A NOAA/NASA Pilot Project for the Preservation of MODIS Data from the Earth Observing System (EOS)

Robert Rank[1], Kenneth R. McDonald[2]

[1] ***NOAA/NESDIS***
5627 Allentown Road #100, Suitland MD 20746 USA
(Robert.Rank@noaa.gov)

[2] ***NASA/Goddard Space Flight Center***
Mail Code 423, Greenbelt MD 20771 USA
(ken.mcdonald@nasa.gov)

Agency Responsibilities

NASA's Earth Observing System (EOS) program was initiated in the 1980's and formally established in 1991 as a U.S. Presidential initiative to be NASA's contribution to the U.S. Global Change Research Program. Its purpose was to enable the study of the Earth as a system and it consisted of three major elements. The first was a constellation of satellites that would acquire long-term records of remotely sensed environmental variables on a global scale. The second was a comprehensive science research program that would produce and validate the satellite data products and use them in studies of Earth system science. The final component was the EOS Data and Information System (EOSDIS) that would provide the data and information systems and services to the EOS program during the active life of the individual EOS missions. By interagency understandings and agreements, the long term preservation and stewardship of the EOS data are the responsibility of National Oceanographic and Atmospheric Administration (NOAA) and the U.S. Geological Survey (USGS) for atmospheric and oceanographic data products and for land processes data, respectively.

EOSDIS has been under development since the early 90's. It is a distributed system that is deployed and operated at eight Distributed Active Archive Centers (DAACs) across the U.S. It supports the full range of data and information systems and services for EOS and other NASA Earth science programs including data production, active archive, search and order, data distribution and direct data access. The data product generation occurs both at the DAAC facilities and at remote facilities developed and managed by the EOS scientists. EOSDIS also includes a distributed framework that supports the interconnection and interoperability of these components.

The National Environmental Satellite, Data, and Information Service (NESDIS), a line office within the National Oceanic and Atmospheric Administration (NOAA), is responsible for archiving and disseminating environmental data collected by a variety of ground-based and space-based observing systems. The Comprehensive Large Array-data Stewardship System (CLASS) is NOAA's planned mechanism for securely archiving large-volume data and data products, and for making this data available to researchers, commercial users, and the public. The volumes of data collected by future satellite-based instruments and observation systems (as well as the growing customer demand for this data) will overwhelm the existing archive and distribution systems without a dramatic expansion in capacity: CLASS is the vehicle for the necessary expansion.

There is reason to expect that the information technology advances we have seen in the last ten years will continue in the future. With these advances, NOAA has made significant progress in its ability to archive and provide access, and will continue to leverage these advancing technologies through effective stewardship of current resources. Management of the increased volume of data can be accomplished only through a rapid expansion in storage capacity, increased communications bandwidth, and automation of the means of data ingest, quality control, and access. The CLASS program will act as the connection in NOAA's effort to meet these challenges and pave the way to accommodate the additional massive data volumes expected over the next several years.

The ability to ensure on-going scientific stewardship for NOAA's environmental data and information will only be possible through extensive enhancement of NOAA's current data ingest, quality assurance, storage, retrieval, access, and migration capabilities. This goal will be met through the development and implementation of a standardized archive management system, which will be integrated with a robust, large-volume, rapid-access storage and retrieval system that is capable of

storing the incoming large array environmental data, in situ data, and operational products as well as receiving a user's on-line data request, automatically processing the request, and providing the requested data on the most appropriate media. This system will provide standardization in media, interfaces, formats, and processes for the very large datasets produced by satellites and radars. Additionally, the system will facilitate ongoing migration, preservation, and validation to new technology and media. This system is modular in design, built to integrate with automated real-time or near-real-time systems that deliver data. Transaction processing will be implemented to enable an essentially "hands-off" operation and, where appropriate, allow users to pay for data or services through credit card or automated billing.

Placing data on-line for access via the Internet is a high priority in accordance with the Federal Government's eGov initiative. Data storage and retrieval systems will continue to be upgraded to support effective and efficient access with special focus on Internet interfaces, emerging telephony technologies, and on-line data that support the objectives of the CLASS concept of operations and ensure that the Nation has access (including Section 508 compliance) to their data and information.

New satellite observation campaigns are being prepared for launch and operations. The volumes of data to be collected by these campaigns dwarf the data streams managed by existing archive and distribution systems within NESDIS. The size, number, and frequency of data sets to be stored and distributed will require significant expansion of capacity for moving, storing, processing, and distributing data. New and continuing remote-sensing campaigns include GOES, POES, DMSP, NEXRAD, NPP, NPOESS, and METOP; numerous in situ observation programs also contribute to the information processing challenge. CLASS has also been designated as NOAA's system that will support the long-term archive requirements for selected NASA EOS data products. The CLASS concept was developed as a framework to provide integrated data support while accomplishing the needed capacity expansion.

CLASS Overview

The Comprehensive Large Array-data Stewardship System (CLASS) is a National Environmental Satellite, Data, and Information Service (NESDIS) initiative to develop and implement a single Information Technology (IT) system for the archive, access, and distribution capabilities of certain National Oceanic and Atmospheric Administration (NOAA) and non-NOAA data. Over time, CLASS plans to support many disparate data types and data sets. CLASS will be hosted at multiple sites while providing a single interface to the user. This multiple-site capability is intended to improve system availability, scalability and enhance data integrity through replication at geographically disparate sites.

CLASS is built on a combination of existing information storage and access systems. The primary basis for large-volume data storage and management was the Satellite Active Archive (SAA), which provided a comprehensive archive, access and distribution capability for POES data, and some derived environmental data products. Covering a broader range of derived environmental data products at smaller volumes, the National Virtual Data System (NVDS) supports user access and ordering for data from many research projects and observation efforts. Discovery of information is supported by the NOAA Metadata Manager and Repository (NMMR) and web site, which provides pointers to specific data systems based on standard descriptors.

CLASS is a key project within NESDIS, drawing on support from several of the NOAA National Data Centers (NNDCs) and managed by a multi-center CLASS Project Management Team (CPMT) reporting to the NESDIS Office of Systems Development (OSD). The project has been underway since November 2001 and fielded its first operational release in July 2002 (Release 0). Architectural and engineering studies were performed to characterize the architecture of the Archive, Access and Distribution Segment of CLASS, and to select hardware and system software for the high-volume hierarchical data storage systems.

The CLASS Project Management Team (CPMT) must define the detailed policies for lifecycle management. The general organizational objectives are as follows:

- Development of software is geographically distributed, but controlled by a centralized Configuration Control Board.
- Installation of the system will be at two sites, leveraging existing facilities.
- Operations for archive and distribution of data sets are fully automatic and depend upon electronic transfer for source data.

- Operators are able to conduct supporting activities remotely, using a secure interface. Examples of supporting activities are customer help, reconfiguration of hardware for load balancing, and data migration.
- Operators coordinate with the housing facility, but system administration of hardware is the responsibility of the housing facility site management.
- Unless a disaster contingency plan is in progress, each housing facility must have one operational system

The policies cover the following four categories of capabilities and responsibilities:

1. Archiving - take in data, catalog, move to archive
2. Preservation - backup, duplication, and migration
3. Distribution - Internet-interface, user access, order fulfillment
4. Operation - people managing the system, and policies.

EOSDIS Operations

An early version was developed to support access to heritage Earth science data that were acquired prior to the EOS program and has been operational since 1994. These data collections were archived in existing systems at the EOSDIS DAACs and the Version 0 Information Management System (IMS) was developed to provide a common search and order interface to them. The EOSDIS Core System (ECS) was developed to support the full range of data system services for the EOS era of satellites. It provides a production, archive and distribution capability for standard science data products and is operational at four of the EOSDIS DAACs. Additional EOS data products are generated at Science Investigator-led Processing Systems (SIPS) and are transferred to the DAACs for archive and distribution services. The ECS has been operational since 1999.

Today EOSDIS is in its extended operations phase and is successfully supporting a massive collection of Earth science data and information. For the fiscal year ending in September 2004, the metrics of these operations include:

- EOSDIS operations at 8 DAACs and 13 SIPs;
- Archive and distribution support for 65 instruments on 28 missions and science data processing for 11 of those instruments;
- Daily archive growth of over 4 TB and total archive volume of approximately 4 PB;
- Over 2 million distinct users accessing the DAACs and over 200,000 distinct users obtaining data;
- Annual distribution of over 33 million data products to users at a rate of almost 2 TB per day.

NASA is required to maintain data system operations for its data collections for a period of four years beyond the life of each mission to support the reprocessing campaigns that are planned by each of the respective science teams. Terra, the first EOS mission, was recently extended to 2010 and the other EOS missions, Aqua and Aura, are still having a number of years remaining on the planned mission life. However, because of the magnitude of the NASA EOS data collection and to prepare to address the transfer of responsibility for the heritage data, NASA and NOAA have initiated a project to define and implement the interface to migrate EOS data to its long-term archive.

EOS Long Term Archive Pilot Project

A Pilot Project is being conducted with the National Aeronautics and Space Administration (NASA) to begin incorporating Earth Observing System (EOS) data into the NOAA/CLASS Archive Access and Distribution Segment (AADS) using Moderate Resolution Imaging Spectroradiometer (MODIS) Level 0 and 1B data from the Terra and Aqua missions. Subsequent, follow-on tasks to this NOAA and NASA EOS Long Term Archive (LTA) – MODIS Pilot Project will move this effort from a pilot study to an operational system. The pilot project covers developing a pilot process, including the requirements, design, and implementation, and responsibilities associated with preserving and servicing the data products as they migrate between NASA organizations and NOAA. This pilot project is an important step in archiving valuable NASA satellite information in CLASS.

The NOAA/NASA MODIS Pilot Project involves a number of different organizational elements within NASA and NOAA and each are represented on the project team. The NASA Earth Science Data and Information Project (ESDIS) manage the development and operation of EOSDIS and are coordinating the NASA effort on the pilot project. The MODIS data is archived at the DAAC located at the Goddard Space Flight Center and these DAAC personnel are responsible for implementing the NASA side of the interface. Members of the NASA team also include representatives of the MODIS Science Data Support Team (SDST), NASA HQ and GSFC's network infrastructure support staff. NOAA will manage its effort within the CLASS project. The CLASS Program Manager will determine the overall priority of the project within the available resources. The CLASS Program Manager has delegated the responsibility for the project management of this pilot study to the CLASS EOS Project Manager who reports status on a weekly basis to the CLASS Project Management Team (CPMT). The CLASS EOS Project Manager and the NASA ESDIS Project Manager will interact directly. Both teams have committed to meet the agreements and specifications that are documented in the Submission Agreements and Interface Control Documents. Individual team members share information as needed to develop the requirements, create the Submission Agreements and Interface Control Documents, and implement the pilot.

ISO/CCSDS Standards for OAIS

In developing CLASS, NOAA is following the guidelines of the International Organization for Standardization (ISO), which has requested that the Consultative Committee for Space Data Systems (CCSDS) Panel 2 coordinate the development of standards in support of the long-term preservation of digital information obtained from observations of the terrestrial and space environments. The initial effort has been the development of a Reference Model for an Open Archival Information System (OAIS), which establishes a common framework for functional and information modeling concepts applicable to any archive. It is specifically applicable to organizations that have a responsibility to provide long-term access to digital information. This Reference Model concentrates on developing a framework that will allow data centers to identify common approaches to packaging data that Earth science data producers can use to develop data collections that facilitate their ingest into the archives.

Data Providers produce products supported by validated science requirements and request that those products be archived and made available to a designated community by the archive. Submission agreements are negotiated for each product. Those submission agreements lead to allocated requirements for archive, access and distribution development. Hopefully, those requirements overlap, so that the development effort for each new product decreases with time. The Core CLASS System requirements should facilitate that evolution.

Mandatory Responsibilities:

1. Negotiate for and accept appropriate information from information Producers.
2. Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.
3. Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided.
4. Ensure that the information to be preserved is Independently Understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.
5. Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.
6. Make the preserved information available to the Designated Community.

Within the Submission Agreement, one or more Data Submission Sessions are specified. There may be significant time gaps between the Data Submission Sessions. A Data Submission Session will contain one or more Submission Information Packages (SIPs) and may be a delivered set of media or a single telecommunications session.

The Data Submission Session content is based on a data model negotiated between the Archive and the data provider in the Submission Agreement. This data model identifies the logical components of the SIP (e.g., the Content information, Preservation description information, Packaging information, and Descriptive information) that are to be provided and how (and whether) they are represented in each Data Submission Session. All data deliveries within a Submission Agreement are recognized as belonging to that Submission Agreement and will generally have a consistent data model, which is

specified in the Submission Agreement. A Submission Agreement also includes, or references, the procedures and protocols by which the Archive will either verify the arrival and completeness of a Data Submission Session with the Producer or question the Producer on the contents of the Data Submission Session.

The NOAA/NASA MODIS Pilot Project provides an excellent case study to exercise the concepts of the OAIS framework but with some additional complexity that is not directly addressed in the standard. In this case, the archives of the EOSDIS are the providers of the data corresponding to the OAIS Producer and CLASS is the OAIS archive. Both EOSDIS and CLASS have their respective Management components and their own requirements, policies, standards and conventions. In addition, as the data migrates from EOSDIS to CLASS, the Consumers of that data will of necessity transition as well. These complicating factors only elevate the need for the standard mechanisms and procedures that are outlined in the OAIS Reference Model. The Reference Model is still silent regarding some of the issues that we have raised about what data producers need to understand. Perhaps the most important of these is the issue of preserving the semantic (i.e., the scientific) content of Earth science data. While data formats are important, they are considerably more transient than the underlying sampling structure that is unique to scientific data. It would appear that the data centers have a responsibility for bringing these semi-conscious structures into clear visibility. The most practical approach may be to develop a community consensus on a (hopefully) small set of standards in each discipline that provides data. Over time, the visibility of the standards may evolve towards a community consensus across the whole realm of the Earth sciences.

Initial Contact between Data Producer and Archive

The OAIS Reference Model notes, "The first contact between the OAIS and the Producer is a request that the OAIS preserve the data products created by the Producer. The OAIS, the Producer or Management may initiate this contact. The Producer establishes a Submission Agreement with the OAIS, which identifies the SIPs to be submitted and may span any length of time for this submission. Some Submission Agreements will reflect a mandatory requirement to provide information to the OAIS, while others will reflect a voluntary offering of information. Even in the case where no formal Submission Agreement exists, such as a World Wide Web (WWW) site, a virtual Submission Agreement may exist specifying the file formats and the general subject matter the site will accept." The OAIS Reference Model (RM) does define several types of Information Package, one of which is the Submission Information Package, where the RM provides the following generic notion "The Submission Information Package (SIP) is that package that is sent to an OAIS by a Producer. Its form and detailed content are typically negotiated between the Producer and the OAIS."

Submission Agreements describe relationships between data providers and archives. They include information on:

- Human Contacts (technical, metadata)
- Designated Community (w/ representatives), search interfaces
- Data transfers / protocols, validation, errors and actions
- Data formats / standards
- Data quality information, lineage
- Parameters / ranges
- Performance

There is also a notion that it is generally wise to tell the Data Producer that the Archive is expecting to follow the OAIS Reference Model in its operations and that there is a formal negotiation of content, roles, and responsibilities that will follow the initial contact. To assist the Data Producer in preparing for the negotiation process, it is helpful to have a draft version of the Submission Agreement Negotiation Schedule that can be revised as the negotiation proceeds.
This initial contact should probably be kept simple. The steps involved in this use case are as follows:

1. Under normal circumstances, the Archive should disclose to the Data Producer the process of negotiating a Submission Agreement, including the Schedule for the negotiation.
2. During the discussion, the Archive Representative should keep minutes of the meeting that can be included in the Notes directory after that has been set up in the Archive's Submission Agreement History Store.
3. At the conclusion of this contact, the Archive should open a Submission Agreement History directory with subdirectories associated with each type of documentation. At this point, it is probably overkill to record any documents in the subdirectories, although blank or sample templates for documents might be copied over from a store.

Obtain Collection Description

From a simple perspective, the Producer Collection Description is intended to be a brief description, say one page, created by a Data Producer of the content and expected use of a collection to be accepted by the Archive. It is particularly helpful to obtain some preliminary information regarding the number and size of files, as well as the Data Producer's expectations regarding the Designated User Community.

The Collection Description is intended to be a brief description (one or two pages) of the Data Producer's view of what needs to be incorporated into the Archive. The actor with the most knowledge of this information is almost certainly the Data Producer Representative. We anticipate that he (or she) has a strongly developed view of what will be in the collection description.

To help the Data Producer Representative document this view, the Archive Representative can ask the following questions:

- What kinds of data will be stored in the Archive's collection?
- To what level are the data in this collection processed (Level-0, Level-1/1b, RDR, SDR, EDR, CDR, other)?
- Are data still being added to the collection?
- If yes, what is their source, and how are they transferred?
- What is the existing data volume and roughly what period of time does this cover?
- What is the rate of data being added to the collection?
- About how many files will be in each sub-collection?
- About how big are the files in each sub-collection?
- How many Submission Sessions are there likely to be?
- How long will it take to ingest the data in a Submission Session?
- What kinds of Quality Assurance work is expected on each file coming in?
- What kind of metadata and provenance information will be supplied with each file?
- What services does the Data Provider expect, need, or require for the collection? - Reformatting, manual metadata input, etc.
- What user groups are there likely to be for the data in the collection?
- How many users are there likely to be in each user group?
- Is there any experience we could write down to help quantify how many files each user group is likely to order?
- Are the users likely to need or want data on media?

After the session ends, the Archive should do the following:

- Record the minutes of the meeting in the notes section of the Submission Agreement History.
- Set up a deadline for receiving a text version of the answers to these questions - in so far as the Data Producer may be able to provide it.
- When the text arrives, ingest it into the Submission Agreement History under the Obtain Collection Description director

CLASS Architecture

The CLASS Architecture Report refers to agreements between CLASS and data providers as Interface Control Documents (ICDs). CLASS has proposed that these items be called CLASS Submission Agreements (CSAs) instead. The role of these agreements is discussed in several of the use case descriptions.

Scenario: Ingest data (standard workflow). The data supplier and CLASS have previously established an ICD for data transfer defining network channels, file and data formats, processing workflow, and acknowledgment protocol.
Scenario: Ingest independent metadata. The data supplier and CLASS have previously established an ICD for metadata transfer regarding observation data that is maintained outside of CLASS. The ICD defines the format and content of the metadata as well as data transfer channels, processing workflow, and acknowledgement protocol.
Scenario: Establishing an ICD for new data stream. The CPMT and Archive Requirements Working Group (ARWG) have approved addition of a new data stream. The details of information exchange must be defined and documented. The process results in an interface control document (ICD).

CLASS and the cognizant data manager negotiate the communication channels (e.g., ftp drop-box procedures and addresses), expected data volumes, error conditions and actions, ingest processing expectations, data formats, metadata types and structure, and any other specific requirements. The ICD negotiation process ensures that expectations and commitments have been clearly understood, and the ICD becomes part of the formal CLASS documentation. Once complete the ICD is placed under configuration control.

Metadata Requirements

The CLASS ADS is required to provide Federal Geographic Data Committee (FGDC) compliant metadata that describe CLASS holdings. The CCSDS recommendation points out that much of that information is "more easily available or only available at the time when the original information [science data] is produced" thus stressing the need for data producers to play an active role in creating and maintaining standards compliant metadata. Metadata are necessary for many purposes, such as calibration and data processing. In addition to technical characteristics, additional metadata would be needed to allow researchers to find needed data in response to ill-formed requirements and take only necessary subsets of data. This requires open standards that allow for a great degree of flexibility in data storage and retrieval. The experience of the NESDIS Data Centers is certainly consistent with this need. These data centers now find it difficult or impossible to create metadata for older data holdings that were not well documented when they were originally provided to the data center. In light of these observations, the CLASS submission agreements should clarify producer's metadata requirements. Whenever possible, those requirements should be addressed in standards compliant ways. For example, when contact information is provided, the FGDC contact template should be used as an indicator of the required information. When parameters in the data products are described, those descriptions should include the FGDC required information at a minimum. This approach will certainly lead to improvements in the quality of metadata available to CLASS users and in the scientific usability of the data that CLASS provides.

The FGDC Metadata Content Standard includes a mechanism for extending the standard to include specialized fields for specific datasets or applications. Producer metadata requirements that cannot be met using the standard fields should be addressed by developing extensions to the standard. A set of extensions has recently been created for Remote Sensing Data that will be applicable to many datasets archived by CLASS. These extensions need to be considered in the metadata discussions for any remote sensing data / products.

Climate observations that are well documented, and have good metadata about the systems and networks used to make them, become more valuable with time. The creation of climate-quality data records (CDRs) is a fundamental objective of the global observing system for climate. International standards and procedures for the storage and exchange of meta-data need to be developed and implemented for many climate observing system components, including those of the operational satellite community. It is essential that all such data be properly archived and managed with the full expectation that they will be reused many times over in the future, often as a part of reprocessing or reanalysis activities. Good stewardship of the data also requires that data be migrated to new media as technology changes, be accessible to users, and be made available with minimal incremental costs.

Finalized Submission Agreement

A Finalized Submission Agreement is a document that identifies that a Submission Agreement is in place and has been accepted by the Data Producer and the Archive. The Finalized Agreement references the most current versions of all the subsidiary components of the Submission Agreement, and identifies the conditions under which the agreement may be terminated and the process for disposing of the Data Producer's items stored in the Archive. The Agreement needs to provide at least one Authorized Data Producer Representative, his (or her) user ID, a trust authority that can verify ftp submissions, and one or more IP addresses that this Representative can use in submitting authorization information to the Archive. The Finalized Submission Agreement must include an agreed to allocation of Storage Capacity for the Data, Metadata, and Provenance History portions of the Archive Storage Partition that will be allocated to this Data Producer. The Finalized Submission Agreement must include a list of the Archive Services and their attributes. There must be a Finalized Submission Agreement before further activities relating to Archive Submission can commence.

Use of the OAIS Reference Model – Benefits and Challenges

The use of the OAIS Reference Model has served the NOAA/NASA MODIS Pilot Project in a number of ways. In what might first appear to be a simplistic example, the reference model has given the data provider and the archive a common

terminology with which to frame the discussion. It has also identified a common set of functions that are required to archive data and information and a set of processes to establish the specific requirements that are associated with those functions. Further, it identifies a set of documentation to capture and record those requirements and specifications. Although some of these functions and processes have had to be tailored and extended to meet the needs of the pilot project, the reference model has provided an excellent foundation on which to build.

One source of difficulty in the direct application of the OAIS Reference Model to the long-term archive of EOS data in CLASS is that in this case, the data provider is not the original data producer but rather is an existing, operational archive itself. The active archives of EOSDIS have been developed over the last decade and have established their own approaches, designs, processes and documentation to support the archive of data received from the source providers and to disseminate that data to users. This requires the additional step of mapping and adapting the conventions of EOSDIS to the OAIS model being used by CLASS. Additionally, EOSDIS was designed to support the active archive function that includes an ongoing interaction with the EOS science teams as they refine their data processing algorithms, reprocess the raw data and submit updated versions of the science products to the archive. During this phase in the data life cycle, the focus is on production of the best science products and not necessarily on their long-term preservation. Some of the steps that ensure the proper stewardship of the data, including the generation of documentation and restructuring data collections must be done after the fact.

Another challenge in the long-term archive of EOS data is the shear magnitude and complexity of the science products and the issues associated with the decisions on what to transfer and how to organize the information that accurately captures and conveys that complexity. In the MODIS pilot project, the decision has been made to initially transfer the lower level instrument data. However, to use that data to generate higher level science products requires orbit and attitude data for the Terra and Aqua satellites that carry the MODIS instrument. These ancillary data sets come from different sources and are in different formats for each of the satellites and are stored in EOSDIS in both raw and processed forms. The ancillary data must be linked with instrument data and the processing software used to generate the higher-level science products. Additionally, that processing software, and all of the information on the computer environment in which it runs must also be stored in the archive. Finally, for true data preservation all of the expertise that is required to generate and interpret the data and associated products must also be captured and preserved.

Summary

The lifecycle of NASA's Earth science data has been the subject of several agency and interagency studies and reports and has implications and effects on agency charters, policies and budgets and on their data system's requirements, implementation plans and schedules. Ensuring that long-term satellite data be preserved to support global climate change studies and other research topics and applications presents some major challenges to NOAA and its partners. It is critical that NOAA and NASA work closely and constructively together to accomplish these important objectives. The NOAA/NASA MODIS Pilot Project is a focused effort to establish an interagency team to begin to address these challenges.

The OAIS Reference Model has been incorporated into NOAA's approach in the development of CLASS and is being employed in the pilot project as a forerunner to NOAA's assumption for the responsibility as the long-term archive for EOS data. The project team is using the OAIS mechanisms to establish and document the interagency transfer of data and is finding those concepts very useful in structuring the discussions and capturing a complete set of goals and requirements. The pilot project is the first opportunity that the team has had to exercise the OAIS mechanisms and is learning the details of the reference model as the effort proceeds. The team is also learning that even for the few MODIS data sets that are currently being addressed, the long-term preservation and stewardship of EOS data and all of its associated information is an extremely complex problem. The OAIS reference model is helping the pilot project and the experience and lessons learned by the project team should assist future users of the model.

References

[1] Consultative Committee on Space Data Systems (CCSDS), 2002: "Reference Model for an Open Archival Information System (OAIS)", CCSDS 650.0-B-1 Blue Book, January 2002.
[2] "A Procedure for Developing Engineering Requirements", Bruce R. Barkstrom, Atmospheric Sciences Data Center, NASA Langley Research Center, 05/05/04

[3] “Digital Archive Issues from the Perspective of an Earth Science Data Producer”, Bruce R. Barkstrom, Atmospheric Sciences Data Center, NASA Langley Research Center
[4] “Comprehensive Large Array-data Stewardship System (CLASS), Data Product Submission Agreements”, Ted Habermann, NOAA/ National Geophysical Data Center/Enterprise Data Systems Division, December 4, 2002
[5] “Comprehensive Large Array-data Stewardship System (CLASS), Guidelines for Submission of Data to the Archive as part of the Archive, Access and Distribution Segment”, Robert Rank, NOAA/NESDIS/OSD, December 9, 2003
[6] “1999 EOS Reference Handbook”, Michael D. King and Reynold Greenstone, editors, EOS Project Science Office, NASA Goddard Space Flight Center