Evolution of NASA’s Earth Observing System Data and Information System (EOSDIS)

Mary Ann Esfandiari
Deputy Associate Director for EOS Operations
NASA/Goddard Space Flight Center

May 23, 2006
Earth Observing System Data Management

• NASA’s Earth Observing System Data and Information System (EOSDIS) is a petabyte-scale archive of environmental data that supports global climate change research

• EOSDIS provides for
  – Data ingest
  – Data processing
  – Data distribution
  – Archive management

This MODIS image shows the wide sediment plume of the Yangtze River as it empties into the East China Sea.
Credit: Jacques Descloitres, MODIS Land Science Team
Image Date: 09-16-2000
EOSDIS Manages Data From 22 Instruments on 9 Spacecraft
Earth Observing System Data Management

• NASA’s Earth Observing System (EOS)
  – Mission is to collect Earth remote sensing data for a 15 year global change research program

• EOS Data and Information System (EOSDIS) Overview
  – Designed to receive, process, distribute and archive several terabytes of science data on a daily basis
  – Provides a distributed information framework (data centers, system elements, interfaces and data model) supporting EOS investigators and other users
  – Interoperates with data archives of other agencies and countries
  – Distributed Active Archive Centers (DAACs)
  – Science Investigator-led Processing Systems (SIPS)

• Web Site: http://eos.nasa.gov/eosdis
EOSDIS Data Centers
Nine Centers Geographically Distributed

- **ASF**
  - SAR Products
  - Sea Ice
  - Polar Processes

- **NSIDC**
  - Cryosphere
  - Polar Processes

- **LP DAAC-EDC**
  - Land Processes & Features

- **SEDAC**
  - Human Interactions in Global Change

- **GES DAAC-GSFC**
  - Upper Atmosphere
  - Atmospheric Dynamics
  - Ocean Color
  - Hydrology
  - Global Biosphere
  - Radiance Data

- **ASDC-LaRC**
  - Radiation Budget
  - Clouds
  - Aerosols
  - Tropospheric Chemistry

- **PODAAC-JPL**
  - Ocean Circulation
  - Air-Sea Interactions

- **GHRC**
  - Hydrologic Cycle
  - & Severe Weather

- **ORNL**
  - Biogeochemical Dynamics
  - EOS Land Validation
EOSDIS Context
EOSDIS Scale

• EOSDIS managing extraordinary rates and volumes of scientific data
  – Terra alone produces 194 GB of “raw” data/day; almost as much as the Hubble Space Telescope (HST) in one year
  – Adding 4 TB of data per day

• Current EOSDIS data archive volume is approximately 4.9 PB
  – Distributing 2.5 TB of data per day to users
  – Over 36 million products distributed to ~ 2 million distinct users per year
EOSDIS Evolution

- In early 2005, NASA embarked on an EOSDIS Evolution Study
- Address multi-faceted goals/issues:
  - Manage archive volume growth
  - Improve science need response and data access
  - Reduce recurring costs of operations and sustaining engineering
  - Update age of systems and components
  - Move towards more distributed environment
- A vision for the 2015 timeframe was developed to guide conduct of study
- EOSDIS Evolution “Step 1” Plan approved by NASA Headquarters in late 2005.
# EOSDIS Evolution 2015 Vision Tenets

<table>
<thead>
<tr>
<th>Vision Tenet</th>
<th>Vision 2015 Goals</th>
</tr>
</thead>
</table>
| **Archive Management**      | - NASA will ensure safe stewardship of the data through its lifetime.  
- The EOS archive holdings are regularly peer reviewed for scientific merit. |
| **EOS Data Interoperability** | - Multiple data and metadata streams can be seamlessly combined.  
- Research and value added communities use EOS data interoperably 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.                                                                                                      |
Features of EOSDIS Evolution (1 of 3)

• Migration to commodity-based hardware
  – reduces maintenance and technology refresh costs

• Transfer responsibility for archive and distribution of MODIS (Moderate Resolution Imaging Spectroradiometer) Level 0, Level 1, and Atmosphere (Level 2) data from Goddard DAAC to MODIS data processing facility (MODAPs)
  – Move from archiving all products to on-demand production of Level 1
  – Decrease archive size and slow future growth; facilitates transition to disk-based archive (all data on-line)
  – Closer relationship/control by science community expected to be more responsive to science needs, products, tools
Features of EOSDIS Evolution (2 of 3)

• Rearchitect EOSDIS Core System (ECS) to reduce footprint; simplify system (removes 750K SLOC (current baseline 1.2M SLOC) and 15 S/W Components)
  – Reduced operations/sustaining engineering costs
  – Improved performance

• At Goddard and Langley DAACs, consolidate data holdings into single, in-house developed system at each DAAC (currently operating ECS and heritage systems at both DAACs)
  – Reduction in operations costs due to elimination of multiple systems
Features of EOSDIS Evolution (3 of 3)

- Earth Science ClearingHOuse (ECHO)
  - Independent reviews conducted
  - Prioritize efforts to focus on achieving higher operational maturity (ingest, reconciliation, search and order, performance, error handling)
  - Global Change Master Directory (GCMD) portal to ECHO underway
  - WIST The Warehouse Inventory Search Tool (WIST) (general ECHO client that searches across all of ECHO’s holdings) fully operational, and users are routinely accessing EOS data via WIST/ECHO.

- ROSES/ACCESS solicitation out for ECHO clients

- Current Holdings
  - Collections 2,237
  - Granules 56 million
  - Browse 14 million
Key Benefits of Proposed “Step 1” Plan

✓ Maximize Science Value
  - Data access easier and data products quickly available to science community
  - MODIS data more closely integrated with science community
  - Potential pathfinder for migration of other data into science communities

✓ Substantial Cost Savings
  - Addresses operational and sustaining engineering
  - Takes advantage of current IT advances
  - Investments provide return on value within 3 years

✓ Manageable Risk
  - Minimizes software development efforts
  - Builds upon existing systems
  - Utilizes steps within plan as proof of value before proceeding
  - Reduces footprint for EOSDIS Core System
EOSDIS Data Science Examples

• Arctic sea Ice
  – Animation (20 second) showing the Spring retreat and subsequent Autumn advance of sea ice over the Arctic from 10/1/2002 through 4/21/2003 using AMSR-E data
  – Highlights fissures in the sea ice.

• MODIS High Resolution Cloud Detection
  – Improved resolution from GOES to MODIS
  – National Weather Service, NOAA and NASA working together to improve severe weather prediction

• 30 day fire potential forecast
  – Historical data mined 2001-2003 to develop model
  – Correlated with actual fire data from MODIS – black dots
NASA and National Weather Service formed Short-term Prediction Research and Transition Center (SPoRT) to help forecasters better incorporate satellite data into their decision making process. Shows specialized products using MODIS data.

- MODIS instrument provides higher resolution (250m-1,000m) measurements of cloud cover, and cloud surface detail and is being used to estimate cloud thickness, cloud height, and fog over specific regions.

- Together with NOAA and NWS, NASA is now providing near real-time precipitation and cloud data to help forecasters better anticipate the impact of storm systems far off shore to be detected by land-based radars.
Fire Potential Forecast

30-Day Fire Forecast Map for 3/31/2004

Log of Forecast Probability

-10 -9 -8 -7 -6 -5 -4 -3 -2 -1
**EOSDIS Future**

- Taking steps to facilitate data mining

- Evolution viewed as a very positive change;
  - Step 1 changes only partially address the 2015 vision
  - Substantial work yet to be done to fulfill the vision

- Paving the way to 1) Increased data usage; 2) Improved data access; 3) Increased participation

- NASA’s EOSDIS is one of the richest sources of Earth Science data and its mined contents hold the keys to knowledge and understanding of long term climate variability and the major processes that drive the Earth’s systems

- Door is OPEN and the opportunities are bright to Mine, Explore, Discover, Understand, Solve, and Amaze
## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIRS</td>
<td>Atmospheric Infrared Sounder</td>
</tr>
<tr>
<td>AMSR-E</td>
<td>Advanced Microwave Scanning for EOS</td>
</tr>
<tr>
<td>ASDC</td>
<td>Atmospheric Sciences Data Center</td>
</tr>
<tr>
<td>ASF</td>
<td>Alaska SAR Facility</td>
</tr>
<tr>
<td>AVHRR</td>
<td>Advanced Very High Resolution Radiometer</td>
</tr>
<tr>
<td>DAAC</td>
<td>Distributed Active Archive Center</td>
</tr>
<tr>
<td>DoD</td>
<td>Department of Defense</td>
</tr>
<tr>
<td>DMSF</td>
<td>Defense Meteorological Satellite Program</td>
</tr>
<tr>
<td>ECS</td>
<td>EOSDIS Core System</td>
</tr>
<tr>
<td>EDOS</td>
<td>EDS Data and Operations System</td>
</tr>
<tr>
<td>EOS</td>
<td>Earth Observing System</td>
</tr>
<tr>
<td>EOSDIS</td>
<td>EOS Data and Information System</td>
</tr>
<tr>
<td>ESE</td>
<td>Earth science Enterprise</td>
</tr>
<tr>
<td>ESDIS</td>
<td>Earth Science Data and Information System</td>
</tr>
<tr>
<td>GES DAAC</td>
<td>GSFC Earth Sciences DAAC</td>
</tr>
<tr>
<td>GHRC</td>
<td>Global Hydrology Resource Center</td>
</tr>
<tr>
<td>GSFC</td>
<td>Goddard Space Flight Center</td>
</tr>
<tr>
<td>JPL</td>
<td>Jet Propulsion Laboratory</td>
</tr>
<tr>
<td>LaRC</td>
<td>Langley Research Center</td>
</tr>
<tr>
<td>LP DAAC</td>
<td>Land Processes DAAC</td>
</tr>
<tr>
<td>LTA</td>
<td>Long Term Archive</td>
</tr>
<tr>
<td>MODIS</td>
<td>Moderate Resolution Imaging Spectroradiometer</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NCEP</td>
<td>National Centers for Environmental Prediction</td>
</tr>
<tr>
<td>NESDIS</td>
<td>National Environmental Satellite, Data, and Information Service</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic And Atmospheric Administration</td>
</tr>
<tr>
<td>NPP</td>
<td>NPOESS Preparatory Project</td>
</tr>
<tr>
<td>NPOESS</td>
<td>National Polar Orbiting Environmental Satellite System</td>
</tr>
<tr>
<td>NRTPE</td>
<td>Near Real Time Processing Effort</td>
</tr>
<tr>
<td>NSIDC</td>
<td>National Snow and Ice Data Center</td>
</tr>
<tr>
<td>ORNL</td>
<td>Oak Ridge National Laboratory</td>
</tr>
<tr>
<td>PB</td>
<td>Peta Byte</td>
</tr>
<tr>
<td>PO DAAC</td>
<td>Physical Oceanography DAAC</td>
</tr>
<tr>
<td>SAR</td>
<td>Side Aperature Radar</td>
</tr>
<tr>
<td>SEDAC</td>
<td>Socio-Economic Data Applications Center</td>
</tr>
<tr>
<td>SSM/I</td>
<td>Special Sensor Microwave/Imager</td>
</tr>
<tr>
<td>TB</td>
<td>Tera Byte</td>
</tr>
<tr>
<td>TRMM</td>
<td>Tropical Rainfall Measuring Mission</td>
</tr>
<tr>
<td>USGS</td>
<td>United States Geological Survey</td>
</tr>
</tbody>
</table>