC. Along with this development, all the data at the GES DISC were made accessible on line. The fourth was the development of the Archive Next Generation (ANGe) at the ASDC at LaRC. This replaced the Langley TRMM Information System (LaTIS) which was used for processing, archiving and distributing CERES data from TRMM, Terra and Aqua missions. Also, the processing system for the Terra Mission Multi-Angle Imaging Spectrometer (MISR) instrument was migrated to a Linux cluster. The fifth activity was the completion and deployment of the EOS Clearing House (ECHO) as a robust operational middleware system.

Each of these activities will be discussed briefly in the following subsections.

7.6.1 ECS Re-architecting

In 2005, the EOSDIS Core System (ECS) was deployed at four EOSDIS Data Centers that perform ingest, processing, archive and distribution of EOS data. The system architecture consisted of two loosely coupled systems: the original Science Data Processing System (SDPS) which provided ingest, archive, processing and distribution functions utilizing a large scale, multi-petabyte tape archive, and the newer Data Pool which provided data access and distribution via a large, shared disk store. The design of the original SDPS (circa 1995) was oriented towards the complexities of managing a tape archive; limited computing resources (CPU, memory, small direct-attached caches); and a sophisticated, type-extensible data model. The system was complex and large (over 1 million lines of custom C++ code plus scripts and database-stored procedures). The SDPS hardware architecture was based on enterprise-class SGI and Sun UNIX servers with direct attached storage and host-centric file systems. This hardware suite became expensive to maintain, and required custom code to be supported on two different operating systems (IRIX and Solaris). The Data Pool design leveraged new technology and experience with actual EOSDIS operations, utilizing a simplified data model, Order Management services based on the data being available online, and a hardware architecture built around a Storage Area Network (SAN).

The Evolution approach for ECS has many new features affecting the software, hardware and maintenance processes.

Data ingest and distribution functions were re-implemented as Data Pool services. This enabled retirement of the legacy Storage Management and Data Distribution subsystems. User search and order functions were

allocated to the evolving ECHO infrastructure, enabling retirement of SDPS user support tools and gateways. The complex Science Data Server database and custom software were re-placed with a greatly streamlined Archive Inventory Management database. All metadata are now stored in XML. The result of these changes was a significant reduction in the custom code to be maintained, from 1.2 million source lines of code (SLOC) to approximately 400K SLOC.

At each data center, all custom code applications and commercial-off-the-shelf software are now running on new hardware as a single blade cluster. All storage is now provided by the SAN, eliminating most of the network data transfers between hardware platforms. The Data Pool SAN capacity was increased to accommodate ingest buffers, and increased storage and distribution capacity. Databases were consolidated onto a single Linux-based database server at each data center.

The re-architected ECS introduced many changes to its suite of commercial-off-the-shelf software. The Hierarchical Storage Manager product was replaced with a newer product that is supported on Linux, runs on commodity hardware platforms, and reduces the archive administration workload on operators. This also enabled a significant reduction in licensing costs, and set the stage for the eventual migration to a totally on-line (disk-based) archive. Operating system maintenance has been simplified by reducing to one (i.e., Linux).

Experience in operating the ECS for 10 years identified many opportunities for improving operations efficiency through automation of operator-intensive tasks. This included automation of recovery from failure conditions; improved operator interfaces to simplify operations; and automation of human-intensive data management tasks.

The above evolution was implemented in a phased approach to minimize impact to operations. The new hardware architecture was deployed to the data centers while operations continued on the legacy configuration. All custom code was ported to run on the Linux operating system. The initial software release (which transitioned data ingest and distribution functions to Data Pool services) was implemented so that it could run on both the new Linux-based, commodity hardware architecture, and the legacy hardware. After the new software architecture was stable on the new hardware architecture, the legacy hardware was removed. The development and test facilities were transitioned first (in 2006) and the EOSDIS Data Centers migrated to the blade/SAN architecture in 2007. The second major software release, which enabled retirement of the legacy Science Data Server and transitioned all user search and order functionality to ECHO, occurred in mid-2008.

The goal of reducing recurring costs was achieved. With a reduction of 33% of the annual ECS operations and system maintenance costs, the investment in new hardware and software was repaid in the first year of the Evolution activity.

7.6.2 LAADS/MODAPS

The MODIS Adaptive Processing System produces the base level, land and atmosphere data products from the MODIS science instrument data. Before the Evolution activity MODAPS served as an EOSDIS SIPS (See Sect. 7.2.3) and the GES DISC processed Level 1 MODIS products, and archived and distributed the Level 1 and atmospheric products. Land products were archived and distributed by LP DAAC and the Snow and Ice products by NSIDC. EOSDIS Evolution resulted in migration of the archiving and distribution functions for MODIS Level 1 and atmospheric data from the GES DISC to MODAPS. By incorporating the MODIS data archiving and distribution functions into MODAPS the EOSDIS gained much efficiency. The part of MODAPS that performs these functions is referred to as the Level 1 and Atmospheric data Archiving and Distribution System (LAADS). All the other functions for MODIS data have remained at the same EOSDIS elements as indicated above.

The archive at MODAPS was planned as a fully disk-based archive. The Level 1 products that contributed a high production volume (54% of daily production) would be held on line for a short period for users to download and be produced on demand after that initial period. Benefits of this transition included:

- Reduction in archive growth through on-demand processing
- Faster access to products, reduced reprocessing time from all on-line storage
- Reduced costs due to use of commodity disks and simplification of operations
- Closer involvement and control by the science community, greater responsiveness to scientific needs, products, tools, and processing

Details about the implementation of LAADS can be found in (Masuoka et al. 2007). A few key points are mentioned below.

The foundation for LAADS is the MODAPS processing software. This software is now augmented with features that support product search, ordering and distribution. The production systems send the generated data product files to LAADS using scripts for ingestion. As a part of the ingest process, the LAADS archive data base is updated. Users can search for

data based on spatial and temporal criteria as well as quality thresholds. Inexpensive high-capacity disk drives have enabled MODAPS to significantly increase the processing rates and to improve distribution of Level 0 data as well as higher level data products to the user community by storing all except Level 1 products on line in the disk archive. Large products, produced early in the MODIS processing chain, such as the Level 1A (unpacked instrument counts) and Level 1B (calibrated radiances) are produced on demand. Such products may be regarded as "virtual products".

MODAPS also offers options including reprojection, masking, subsetting and reformatting for transforming both archived and virtual products in forms better suited for an individual user's needs. When a user requests an on-demand or custom product, jobs are launched on a cluster of compute servers assigned to LAADS and the results placed in an on-line directory for retrieval by the requestor. The products generated on demand are held in the on-line archive for a period of a few days to enable the requesting user (and any other interested user) to download the files. Also, with even further reductions in on-line storage costs, it is becoming feasible to hold larger percentages of the Level 1 products on line thus reducing the need for on-demand production.

The products in the on-line archive can be obtained by users through a variety of means. Interactive search and order can be performed through the LAADS web server. A file transport protocol (FTP) server is available for scriptable access to the entire archive. Some servers support machine-to-machine access protocols. Also, users can perform cross-instrument and cross-data center searches and order products through ECHO and the Warehouse Inventory Search Tool (WIST). Of these access methods, the most popular means of obtaining data products is via FTP.

The MODAPS evolution and the implementation and transition to operations of LAADS from GES DISC occurred as a gradual process during February 2006 and January 2007 to ensure no disruption to the users. Before the transition, the data were archived in robotic tape silos. The average monthly distribution during the 6 months prior to transition was 3 million files and 30 terabytes per month. Soon after the transition, with all data on line, these numbers increased to 7 million files and 48 terabytes, respectively, per month. In late 2008, these numbers were at over 150 million files and 80 terabytes per month.

7.6.3 GES DISC

From its inception as one of the original EOSDIS DAACs the Goddard Earth Sciences (GES) Data Information Services Center (DISC), devel-

oped and managed two data management systems concurrently. The first was the Version 0 system for TRMM and pre-EOS heritage data. The second system was the EOSDIS Core System used for NASA's Terra, Aqua, and Aura missions. The evolution plan at the GES DISC was to reduce operations to a single data management system.

The GES DISC evolution was based on an in-house developed software system and adoption of on-line storage using commodity based hardware. The data center staff initially developed the Simple, Scalable, Script-based Science processor Archive (S4PA) system to replace the V0 system. The S4PA is a simplified data archive architecture where data resides on commodity disks. Its modular design permits its reuse with replacement of application-specific components. The transitions from ECS to S4PA were handled incrementally to ensure no impact to ongoing operations. Given the reduction in volumes due to transfer of MODIS data archiving to LAADS/MODAPS as indicated above, the ECS robotic silos were phased out. This represents a cost savings to the GES DISC, and to the EOSDIS Core System as well.

The benefits of these changes included:

- Reduced operations costs due to consolidation of multiple systems into one software system
- Increased data center automation due to a single management system with simpler operational scenarios
- Reduced sustaining engineering costs due to use of simpler, scalable software and reduction in dependency on high maintenance commercial-off-the-shelf products
- Improved data access due to increased on-line storage and commodity disks/platforms
- Risk mitigation for the LAADS/MODAPS transition effort

The size of the GES DISC archive was significantly reduced with the transition of MODIS data management responsibility to MODAPS. The remaining science data at the GES DISC was archived on line, eliminating the need for tape silo storage. By the end of 2007 the GES DISC completed migration of data to S4PA.

With the data stored on line, users have greater flexibility for access to data. Users can navigate to the data of interest through the hierarchical structure of S4PA or write scripts to acquire bulk data. GES DISC also offers services such as OPeNDAP (Cornillon et al. 2003), OGC Web Map Service and Web Coverage Service, and on-line analysis capabilities using Giovanni, which is a Web-based application developed by the GES DISC that provides capabilities to visualize, analyze, and access Earth science

remote sensing data without having to download the data (Acker and Leptoukh 2007). Users can search for data by navigating through the hierarchical structure as indicated above, a web based hierarchical navigation tool or a free-text (Google-like) tool called Mirador. The GES DISC also supports cross-data center searches through the WIST client using the ECHO middleware. Since the data are archived on disks, GES DISC can tailor its services to particular missions or measurements and provide discipline specific services. More details about the evolution of GES DISC can be found in a paper (Kempler et al. 2009).

7.6.4 Langley ASDC

The Langley Atmospheric Science Data Center, an original EOSDIS DAAC, employed two data management systems over time to meet its needs for processing, archiving and distribution functions. The first, called the Langley Tropical Rainfall Measuring Mission (TRMM) Information System (LaTIS), was used for processing, archiving and distributing the Clouds and the Earth's Radiant Energy System (CERES) instrument data products and to archive and distribute all the pre-EOS data products held at this data center. The second system, the ECS, was used for processing, archiving and distributing the data products from the Multi-angle Imaging Spectro-Radiometer (MISR) instrument on the EOS Terra spacecraft, and for archiving and distributing data products from several other EOS instruments.

The ASDC's evolution strategy consisted of replacing LaTIS with a modern, advanced, scalable system developed in-house called Archive Next Generation (ANGe) (Ferebee et al. 2007). ANGe is designed to increase automation and reduce manual operations in the archiving and distribution functions, and be expandable for additional science data. ASDC also upgraded their ECS implementation with the re-architected version of ECS (described in Sect. 7.6.1 above). The benefits of these changes included:

- Reduction in sustaining engineering costs due to reduction in dependency on high maintenance commercial-off-the-shelf products
- Increased system automation
- Improved data access due to planned use of increased on-line storage and commodity disks/platforms

ANGe began to successfully archive and distribute CERES and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) datasets in 2008. In addition to ANGe development at the ASDC, the

software for processing MISR data has been migrated from the SGI systems to Linux clusters to increase efficiency and scalability and to reduce maintenance costs.

7.6.5 ECHO

The development of the EOS Clearing House (ECHO) was underway before the Evolution activity began, but it incorporated enhancements to meet Evolution objectives. ECHO is implemented as a series of releases adding or enhancing capabilities. The extension of ECHO capabilities is important for the EOSDIS Evolution to meet the objectives of data interoperability, data access, and preserving the pedigree of derived data products. Since the EOSDIS Evolution activity began, ECHO has added capabilities including Collection and Granule Browse data insert, update and delete; Enhancements to Access Control Lists (to support multiple collections per provider); Mechanism for Clients to perform Spatial Query based on Latitude/Longitude; Line Item Order Status; Reorganization of Web Services Applications Program Interface (API) to improve usability; Framework for error handling; and changes to improve maintainability and performance. Additional functionality under development includes support for asynchronous queries by ECHO Clients; support for Product Specific Attributes; and support for 2-dimensional coordinate-based search (e.g., path/row).

Future versions will include a new Web Service Order interface; new capabilities in the areas of metrics reporting, event notification, and data partner data reconciliation; improved performance from metadata transmission to ECHO ingest; and improvements to better ensure data integrity.

7.7 Progress towards Vision 2015

The Evolution planning teams characterized the Vision for 2015 in seven tenets, each representing a set of objectives guiding evolution success. These tenet goals, presented in Table 7.5.1, provide a mechanism for gauging the results to date. After nearly three years, the implementation activity discussed in Sect. 7.6 shows significant progress in meeting the goals of each vision tenet.

The evolution of EOSDIS resulted in progress towards the Vision for 2015 by implementing changes that maximize science value and achieve cost savings. At this stage in its evolution, EOSDIS makes data access easier and data products more quickly available to the science community by

increasing the amount of data available on line. EOSDIS data has also become more closely integrated with the science community, especially with MODIS data for the atmospheres community. Substantial cost savings have been achieved by replacing operations with automation, seeking less costly sustaining engineering approaches, and taking advantage of current information technology advances in hardware and automation.

Progress towards each Vision tenet is discussed below.

Archive Management has been strengthened through upgrades to the hardware and software at each site. More data is available on line. The LAADS/MODAPS data center and the GES DISC evolved away from tape archive based systems and the re-architected ECS at other data centers increased the use of data pools. EOSDIS data centers have tailored their processing and archiving software and systems to be more efficient. The data centers now have the ability to review the archive collection. This ability to better manage the archive supports the goal of long-term data stewardship.

EOS Data Interoperability is enhanced by making more data available on line, thus decreasing the access time to the science data and products.

The ease of access to EOSDIS data is evidenced through a dramatic increase in data distribution to end users. EOSDIS has experienced increases in product distribution of approximately 50% in both FY2007 and FY2008. In FY2006 the number of products distributed increased by less than 20% over the previous year. While some of this increase may be the result of a general world-wide growing interest in Earth science-related problems (e.g., climate change), the magnitude of the increase suggests that data are becoming easier to access.

Other aspects of interoperability are needed to achieve the 2015 Vision. EOSDIS can now focus on defining ways for combining multiple data and metadata streams seamlessly, and can address data interoperability with other relevant data systems. While on-line availability facilitates making processing and data mobile, it takes more effort and coordination with the science community to achieve this fully.

Future Data Access and Processing objectives are being met by archiving data on line and processing on demand, which supports provision of services that customize data access in the amounts and the form needed by science users. Because of the vast increase in processor speed EOSDIS is now able to process on demand. The ability to process on demand also reduces the size of the archive to be maintained. For example, EOSDIS ingests a large amount of data from the MODIS instrument alone, which progresses through multiple levels of processing to become useful products. At one time storing the initial processed (Level 1) data required a large archive for data that are not highly sought after nor uniformly re-

quested from across the entire global coverage area. By not archiving these intermediate data products, but ensuring availability by reprocessing lower level products as needed, EOSDIS saves storage space at the modest cost of some reprocessing. Also, the design of the evolved EOSDIS permits more agile decisions on processing versus storage based on changes in hardware technology and the resulting reductions in cost.

The ECHO middleware provides a robust and common means to access EOS data. Beginning in June 2008 EOSDIS Data Centers began transitioning from the EOSDIS legacy user interface system (EDG) to the EOSDIS Warehouse Inventory Search Tool (WIST) system. From a general user perspective, the access to data depends little on where it is physically located, or even the means to prepare it for delivery, as long as the data are made available in a reasonable amount of time.

The ability to track the <u>Data Pedigree</u> improves with the focus on metadata and the success of the evolving EOSDIS Core data model. More attention is needed for preserving and ensuring access to the various versions of the software used to generate the data products.

<u>Cost Control</u> was improved with a focus on identifying and evolving the components that were cost drivers. All data center sites began a process to transition from expensive workstations to commodity hardware and the replacement of expensive commercial-off-the-shelf tools with less expensive tools while retaining essential functionality. This included new maintenance strategies (e.g., purchase of less costly computing platforms with a more frequent refresh cycle rather than pay high maintenance costs); and increased automation leading to operational cost savings. Targeting specific data centers for initial improvements produced earlier and larger cost savings. The other data centers followed with self-directed upgrades to equipment replacement, software upgrades, and archive holdings' cleanup in parallel with the formal evolution process.

<u>User Community Support</u> improved by moving control of the data and supporting services closer to the users and science teams. A specific example from the EOSDIS Evolution is the support to the atmospheres community by combining the MODIS archive and distribution with the data processing function. This closer tie between the user community and data providers enables EOSDIS to be responsive to science requirements, and influences the product definition, tool development, and processing needs to the benefit of the users. The EOSDIS Data Centers have increased the number of on-line tools and services such as visualization and subsetting. All user communities, including the general public, are served by improved interfaces, the upgraded catalog and inventory tools, and the easier access to data on line.

Information Technology (IT) Currency is being realized in the upgrades and simplifications provided by the re-architected ECS and the data center upgrades, improving the flexibility to meet expectations of a more sophisticated user community. The entire EOSDIS Evolution of Elements activity is an example of the NASA commitment for continuous technology assessment and infusion.

7.8 Summary

In this chapter, we have provided a discussion of the evolution of EOSDIS. As NASA's major data system capability for managing Earth science data, EOSDIS has been evolving since its conception in the early 1990's. Many changes have occurred along the way. Starting with a centralized design involving two data centers, it was changed to have a more geographically distributed set of eight Distributed Active Archive Centers (DAACs), where each was focused on a specific set of Earth science disciplines. The design of the system where all the EOS standard products were to be generated at the data centers using the EOSDIS Core System and instrument team provided software was replaced by an implementation using Science Investigator-led Processing Systems (SIPSs) to generate the data products. Currently there are twelve EOSDIS Data Centers and fourteen SIPSs in EOSDIS. The standards, discussed in Sect. 7.4 of this chapter, have played an important role in the successful operation of these elements as well as in interactions with the user community.

The latest focused effort for evolution of EOSDIS started as a formal study initiated by NASA in 2004 with the goals of increasing efficiency, operability, data usability, services and tools availability, and improved support for end users. The Study Team and the Technical Team working on this evolution study arrived at a Vision for 2015, with the Technical Team following up with an initial implementation plan. This implementation plan was established by late 2005, and the implementation was carried out during 2006 through 2008. The major elements involved in this implementation were: the EOSDIS Core System, MODAPS, the GES DISC, the Langley ASDC, and ECHO. The simplified, re-architected ECS is now operating at three EOSDIS Data Centers – Langley ASDC, LP DAAC and NSIDC. As of this writing, most of the implementation has been completed, and the systems are in operation. Significant progress was made towards the Vision for 2015 goals.

Despite all of this progress, much work remains to ensure that EOSDIS remains a vibrant tool to serve the user community as technology changes

over the next several years. The rapid changes in information technology will provide both challenges and opportunities. The challenges will be due to increased expectations on the part of users concerning innovative uses of data and information to derive knowledge. The technologies will provide opportunities for EOSDIS and its elements to improve their capabilities to serve the community in innovative ways. A strategy for active infusion of technology is essential for the continued success of EOSDIS.

Acknowledgments

References

Acker JG, Leptoukh G (2007) Online analysis enhances use of NASA earth science data. Eos, Trans., Amer. Geophysical Union, 88 (2), 14, 17

Cleave ML (2004) Evolution of EOSDIS Elements Study Charter Amended. NASA http://eosdis-evolution.gsfc.nasa.gov/

Cornillon P, Gallagher J, Sgouros T (2003) OPeNDAP: Accessing data in a distributed, heterogeneous environment, Data Science Journal [Online] 2 (0), pp.164-174. http://www.jstage.jst.go.jp/article/dsj/2/0/2_164/_article

ECHO (2008) NASA EOS Clearing House. Online: http://www.echo.eos.nasa.gov/

EEE Study Team (2005) Evolution of EOSDIS Elements, Study Team Briefing to NASA. Online: http://eosdis-evolution.gsfc.nasa.gov/

EOS Project Science Office NASA GSFC (1990) EOS Reference Handbook

ESIPFED (2006) Federation of Earth Science Information Partners. Online: http://www.esipfed.org/

Feldman G (2007) Ocean Color Web. Online: http://oceancolor.gsfc.nasa.gov/, NASA (dynamically updated web page)

Ferebee MT, Cordner DE, Ritchey NA, Hunt LA, Piatko P, Haberer SJ, Wang FY (2007) Finding and accessing data at the NASA atmospheric science data center. IGARSS 2007, Barcelona

FGDC (2002) Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata. FGDC Document Number FGDC-STD-012-2002

HDF Group (2008) HDF-EOS Tools and Information Center. Online: http://hdfeos.net/index.php

Kempler SJ, Lynnes C, Vollmer B, Alcott G, Berrick S (2009) Evolution of Information Management at the GSFC Earth Sciences (GES) Data and Information Services Center (DISC): 2006 – 2007. IEEE TGARS (in press)

Maiden ME, NewDISS Team (2000) NewDISS: A 6-to 10-year Approach to Data Systems and Services for NASA's Earth Science Enterprise, Draft, Version 1.0

Masuoka E, Wolfe R, Sinno S, Ye G, Teague M (2007) A disk-based system for producing and distributing science products from MODIS. IGARSS 2007, Boston, MA

NASA (2006) NASA Strategic Plan. NASA Headquarters, Washington, DC 20546, NP-2006-02-423-HQ

NASA, CSU (2007) CloudSat Data Processing Center. Online: http://www.cloudsat.cira.colostate.edu/

NASA GSFC, Science Advisory Panel for EOS Data and Information (1990) Panel Comments on EOSDIS (Phase B) Final Design Review, February 12-16, 1990

NASA GSFC (1989) UARS Ground Data Processing Description Document

National Research Council (1995) A Review of the U.S. Global Change Research Program and NASA's Mission to Planet Earth/Earth Observing System. National Academy Press, Washington, D.C.

Ramapriyan HK (2003) NASA's Earth Science Data Systems – Past, Present and Future. IGARSS 2003, Toulouse, France

Ramapriyan HK, McConaughy GR (1991) Version 0 EOSDIS - An Overview. Technical Papers, ACSM-ASPRS Annual Convention, Vol. 3, pp 352-362

Stocker EF (2003) A precipitation processing system for the Global Precipitation Measurement mission. Proc. Int. Geoscience and Remote Sensing Symposium. (IGARSS 2003), Toulouse, France, IEEE, 1704–1706

Acronyms

ACCESS	Advancing Collaborative Connections for Earth System Science
ACRIM	Active Cavity Radiometer Irradiance Monitor
ACRIMSAT	Active Cavity Radiometer Irradiance Monitor Satellite
AIRS	Atmospheric Infrared Sounder
AMSR-E	Advanced Microwave Scanning Radiometer - EOS
AMSU	Advanced Microwave Sounding Unit
ANGe	Archive Next Generation
API	Application Programming Interface
ASDC	Atmospheric Sciences Data Center
ASF	Alaska Satellite Facility
ASTER	Advanced Spaceborne Thermal Emission and Reflection
CALIOP	Cloud-Aerosl Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CDDIS	Crustal Dynamics Data and Information System
CERES	Clouds and the Earth's Radiant Energy System
CPR	Cloud Profiling Radar
DAAC	Distributed Active Archive Center
DB	Direct Broadcast
DIF	Directory Interchange Format
DISC	Data and Information Services Center
DORIS	Doppler Orbitography and Radiopositioning Integrated By Satellite
ECHO	EOS ClearingHOuse
ECS	EOSDIS Core System
EDC	EROS Data Center
EDG	EOS Data Gateway
EDOS	EOS Data and Operations System
EEE	Evolution of EOSDIS Elements
EOC	EOS Operations Center
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
EPGS	EOS Polar Ground Stations
EROS	Earth Resources Observation Systems
ESDIS	Earth Science Data and Information System
ESDSWG	Earth Science Data System Working Groups
ESE	Earth Science Enterprise
ESIPs	Earth Science Information Partners
ESSP	Earth System Science Pathfinder
FGDC	Federal Geographic Data Committee
FOS	Flight Operations Segment
FTP	File Transfer Protocol
GCMD	Global Change Master Directory

GES	GSFC Earth Sciences
GHRC	Global Hydrology Resource Center
GLAS	Geoscience Laser Altimeter System
GPM	Global Precipitation Mission
GSFC	Goddard Space Flight Center
HDF	Hierarchical Data Format
HIRDLS	High-Resolution Dynamics Limb Sounder
HSB	Humidity Sounder for Brazil
ICESat	Ice, Cloud and Land Elevation Satellite
IMS	Information Management System
IT	Information Technology
JMR	Jason Microwave Imager
JPL	Jet Propulsion Laboratory
LAADS	Level 1 and Atmospheric data Archiving and Distribution System
LaTIS	Langley TRMM Information System
LIS	Lightning Imaging Sensor
LP DAAC	Land Processes DAAC
MEaSUREs	Making Earth Science Data Records for Use in Research Environments
MISR	Multi-angle Imaging Spectrometer
MLS	Microwave Limb Sounder
MODAPS	MODIS Adaptive Processing System
MODIS	Moderate-Resolution Imaging Spectroradiometer
MOPITT	Measurements of Pollution in the Troposphere
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCSA	National Center for Supercomputing Applications
NewDISS	new Data and Information Systems and Services
NISN	NASA Integrated Services Network
NOAA	National Oceanic and Atmospheric Administration
NRC	National Research Council
NSIDC	National Snow and Ice Data Center
OBPG	Ocean Biology Processing Group
OGC	Open GIS Consortium
OMI	Ozone Monitoring Instrument
ORNL	Oak Ridge National Laboratory
PB	Peta Byte
PI/TL	Principal Investigator/Team Leader
PO.DAAC	Physical Oceanography DAAC
QuickScat	Quick Scatterometer
REASoN	Research, Education and Applications Solutions Network
S4PA	Simple Scalable Script-Based Science Processor Archive
SAGE	Stratospheric Aerosol and Gas Experiment
SAN	Storage Area Network
SAR	Synthetic Aperture Radar
SDPS	Science Data Processing Segment

SeaWinds	Seawinds Scatterometer (For Flight On ADEOS II)
SEDAC	Socio-economic Data Applications Center
SEEDS	Strategic Evolution of Earth Science Enterprise (ESE) Data Systems
SIM	Spectral Irradiance Monitor
SIPSs	Science Investigator-led Processing Systems
SLOC	Source Lines of Code
SOLSTICE	Solar Stellar Irradiance Comparison Experiment
SORCE	Solar Radiation and Climate Experiment
TB	Terabyte
TDRS	Tracking and Data Relay Satellite
TES	Tropospheric Emission Spectrometer
THG	the HDF Group
TIM	Total Irradiance Monitor
TRMM	Tropical Rainfall Measuring Mission
TSDIS	Tropical Rainfall Mapping Mission Science Data and Information System
UARS	Upper Atmosphere Research Satellite
U.S.	United States
USGS	U.S. Geological Survey
UWG	Users Working Group
V0	Version 0
WCS	Web Coverage Service
WIST	Warehouse Inventory Search Tool
WMS	Web Mapping Service
WSC	White Sands Complex
WWW	World Wide Web
XPS	XUV Photometer System