The NASA Earth Observing System Data and Information System (EOSDIS)

NASA Earth Science Data & Information Systems Project
June 04, 2014
# Data Sources

<table>
<thead>
<tr>
<th>Type</th>
<th>Example Missions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite/on-orbit Missions</td>
<td>Terra, Aqua, Aura, Suomi-NPP</td>
</tr>
<tr>
<td>Airborne missions</td>
<td>IceBridge, Earth Ventures (5+ missions)</td>
</tr>
<tr>
<td>In Situ Measurement missions</td>
<td>Field campaigns on land (e.g., LBA-ECO) and in the ocean (e.g., SPURS)</td>
</tr>
<tr>
<td>Applications support</td>
<td>Near-real time creation and distribution of selected products for applications communities</td>
</tr>
<tr>
<td>Earth Science Research support</td>
<td>Research products from ROSES efforts like MEaSUREs. This also includes data from older, heritage missions (prior to EOS Program) that the DAACs rescue – e.g., Nimbus, SeaSat</td>
</tr>
</tbody>
</table>
Role of EOSDIS

“Advance knowledge of Earth as a system to meet the challenges of environmental change, and to improve life on our planet.” -- 2014 NASA Strategic Plan

- NASA’s Earth Science Data Systems directly support this objective by providing end-to-end capabilities to deliver data and information products to users

- NASA’s Earth Science Data Policy promotes usage of data by the community
  - No period of exclusive access
  - Data available at no cost to all users on a non-discriminatory basis, except where agreed upon with international partners

- EOSDIS provides:
  - Interoperable Distributed Data Archives
  - Science Data Processing
  - Data Management
  - On-Line Data Access Services
  - Earth Science Discipline-Oriented User Services
  - Network Data Transport to distributed System Elements
# Earth Science Data Operations

## Mission Operations

<table>
<thead>
<tr>
<th>Data Acquisition</th>
<th>Flight Operations, Data Capture, Initial Processing, Backup Archive</th>
<th>Data Transport to Data Centers/SIPSs</th>
</tr>
</thead>
</table>

- **EOS Spacecraft**
- **Tracking and Data Relay Satellite (TDRS)**
- **White Sands Complex (WSC)**
- **EOS Polar Ground Stations**
- **Direct Broadcast (DB)**
- **Direct Broadcast/Direct Readout Stations**

## Science Operations

<table>
<thead>
<tr>
<th>Science Data Processing, Data Management, Interoperable Data Archive, and Distribution</th>
<th>Distribution and Data Access</th>
</tr>
</thead>
</table>

- **EOSDIS Data Centers**
- **NASA Integrated Services Network (NISN) Mission Services**
- **Instrument Teams and Science Investigator-led Processing Systems (SIPSs)**
- **Research**
- **Education**
- **Value-Added Providers**
- **Interagency Data Centers**
- **Earth System Models**
- **International Partners**
- **Decision Support Systems**

## Value-Added Providers

- **Earth System Models**
- **Decision Support Systems**
Earth Science Data and Information System (ESDIS) Project - Code 423

Project Manager
Dawn Lowe

Deputy Project Manager/Operations
Jeanne Behnke

Deputy Project Manager/Technical
Jeff Walter

Deputy Project Manager/Resources
Diane Hronek

Assistant Project Manager
H.K. “Rama” Ramapriyan

System Security
Clayton Sigman

System Manager
Karen Michael

Financial Manager
Chikia Barnes

System Architect
Kevin Murphy

Science Systems Development Office
Manager - Dan Marinelli
Deputy - Andy Mitchell

SNPP Science Data Segment
Alfreda Hall
Eunice Eng
Evelyn Ho

Science Operations Office
Manager - Jeanne Behnke
Deputy - Drew Kittel

Systems Engineering
Steve Berrick

Development & Sustaining Engineering
Luther Lighty
Ryan Boller

Network System Engineer
Kevin Kranacs

DAAC Operations
Nate James
Dave Batchelor
vacant (2)

January 2014
Science Processing Systems for EOSDIS

EOSDIS science processing systems (SIPSs, SCFs, and DAAC-based) perform forward processing of standard products, and reprocess data to incorporate algorithm improvements.

Production Data for ECS DAACs
DAACs ingest, archive, process and distribute data to users.
• The DAAC is located at the University of Colorado in Boulder, CO at the National Snow and Ice Data Center.
• Ingests, archives and distributes spaceborne data from AMSR-E, ICESat/GLAS, MODIS on Terra and Aqua (via ECS)
  • Preparing for SMAP launch 2014
• Ingests, archives and distributes airborne data from NASA’s ICEBridge missions
• Provides tools and services for discovery of cryosphere data
• Systems consist of a mix of servers, 465 TB of disk storage, and 157 TB of tape storage (for back-up)
The LP DAAC is located at the U.S. Department of the Interior, U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center in Sioux Falls, SD.

Processes ASTER data from Terra including the ASTER Global DEM via “S4PM” (Open source system developed by the GES DISC).

Archives and distributes ASTER data via ECS.

Archives and distributes MODIS from Terra and Aqua via ECS.

Provides tools and services for discovery and analysis of NASA’s land cover and land use data.

Systems consist of a mix of servers, 2.4 PB of disk storage, and 1 PB of tape storage (for backup).
Atmospheric Science Data Center (ASDC)

- The ASDC is in the Science Directorate at NASA Langley Research Center, in Hampton, VA.
- Processes production science data products for CERES and MISR science teams via "S4PM"
- Archives and distributes data from MISR, SAGE III, MOPITT, and TES via ECS
- Archives and distributes CERES via locally developed ANGe
- Provides sensor-specific search tools as well as more general tools and services such as atmosphere product subsetting
- Systems consist of a mix of servers, 1.3 PB of disk storage, and 931 TB of tape storage (for backup)
### ECS Metrics for DAACs

#### Archive Volume

<table>
<thead>
<tr>
<th>DAAC Total</th>
<th>2014 Granules</th>
<th>GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASDC</td>
<td>20,368,172</td>
<td>900,698</td>
</tr>
<tr>
<td>LPDAAC</td>
<td>74,703,267</td>
<td>2,180,462</td>
</tr>
<tr>
<td>NSIDC</td>
<td>17,706,152</td>
<td>250,894</td>
</tr>
</tbody>
</table>

#### Daily Ingest

<table>
<thead>
<tr>
<th>DAAC Total</th>
<th>2014 Granules</th>
<th>MB</th>
<th>Peak Day Granules</th>
<th>MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASDC</td>
<td>5,091</td>
<td>293,389</td>
<td>22,273</td>
<td>2,322,227</td>
</tr>
<tr>
<td>LPDAAC</td>
<td>22,051</td>
<td>491,504</td>
<td>149,291</td>
<td>5,460,095</td>
</tr>
<tr>
<td>NSIDC</td>
<td>9,364</td>
<td>183,925</td>
<td>208,875</td>
<td>1,927,491</td>
</tr>
</tbody>
</table>

#### Daily Distribution

<table>
<thead>
<tr>
<th>DAAC</th>
<th>2014 Granules</th>
<th>MB</th>
<th>Peak Day Granules</th>
<th>MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASDC</td>
<td>34,586</td>
<td>1,286,579</td>
<td>228,162</td>
<td>9,010,713</td>
</tr>
<tr>
<td>LPDAAC</td>
<td>483,125</td>
<td>5,431,609</td>
<td>2,000,778</td>
<td>34,526,159</td>
</tr>
<tr>
<td>NSIDC</td>
<td>95,750</td>
<td>513,541</td>
<td>861,129</td>
<td>5,488,084</td>
</tr>
</tbody>
</table>

Archive – total size of the collection of data in the ECS system at each DAAC. As new data is processed, it is added to the DAACs. Some interim files and datasets may be deleted on a daily basis.

Ingest – this represents the number of files that are put into the archive on a daily basis.

Distribution – this represents all data distributed out of the ECS system on a daily basis, typically through FTP or HTTP at each DAAC.
Elements and Capabilities

- **Science Data Processing System**: The Science Data Processing System is known also known as the EOSDIS Core System (ECS) supporting the ingest, archive and distribution of data at 3 DAACs.

- **EarthData**: the EOSDIS website (*Earthdata.nasa.gov*) provides a focal point for cross-DAAC and EOSDIS content and news sharing, and access to Earth science data and services

- **ECHO/Reverb**: The EOS ClearingHouse (ECHO) is a metadata catalog of NASA's EOS data and a registry for related data services (e.g. reformatting, pattern recognition). Reverb allows users to search science data holdings, view browse images, and submit orders via ECHO to the appropriate data providers

- **Common Metadata Repository**: The Common Metadata Repository (CMR) builds on the work done by ECHO and the GCMD to provide a unified, authoritative repository for NASA's Earth Science metadata.

- **User Registration System**: provides a centralized and simplified mechanism for user registration and account management for all EOSDIS system components.
Science Data Processing System (SDPS) Context

SDPS Context Diagram
Q1 2014 v1
Earthdata is aimed at improving EOSDIS web presence and services, as well as representing EOSDIS as an efficient and interoperable program to support NASA’s Earth science research.

Earthdata reorganized the information into an intuitive, logical structure facilitating users’ discovery of the wide array of services and content offered throughout EOSDIS (refer to http://earthdata.nasa.gov).
“ECHO Architecture”

**ECHO is NASA’s middleware layer between Earth science data and users via a service-oriented architecture. Designed to improve the discovery and access of NASA data.**

**ECHO Architecture**

- **API SERVICES**
- **DATA / SERVICE REGISTRY**
- **CATALOG**
- **INGEST**

**ECHO Clients**

- Reverb
  - Primary Web Based Client

**Search, Browse, Order**

- Acts as an order broker between end users and EOSDIS Data Centers who provide metadata for their data holdings and other Earth science-related data holdings.

- User-defined specialized “clients” can be easily developed to give science data users of their community access to data and services using ECHO’s open APIs.

**Metadata, Browse, Orders**

- CCDIS
- GES DAAC
- ASF
- NSIDC
- ORNL

- OBPG
- PO.DAAC
- GHRC

- SEDAC
- LAADS
- LP DAAC

**ECHO is NASA’s middleware layer between Earth science data and users via a service-oriented architecture. Designed to improve the discovery and access of NASA data.**
ECHO Holdings and Availability

- ECHO maintains a +98% uptime in all operational systems

As of May 24, 2014

- Current Holdings:
  - Datasets: 3,855 total
  - Granules: 157 million
  - Browse: 76 million total

- Registered Users: 86,322 total
Recent ECHO Improvements

- ECHO was re-architected to provide full redundancy of all critical components, virtually eliminating need for preventive maintenance downtime, and increased stability.
  - Average operational availability: 99%

- ECHO search performance has been significantly improved in the past 5 years: Average query time now 4 seconds (from 30 seconds)
  - Will be implementing sub-second performance

- Average Ingest performance improved by a factor of 10

- ECHO is working to simplify efforts of Client developers:
  - Added 2 new API formats last year: OpenSearch and REST
  - Benefits: Lightweight architecture and ease-of-use for Client developers
Reverb is the primary web-based client for discovering and ordering cross-discipline data from all of ECHO’s metadata holdings.

Reverb allows users, including those without specific knowledge of the data, to search science data holdings, retrieve high-level descriptions of data sets and detailed descriptions of the data inventory, view browse images, and submit orders via ECHO to the appropriate data providers.

Users are able to submit queries using spatial and temporal criteria and examine search results for relevancy using built-in tools.

http://earthdata/reverb
End user applications will continue to work
EOSDIS User Registration System (URS)

- Consolidation of Similar Registration Systems into an EOSDIS Wide User Registration System

- Improve the User Experience
  - Simplified and Consistent User Registration & Authentication
  - Integrated with Coherent Web (Earthdata.NASA.gov)
  - [https://urs.eosdis.nasa.gov/](https://urs.eosdis.nasa.gov/)

- Standardized Method of Metrics Collection & Reporting (via EOSDIS Metrics System - EMS)
  - Understand User Demographics and Access Patterns

- Enable Status Change Notifications to Users
  - By access pattern, data product, site, application, etc.

- Establish Framework for Future Capabilities
  - User Tailoring, Customized Views
  - Saved Queries, Order Management

- Goal for phase-in by all DAACs is mid 2015
As of May 24, 2014

- Total Users: 107,818 (Increasing 400+ new users per week)
- Authentications: Averaging 1.2 million weekly
- Operational Systems: ECHO (Reverb), LANCE, Earthdata (including ECE & wiki), EMS, GCMD, LP DAAC, OBPG (MERIS) and ASF (more to come…)

https://urs.eosdis.nasa.gov/
EOSDIS ACSI Customer Satisfaction Survey 2013: Relative Rankings

- EOSDIS sponsors an annual independent customer survey in conjunction with the American Customer Satisfaction Index (ACSI)
- EOSDIS consistently exceeds the Federal Government average
- Ratings in the mid to upper 70s are considered “very good” by the rating organization, the CFI Group
- 2013 Survey results based on 4,146 responses (~4.3%)
- Comments in surveys help define system improvements

![Survey Results Bar Chart]

Federal Government Overall (2013) - 66
NASA EOSDIS - 76
E-Business Internet News & Information - 73
Example EOSDIS Metrics

**EOSDIS Metrics FY2013**
*(Oct 1, 2012 to Sept 30, 2013)*

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique Data Products</td>
<td>6,861</td>
</tr>
<tr>
<td>Distinct Users of EOSDIS Data and Services</td>
<td>1.7 M</td>
</tr>
<tr>
<td>Average Daily Archive Growth</td>
<td>8.5 TB/day</td>
</tr>
<tr>
<td>Total Archive Volume</td>
<td>9.8 PB</td>
</tr>
<tr>
<td>End User Distribution Products</td>
<td>839 M</td>
</tr>
<tr>
<td>End User Average Daily Distribution Volume</td>
<td>22 TB/day</td>
</tr>
</tbody>
</table>

**Distribution by Discipline**
*(Oct 2012 - Sep 2013)*

- Atmosphere, 224,162
- Geodesy/Solid Earth, 121,723
- Land, 109,374
- Ocean, 93,039
- Hydrology, 83,615
- Radiance/Geolocation, 50,790
- Raw Data, 21,860
- Others, 83,413
- Terrestrial Ecology, 51,373
- Raw Data, 21,860

**Multi-year Total Archive Volume (PBs) Trend**

**EOSDIS Product Distribution Trend**

**EOSDIS Science Data**
*Volume Growth*

**EOSDIS Science Data Products Distribution**
*FY2000 through FY2013*
Summary

- EOSDIS is one of the largest science information systems in the world.
- The ECS/SDPS, EarthData, ECHO, Reverb, User Registration are critical components of the EOSDIS.
  - The ECS SDPS archives > 3 petabytes of data across three DAACs, and distributes an average of ~ 7 terabytes of data per day from NASA’s EOS and other missions
  - ECS is highly tuned to support efficient transfer of high data volumes
  - Earthdata, ECHO, Reverb, User Registration support all EOSDIS DAACs, and all EOSDIS data holdings.
- Continued reliable, responsive, and efficient service by EOSDIS to NASA’s Earth science data users is essential.